



**KTH Electrical Engineering**

# **Robust inference of gene regulatory networks**

System properties, variable selection, subnetworks, and design of experiments

Version 2013-4-29

TORBJÖRN E. M. NORDLING

Doctoral Thesis  
Stockholm, Sweden 2013

TRITA-EE 2013:019  
ISSN 1653-5146  
ISBN 978-91-7501-762-4

KTH School of Electrical Engineering  
Automatic Control Lab  
SE-100 44 Stockholm  
SWEDEN

Akademisk avhandling som med tillstånd av Kungliga Tekniska högskolan framlägges till offentlig granskning för avläggande av teknologie doktorsexamen i reglerteknik tisdagen den 28 maj 2013 klockan 13.00 i sal D3 Kungliga Tekniska högskolan, Lindstedtsvägen 5, Stockholm.

© Torbjörn E. M. Nordling, April 2013. All rights reserved.

Tryck: Universitetsservice US AB

## Abstract

In this thesis, inference of biological networks from *in vivo* data generated by perturbation experiments is considered, *i.e.* deduction of causal interactions that exist among the observed variables. Knowledge of such regulatory influences is essential in biology.

A system property—interampattiness—is introduced that explains why the variation in existing gene expression data is concentrated to a few “characteristic modes” or “eigengenes”, and why previously inferred models have a large number of false positive and false negative links. An interampatte system is characterized by strong INTERactions enabling simultaneous AMPLification and ATTEnuation of different signals and we show that perturbation of individual state variables, *e.g.* genes, typically leads to ill-conditioned data with both characteristic and weak modes. The weak modes are typically dominated by measurement noise due to poor excitation and their existence hampers network reconstruction.

The excitation problem is solved by iterative design of correlated multi-gene perturbation experiments that counteract the intrinsic signal attenuation of the system. The next perturbation should be designed such that the expected response practically spans an additional dimension of the state space. The proposed design is numerically demonstrated for the *Snf1* signalling pathway in *S. cerevisiae*.

The impact of unperturbed and unobserved latent state variables, that exist in any real biological system, on the inferred network and required set-up of the experiments for network inference is analysed. Their existence implies that a subnetwork of pseudo-direct causal regulatory influences, accounting for all environmental effects, in general is inferred. In principle, the number of latent states and different paths between the nodes of the network can be estimated, but their identity cannot be determined unless they are observed or perturbed directly.

Network inference is recognized as a variable/model selection problem and solved by considering all possible models of a specified class that can explain the data at a desired significance level, and by classifying only the links present in all of these models as existing. As shown, these links can be determined without any parameter estimation by reformulating the variable selection problem as a robust rank problem. Solution of the rank problem enable assignment of confidence to individual interactions, without resorting to any approximation or asymptotic results. This is demonstrated by reverse engineering of the synthetic IRMA gene regulatory network from published data. A previously unknown activation of transcription of *SWI5* by *CBF1* in the IRMA strain of *S. cerevisiae* is proven to exist, which serves to illustrate that even the accumulated knowledge of well studied genes is incomplete.



*To my mother and grandmother*



---

# Acknowledgements

---

It is my pleasure to finally acknowledge at least a subset of all the people that either directly or indirectly have contributed to the realization of this thesis. First, I would like to thank my supervisor—Elling Jacobsen—who has given me the opportunity to write this thesis. Already during the basic course in control theory 2001, you showed me the importance of being in control and have during the past years made sure I never will forget it. You have continued to inspire me by your deep knowledge of robust control, in particular.

Second, I would like to thank my collaborators: Andreas Tjärnberg, Kristoffer Forslund, Matthew Studham, Sanjit Roopra, and Erik Sonnhammer at Stockholm University. Noriko Hiroi, Akira Funahashi, and Hiroaki Kitano at Keio University. Sven Nelander, Rebecka Jörnsten, Tobias Abenius, Teresia Kling, and Linnéa Schmidt at either Uppsala University, Chalmers University of Technology, or University of Gothenburg. Mikko Hellgren and Jan-Olov Höög at Karolinska Institutet. None of the work that we have done together is included in this thesis, but I have enjoyed every discussion and learnt a lot from working with you. I would also like to thank Sven, Erik, Rebecka, Noriko, and Akira for the perspective and strength you have given me during the toughest hours.

Many thanks to Diego di Bernardo for the invitation to your group in Naples and all group members—Irene Cantone, Mukesh Bansal, Alberto Ambesi-Impiombato, Lucia Marucci, Giusy Della Gatta, Giulia Cuccato, Vincenzo Belcastro, Francesco Iorio, Velia Siciliano, and Maria Pia Cosma—for making it an enjoyable and memorable experience. Things did not go exactly as planned, but I have learnt that they seldom do in research. From now on, I at least know that there are pizzas and the original Margarita. A special thanks to Luca Quagliata for being a good friend and finding me an apartment.

To all of my colleagues at the Automatic Control Lab at KTH, thank you for impressing me with your diverse knowledge, for helping me when L<sup>A</sup>T<sub>E</sub>X doesn't and for making it such a welcoming workplace when I started. Thanks to Cristian Rojas, Alberto Speranzon, Carlo Fischione, Emmanuel Witrant, and Maben Rabi for philosophical debates and moral support. Thanks to Henrik Sandberg, Per Sahlholm, and Winston Garcia-Gabin for interesting scientific discussions. A special thanks to Camilla Trané, Magnus Lindhé, Assad Alam, and Chitrupa Ramesh for spontaneous discussions, help with all unimaginable small things, and sharing office with me. Thanks to Oscar Flärdh, Erik Henriksson, Jonas Mårtensson, Christian Larsson, Märta Barenthin, and Farhad Farokhi for cheering up the lunches and fika corner.

Thanks to Niels Möller, Paul Sundvall, Andreas Stenhall, Nina Unkuri, and Niclas Horney for sharing your knowledge on how to keep a computer running. Thanks also to Karin Karlsson Eklund, Anneli Ström, Kristina Gustafsson, and Hanna Holmqvist for your great spirit and help with all kinds of administrative issues. I nowadays count many of you as my friends.

For the actual production of this thesis I would like to thank Cristian Rojas, Voichita Marinescu, Erik Sonnhammer, Carlo Fischione, Andreas Tjärnberg, Patrik Johansson, Matthew Studham, Camilla Trane, Georg Metsalo, Peter and Lina Lindqvist, Maria Fröling, Maben Rabi, Farhad Farokhi, Sadegh Talebi, Niclas Blomberg, Kristen Liu, Niklas Everitt, Martin Andreasson, Rebecka Jörnsten, and Sven Nelander. Your careful scrutiny of the drafts and thoughtful comments have significantly raised the quality of the text and provided insights that will be useful later on.

This research project was partly supported by the Swedish Research Council, which I am very thankful for. It is fantastic that at least one research financier still dare to let young scientists do real science, having only a title as their guiding star.

I also want to use this opportunity to acknowledge the invaluable support and knowledge provided by my former teachers, in particular, Jarmo Alander, Gunnar Benediktsson, Göran Grimvall, Hans Grandell, Gunilla Eriksson, Camilla Niiranen, Märta Westerback, Christel Bishoff, and Inga-lill Mickos. A special thanks to all my friends, especially, Maria Fröling, Maria Rojas, Peter and Lina Lindqvist, Andreas Thomasson, Henrik Antell, Björn Forss, Cecilia Wik, Padideh Kamalizare, Johan and Yuko Elg, Johan Glimming, Gustav Henter, Martin Norström, Fredrik Feurst, Michael Andersson, Sami Muhonen, Carl Johnzen, Magnus Nilsson, Peter Modin, and my closest friend through the years in Stockholm, Finn Lennartsson. It is a true privilege to call you my friends.

Last, but not least, the people that are closest to me and that have suffered the most from the production of this thesis. Tack, kära Mamma, Pappa, Melli och Lina för er ovillkorliga omtanke och acceptans av min frånvaro. Dear Sunny, without your love and care the world would have been a harsh place.

*Torbjörn E. M. Nordling*  
Stockholm, April 2013.

---

# Contents

---

<b>Abstract</b>	<b>iii</b>
<b>Acknowledgements</b>	<b>vii</b>
<b>Contents</b>	<b>ix</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Motivation and purpose . . . . .	1
1.2 Network inference and variable selection . . . . .	4
1.3 Experiment design . . . . .	9
1.4 Properties of biological systems . . . . .	13
1.5 Publications . . . . .	14
1.6 Outline . . . . .	17
<b>2 From data to linear models of gene regulatory networks</b>	<b>19</b>
2.1 A gene regulatory network—the IRMA example . . . . .	20
2.2 Gene expression data—real-time RT-PCR for mRNA quantification . . . . .	23
2.3 Linear models of gene regulatory networks . . . . .	27
<b>3 Interampattiness—a property of biological networks</b>	<b>33</b>
3.1 Introduction . . . . .	34
3.2 A signals and systems perspective on biological networks . . . . .	38
3.3 Features of strongly interampatte systems . . . . .	45
3.4 Interampattiness depends on network motifs . . . . .	48
3.5 Interampattiness related to time-scale separation . . . . .	51
3.6 Implications of interampattiness for reverse engineering of GRNs . . . . .	51
3.7 Evidence of strongly interampatte bio-networks . . . . .	54
3.8 Conclusions . . . . .	57
<b>4 Robust variable selection for network inference</b>	<b>59</b>
4.1 Introduction . . . . .	60
4.2 Problem formulation and introductory example . . . . .	63

4.3	Network inference is variable selection . . . . .	69
4.4	As many samples as variables are needed to infer sparse networks . . . . .	70
4.5	Inference of existing interactions with confidence . . . . .	74
4.6	Concluding examples . . . . .	77
4.7	Conclusions . . . . .	82
<b>5</b>	<b>Theory of robust variable selection and network inference</b>	<b>85</b>
5.1	Introduction . . . . .	86
5.2	The variable selection problem . . . . .	90
5.3	The network inference problem . . . . .	93
5.4	Introduction of uncertainty sets . . . . .	96
5.5	Robust tools for analysis of variable selection . . . . .	102
5.6	Practically selectable regressors . . . . .	116
5.7	Practically excludable regressors . . . . .	121
5.8	Practically negligible and alternative regressors . . . . .	124
5.9	Geometrical interpretation . . . . .	128
5.10	Testing for practical independence . . . . .	132
5.11	Conclusions . . . . .	136
5.12	Proofs . . . . .	139
<b>6</b>	<b>Inference of subnetworks—impact of latent states</b>	<b>149</b>
6.1	Introduction . . . . .	150
6.2	Problem definition and the IRMA gene regulatory network . . . . .	153
6.3	Conditions for inference of the full network . . . . .	157
6.4	Conditions for inference of a subnetwork . . . . .	160
6.5	Relation to previously published conditions . . . . .	163
6.6	Direct versus indirect interactions . . . . .	168
6.7	Discussion and conclusions . . . . .	170
<b>7</b>	<b>Design of perturbation experiments for network inference</b>	<b>175</b>
7.1	Introduction . . . . .	176
7.2	Published data with insufficient information . . . . .	179
7.3	Informative data for variable selection and network inference . . . . .	181
7.4	System properties necessitate iterative design . . . . .	186
7.5	Fundamental principle for design of experiments . . . . .	187
7.6	SVD design of steady-state experiments . . . . .	190
7.7	Conclusions . . . . .	193
<b>8</b>	<b>Summary and future work</b>	<b>197</b>
8.1	Our contribution towards solving the network inference puzzle . . . . .	197
8.2	Suggested work to realize the potential of network inference . . . . .	205
<b>A</b>	<b>Robust variable selection for network inference in depth</b>	<b>207</b>
A.1	Use of effective rank in network inference . . . . .	208

---

A.2	Illustration of variable selection and parameter estimation . . . .	209
A.3	Additional theorems with proofs . . . . .	210
A.4	Generation of synthetic data for the IRMA example . . . . .	214
A.5	Development of an error model for the IRMA data . . . . .	215
A.6	Robust inference of the IRMA network . . . . .	224
A.7	Additional information on the TOEL example . . . . .	229
A.8	Robust inference illustrated on the TOEL example . . . . .	233
<b>B</b>	<b>Inference of subnetworks in depth</b>	<b>245</b>
B.1	Construction of the extended IRMA example . . . . .	246
B.2	Transfer function basics . . . . .	265
<b>C</b>	<b>Design of perturbation experiments in depth</b>	<b>269</b>
C.1	Additional information on the Snf1 example . . . . .	270
C.2	Non-rejectable network models . . . . .	280
C.3	Simulation of perturbation experiments on the Snf1 example . .	314
	<b>Abbreviations</b>	<b>315</b>
	<b>Index</b>	<b>317</b>
	<b>Bibliography</b>	<b>321</b>



# Introduction

---

“*Wir müssen wissen,  
wir werden wissen.*”

David Hilbert, in an address to the Society of German  
Scientists and Physicians, Königsberg, 1930.

## 1.1 Motivation and purpose

Within a living cell thousands of different molecules interact. The genetic code—DNA—is transcribed into messenger RNA (mRNA), which either is translated into proteins or performs some regulatory functions (Lander, 2011). Proteins are folded, phosphorylated, and form complexes while they interact with DNA, RNA, lipids, metabolites, and other molecules found in a cell. Together they form a dynamic system that integrates both genomic information and environmental cues (Wolkenhauer et al., 2005b; Sontag, 2005; Csete and Doyle, 2002). This integration is essential for regulation and performance of all processes and functions in the cell. Knowledge of the genetic code *per se* only provides the blueprint for the system components (Benfey and Mitchell-Olds, 2008). Depending on the level of abstraction different graphical network representations are commonly used to describe this system, *e.g.* gene regulatory networks (GRNs) capture the regulatory interactions resulting from changes in gene expression (He et al., 2009; Hecker et al., 2009; Karlebach and Shamir, 2008; Christensen et al., 2007; Brazhnik et al., 2002). The structure of these networks reveals the feedback loops, feedforward loops and cascades that are fundamental for normal function and behaviour of the system—and life itself. The system, and its network structure, is altered by disease, *e.g.* cancer arises from multiple spontaneous and/or inherited mutations of the DNA that affects the control of cell growth and division (Hood et al., 2004; Roukos, 2010a). Some diseases, such as Huntington’s disease (Walker, 2007) or sickle-cell anaemia (Olowoyeye and Okwundu, 2010; Orkin and Higgs, 2010), are due to a mutation in a single gene, leading to a defect protein, and could at least in theory be cured by repairing the gene, but most diseases depend on many factors. The molecular interactions therefore need to be studied

from a system perspective to understand how genetic information and environmental cues give rise to function and to enable development of predictive, personalized, preventive, and participatory (P4) medicine (Barabási et al., 2011; Hood and Friend, 2011; Roukos, 2011, 2010b; Wolkenhauer et al., 2009; Ideker and Sharan, 2008). The World health organization estimated that 1.3 billion disability-adjusted life years were lost to disease in 2004 (World Health Organization, 2008), so new cures can have a tremendous impact.

Molecular biologists have since the discovery of DNA by Watson and Crick (1953) investigated which protein activates or represses transcription of which protein coding gene. They have in other words mapped out parts of the GRN—component by component and interaction by interaction—a very time consuming task that mainly yields qualitative information. Network inference offers the ability to automate this task by reverse engineering network models based on *in vivo* measurements of changes in gene expression, copy numbers, metabolite concentrations, protein abundance and phosphorylation, *etc.* (Tegnér and Björkegren, 2007; Cho et al., 2007; Crampin, 2006; Gardner and Faith, 2005; Goncalves and Warnick, 2008). Considering the large number of components, possible interactions, and dynamical modes in any given cell, it is clear that automated knowledge discovery is essential to speed up this task and gain new insight. In addition, both qualitative and quantitative models can be inferred from the quantitative data that today can be generated by perturbation experiments. This data is enabled by measurement techniques, such as whole genome shotgun sequencing (Fleischmann et al., 1995), real-time RT-PCR (Higuchi et al., 1992), DNA microarrays (Schemm et al., 1995), chromatin immunoprecipitation (ChIP-chip) technique (Ren et al., 2000; Iyer et al., 2001; Lieb et al., 2001), protein arrays (MacBeath and Schreiber, 2000; Zhu et al., 2001), and single molecule imaging in living cells (Schütz et al., 2000; Sako et al., 2000), that despite their introduction during the last two decades already have revolutionized biology. In particular, the advances in genome sequencing, which enabled the completion of the human genome project faster than expected (Venter et al., 2001; IHGSC, 2001, 2004), have already changed biology forever (Lander, 2011; Butler, 2010). But the long term impact of the other techniques and their future refinements should be equally important since they already have changed the way biologists do experiments and enabled quantitative modelling of intracellular processes (Thellin et al., 2009; Sako, 2006; Sauro and Kholodenko, 2004). The invention of high-throughput measurement techniques has already led to the emergence of computational/systems biology (Ideker et al., 2001; Kitano, 2002; Chuang et al., 2010). In systems biology it is recognized that the system contains so many interacting components that mathematical models are needed to analyse and comprehend it and in particular its dynamics (Kritikou et al., 2006; Way and Silver, 2007; Kholodenko, 2006; Mogilner et al., 2006). In other words, both the complexity of the system and the large amounts of data that already is coming out of labs make mathematical modelling key to advance biological knowledge and cure diseases. Even though many of the problems encountered in analysis of genetic data and modelling of biological systems fall within established fields, such as statistics, system identification, and pattern recognition, the vast

amount of data in terms of measured variables and scarcity in terms of number of data points make them unconventional (Wang et al., 2008). Improved and alternative methods for almost every aspect of mathematical modelling and network inference, in particular, are therefore needed.

Network inference has been intensively studied for the past decade, but many open problems remain (Hendrickx et al., 2011; He et al., 2009; Hecker et al., 2009). The accumulated biological knowledge is always represented by models—explicitly or implicitly. Our ability to connect different observations and findings, analyse them, and use them depends on the models that we construct. One should, however, remember that all models are approximations of the “true” system and the art of modelling is largely in finding useful levels of approximation. To stress this, Box and Draper (1987) have stated that “all models are wrong, but some are useful.” According to Kolch (2008), “there is a widespread misconception in the wet scientists camp that models just reproduce known behaviour”, and he stresses that models do not even need to be accurate to be useful in systems biology. Indeed, there are good reasons to believe that modelling of the molecular details of a whole organism, organ, or even cell is beyond the current technical capability and not even needed to cure complex diseases, like cancer. This work will therefore focus on inference of causal network models of subsystems responsible for some well defined biological functions or behaviours.

To model the subsystem that generates a function of interest, three basic problems must be solved: First, one needs to determine the components of the system that are essential for generation of the function. Second, one needs to design experiments that yield informative data for network inference. Finally, one needs to infer a model with the correct structure of the subnetwork based on observed data. Aspects of the first and third problems have been widely studied, while the second one largely has been neglected, but none of them have been satisfactorily solved as is evident *e.g.* from the reviews by Hecker et al. (2009); Karlebach and Shamir (2008); Bonneau (2008); Tegnér and Björkegren (2007); Styczynski and Stephanopoulos (2005); Filkov (2005); D’haeseleer et al. (2000). The purpose of this work is to solve the later two of these basic problems for the case with steady-state data recorded from perturbation experiments and thereby get a step closer to a technique for automated knowledge discovery and mathematical modelling, directly applicable in biological research. In this thesis four crucial issues are studied: properties of biological systems, robust variable selection and network inference, inference of subnetworks, and design of perturbation experiments. This work is largely based on the long tradition in automatic control of identifying system models, analysing models, and synthesizing control strategies for engineered systems.

A detailed background, describing the state of the art prior to this work, on network inference and variable selection, experiment design, and properties of biological systems is provided in the following three sections, in order to motivate this thesis. Contributions in terms of scientific publications that this thesis, in part, is based on is listed in Section 1.5. An outline of the thesis is provided in Section 1.6. Eager readers may however jump directly to the chapter of their greatest interest,

because the presentation in each chapter is largely self contained.

## 1.2 Network inference and variable selection

Interacting genes, proteins and metabolites form dynamical systems controlling cellular processes (Wolkenhauer et al., 2005b; Sontag, 2005; Csete and Doyle, 2002). The architecture of these systems varies among organisms, cell types, developmental phases, and environmental and epigenetic conditions (Szalay et al., 2007; Huang et al., 2009). Development of high-throughput technologies has during the last decade enabled a systems approach, which is necessary to gain a better understanding of the complex global behaviour of biological processes (Hecker et al., 2009). While the genetic code provides the blueprint for the system components, it is the context dependent and largely unknown interactions that generate specific biological functions (Benfey and Mitchell-Olds, 2008). Inference of direct causal interactions underlying a function is hence a key problem in systems biology (Brazhnik et al., 2002; Wolkenhauer et al., 2009; Hecker et al., 2009).

Network inference, which is also known as reverse engineering or network reconstruction, has attracted significant interest, with more than 800 articles published and indexed in Pubmed ([www.pubmed.org](http://www.pubmed.org)) during the past decade. Several of these studies claim to have demonstrated how the interactions existing in, in particular, gene regulatory networks and transcriptional networks can be inferred based on gene expression data. This claim is only in a minority of the studies based on data obtained from *in vivo* experiments in which all genes of interest are systematically perturbed and the resulting expression changes in all of them are measured, see *e.g.* Lorenz et al. (2009); Cantone et al. (2009); Gardner et al. (2003). In a majority of the studies the claim is based on data sets with fewer observations than variables and using methods that do not explicitly utilize information about the perturbations, see *e.g.* Faith et al. (2007); Bonneau et al. (2006); Schäfer and Strimmer (2005) and the recent reviews Emmert-Streib et al. (2012); Hecker et al. (2009); Bonneau (2008); Tegnér and Björkegren (2007); Cho et al. (2007). The quality of the obtained models are, however, questionable—if not in fact shown to be poor by the authors themselves. For instance, Lorenz et al. (2009) reported a mere 62% sensitivity and 69% precision with 24% of the predicted interactions having the opposite sign in their model of the Snf1 network in *S. cerevisiae*. Sensitivity measures the percentage of known interactions that are successfully inferred, while precision measures the percentage of inferred interactions that are consistent with known interactions. Moreover, benchmarking studies, such as the Dialogue for Reverse Engineering Assessments and Methods (DREAM), have shown that inference of gene regulatory networks usually results in a large fraction of false positives, *i.e.* inferred interactions absent in the “true” network, and false negatives, *i.e.* missed interactions present in the “true” network (Marbach et al., 2010; Stolovitzky et al., 2009). The consensus of the published works is nonetheless that network inference has potential, even though it is in general not clear what the necessary conditions for

correct reconstruction of the network of interest are and *e.g.* [Krishnan et al. \(2007\)](#) fairly recently conclude that reverse engineering of GRNs from expression data is an indeterminate problem. Poor performance on *in vivo* data is, however, at least partly explained by incomplete information and significant uncertainty. These are two of the following five characteristics of existing data sets used for reverse engineering of GRNs (see *e.g.* [Holter et al., 2000](#); [Alter et al., 2000](#); [Tegnér and Björkegren, 2007](#); [Faith et al., 2007](#); [Cosgrove et al., 2010](#); [Wu and Wu, 2010](#)): The data points are few compared to the number of genes and possible interactions, *i.e.* the dimensionality is high in terms of variables but low in samples. The measurement uncertainty is large both in the perturbations and responses, *i.e.* the signal to noise ratio (SNR) is low and the data is described by a so called errors-in-variables model (see *e.g.* [Söderström, 2007](#)). The variables are nearly collinear, *i.e.* the response matrices are ill-conditioned.

The primary objective in network inference is to determine which interactions exist and which do not. The secondary objective is to determine if existing interactions correspond to activations or repressions. Determination of the strength of the interactions is subordinate and typically neglected. The major focus in network inference has so far been on development of methods for selection of a single sparse network model, *i.e.* a model containing only a fraction of all possible interactions, that fit to available data when fewer data points than the number of nodes in the network are available. A multitude of traditional estimation and variable selection methods, as well as novel inference algorithms have been adopted and developed; see *e.g.* the review articles [Emmert-Streib et al. \(2012\)](#); [De Smet and Marchal \(2010\)](#); [He et al. \(2009\)](#); [Hecker et al. \(2009\)](#); [Karlebach and Shamir \(2008\)](#); [Li et al. \(2008b\)](#); [Bonneau \(2008\)](#); [Tegnér and Björkegren \(2007\)](#); [Goutsias and Lee \(2007\)](#); [Markowitz and Spang \(2007\)](#); [Cho et al. \(2007\)](#); [Styczynski and Stephanopoulos \(2005\)](#); [Doyle and Lauffenburger \(2005\)](#); [Filkov \(2005\)](#); [van Someren et al. \(2002\)](#); [de Jong \(2002\)](#); [D’haeseleer et al. \(2000\)](#). Fundamental differences in the theoretical motivation and principle of these algorithms make evaluation and comparison of them essential but at the same time difficult. The performance of inference algorithms has therefore mainly been investigated through benchmarking on simulated data ([Penfold and Wild, 2011](#); [Marbach et al., 2010](#); [Stolovitzky et al., 2009](#); [Hache et al., 2009](#); [Bansal et al., 2007](#)). Different algorithms were found to perform well in each of these studies and the winning algorithm appears to largely depend on the data set, but it is worth noting that only a small subset of all algorithms that have been proposed was investigated in each of these studies. It is however clear that most algorithms perform rather poorly, *e.g.* [Hache et al. \(2009\)](#) conclude that the performance of the tested methods is not good enough for inference of large GRNs in practice and their set of evaluated algorithms—ARACNe ([Basso et al., 2005](#)), ParCorA ([de la Fuente et al., 2004](#)), GNRevealer ([Hache et al., 2007](#)), Banjo ([Yu et al., 2004](#)), LDST ([Rangel et al., 2004](#)), and GeneNet ([Schäfer and Strimmer, 2005](#))—represents the major model formalisms in this area.

Theoretical limitations of inference based on perturbation data have also been discussed in a few articles, most notably in [Margolin and Califano \(2007\)](#); [Goncalves](#)

and Warnick (2008); He et al. (2009); Hendrickx et al. (2011). Necessary and sufficient conditions for dynamical structure reconstruction of linear time invariant (LTI) systems from knowledge of the transfer function from input to output have been established in Goncalves and Warnick (2008). The dynamical structure function was introduced in Goncalves et al. (2007) and differs from all other model formalisms used in inference of GRNs, so the implications of these conditions on inference are not clear. Implications of latent states on inference based on mutual information are discussed in Margolin and Califano (2007). They use a simple co-regulation model in which two genes are regulated by a single transcription factor (TF) to show that an edge erroneously will be inferred between the co-regulated genes instead of the edges between the TF and the genes, unless the unobserved latent protein state of the TF correlates strongly enough with the TF mRNA. They conclude that it is important to characterize the necessary correlation strength for correct inference, because the evidence for relatively weak correlation between the mRNA and protein of TFs is increasing. The most common model formalisms together with several underlying biological assumptions are discussed in He et al. (2009), who conclude that these assumptions in general are not valid. In a critical assessment of inference of metabolic networks from time-series data, Hendrickx et al. (2011) show that the fastest and/or slowest interactions sometimes cannot be inferred with current techniques and conclude that the required sampling frequency for inference of large metabolic networks is not met today.

Different inference algorithms, model formalisms, and data types typically yield different types of networks, so the interpretation of an interaction varies. *E.g.* inference of physical interactions versus regulatory influences are discussed in Gardner and Faith (2005). They point out that the physical approach seeks the regulatory protein factors that physically bind to the promoter of the regulated transcript, while the influence approach seeks transcripts whose concentration changes can explain the changes of a considered transcript.

Many different model formalisms, such as Boolean networks, Bayesian networks, generalized logical networks, ordinary differential equations (ODEs), and stochastic master equations, see *e.g.* Karlebach and Shamir (2008); de Jong (2002), are used to describe GRNs, but “all models can be interpreted as networks of interacting nodes” as Hecker et al. (2009) phrase it. A fully connected network with  $n$  nodes has  $n^2$  interactions, including self-loops, but according to the dominating view only a fraction of these interactions exist in biological networks and even fewer are active in a particular mode of the system. Biological networks are thus assumed to be sparsely connected. Most genes are regulated by a small number of other genes via transcription factors that binds to the promoter sequence; 2-4 in bacteria (McAdams and Arkin, 1998; Thieffry et al., 1998) and 5-10 in eukaryotes (Arnone and Davidson, 1997), implying that transcriptional networks typically are highly sparse. Also evaluation of the degree distribution, *i.e.* number of nodes with each number of links, of existing gene, metabolic, and protein networks show that they are sparse (Jeong et al., 2000, 2001; Lima-Mendez and van Helden, 2009). On the other hand, Brazhnik et al. (2002) consider the influence of the proteome and

metabolome level on GRNs and give the following three arguments against sparse connections: All genes encoding enzymes that control a metabolite will also influence the transcription of all genes that depend on the metabolite. Transcription depends on metabolic energy and the expression of each gene is therefore affected by all genes that encode enzymes of the energy metabolism. If the availability of RNA polymerase is low, then competition between the polymerase binding sites for polymerase would introduce interactions between genes. In principle any two molecules interact when they collide, but if the behaviour of the system was affected then it would be as random as the collisions. Another argument against full connectivity in large GRNs is that from a systems perspective it would be difficult to ensure robust function. Small changes of the interactions are known to make the overall system unstable even though all the local control loops are stable, unless the system is diagonally dominant (see *e.g.* Skogestad and Postlethwaite, 1996, ch. 10). Even though the degree of sparsity is debated, biological networks are in general considered sparse. The essence of the network inference problem is therefore selection of the active interactions based on experimental data, which is a problem known as feature selection, feature reduction, attribute selection, subset selection, variable selection or model selection in the machine learning, statistics and system identification literature (Guyon and Elisseeff, 2003; Fan and Lv, 2010; Hara and Sillanp, 2009; George, 2000; Hong et al., 2008; Stoica and Selen, 2004). The term variable selection is mainly used here, because the features in biological data sets typically correspond to observed variables. From a parameter estimation perspective the following three reasons imply that network inference is not a parameter estimation problem, even though many variable selection methods require estimation of model parameters for evaluation of the fit to data, *e.g.* use of the Akaike information criterion and Bayesian information criterion (Cedersund and Roll, 2009; Stoica and Selen, 2004; Akaike, 1973; Schwarz, 1978), and some implicitly select the variables during the parameter estimation, *e.g.* LASSO (Tibshirani, 1996; Candès and Plan, 2009). First, in general, model and variable selection precedes parameter estimation and hence parameter estimates, as well as estimates of their variance and significance, are only valid for a selected model structure (Burnham and Anderson, 2002, p. 14). Second, only parameters corresponding to active interactions are nonzero and need to be estimated. The parameter estimator needs the correct degrees of freedom, *i.e.* number of parameters, in order to efficiently use the data, since *e.g.* the total parameter variance grows with the degrees of freedom (Kay, 1993, p. 34,42-43). Third, if the number of parameters exceeds the number of data points, then some parameters are structurally/*a priori* unidentifiable and no unique solution exists to the parameter estimation problem (Ashyraliyev et al., 2009; Faller et al., 2003; Bellman and Åström, 1970). The distinction between variable selection and parameter estimation is particularly important in network inference, since the number of possible interactions, *i.e.* unknown parameters, typically exceeds the number of measurements and the networks are known to be sparse. This renders any formulation as a parameter estimation problem underdetermined, which in general leads to indeterminate parameters, over-fitting, *i.e.* incorporation of noise specific to the data

set in the parameter estimates, and a large set of models with different structure that fits equally well to the data.

Many different solutions to the variable selection problem have been developed and employed in diverse fields, such as machine learning (Liu et al., 2010; Guyon and Elisseeff, 2003), statistics (Fan and Lv, 2010; Fan and Li, 2006; George, 2000; Hara and Sillanp, 2009; Fu and Desmarais, 2010), signal processing (Candes and Wakin, 2008), system identification (Hong et al., 2008; Stoica and Selen, 2004), spectroscopy (Xiaobo et al., 2010; Koljonen et al., 2008), intrusion detection (You et al., 2006), econometrics (Owen, 2003), pharmacology (Li et al., 2008a), bioinformatics and systems biology (Zeng et al., 2009; Hecker et al., 2009; Schwender et al., 2008; Smit et al., 2008; Saeys et al., 2007), so one also needs to look at works outside of systems biology. The importance of variable selection has increased over the past decade due to high-throughput measurement techniques, such as DNA sequencing and microarrays, and digitalization of information, see *e.g.* Fan and Li (2006); Kohavi and Provost (2001). The number of considered variables has grown by a factor 1000 since 1997 when few domains explored more than 40 variables (Guyon and Elisseeff, 2003). This has spurred a tremendous ongoing research effort, in particular in the bioinformatics and statistics community. The following conclusions taken from recent reviews and overviews clearly show that variable selection is still largely an unsolved fundamental problem of science. First note that brute-force search for an optimal variable subset is an NP-hard problem and thus in general not feasible (Yusta, 2009; Cotta et al., 2004; Kohavi, 1995; Cover and Van Campenhout, 1977). Current selection methods are “either computationally feasible but far from optimal, or they are optimal or almost optimal but cannot cope with the computational complexity of feature selection problems of realistic size” (Yusta, 2009). The most relevant variables are often suboptimal for prediction, while seemingly irrelevant variables taken together can provide good prediction (Fu and Desmarais, 2010; Guyon and Elisseeff, 2003; Kohavi and John, 1997). On the other hand, Freedman’s paradox states that if the number of explanatory variables is large relative to the number of data points, then a highly significant subset for prediction of any target variable can be selected, even if the target variable is unrelated (Burnham and Anderson 2002, p. 17; Lukacs et al. 2009; Freedman 1983). Fan and Fan (2008) have recently demonstrated that classification using any linear discriminants can perform as poorly as random guessing due to noise accumulation in population centroids estimated in a high-dimensional variable space and that selection of a subset of important variables is necessary for correct classification. High dimensional statistical learning requires development of robust and user-friendly algorithms and analysis of them and existing techniques, solving issues such as selection of data-driven penalty functions and algorithm parameters, group variable selection, incorporation of information on covariates, characterization of optimality properties, and lack of confidence in selected models and estimated parameters (Fan and Lv, 2010). Collinearity of variables causes over-fitting and selection of non-informative variables, in particular in high dimensions where the number of variables exceeds the number of observations (Fan and Lv, 2010). Many methods for selection of subsets of variables are sensitive

to noise or small perturbations in the data (Guyon and Elisseeff, 2003). Current variable selection algorithms for human promoter recognition identifies less than half of the existing transcription start sites accurately and provide more false than true positives, partly due to poorly selected features and computational problems (Zeng et al., 2009). From the above it is clear that, in particular, low signal to noise ratios, errors-in-variables, near collinearity, high dimensionality, and few data points, *i.e.* common properties of current biological data sets, are problematic and no current theory or method can cope with the combination of these in a robust manner.

In summary, both inference of biological networks and variable selection are in general unsolved problems of great importance. A robust solution of both problems, which guarantees desired confidence even for data with the common limitations of current biological data sets, is proposed in Chapter 4 of this thesis, while the theoretical foundation is established in Chapter 5. Conditions for correct inference of pseudo-direct causal interactions, *i.e.* directed regulatory influences that are not mediated by any included node, are established in Chapter 6.

### 1.3 Experiment design

The potential of network inference to reveal interactions among genes based on gene expression data obtained from *in vivo* experiments has been established in numerous studies (see *e.g.* Hecker et al., 2009; Bonneau, 2008; Tegnér and Björkegren, 2007; Cho et al., 2007; Crampin, 2006; Gardner and Faith, 2005). Realization of this potential is, however, hampered by poor quality of available *in vivo* data, *i.e.* lack of information. Current data sets typically suffer from few data points compared to the high number of genes and possible interactions, *i.e.* high dimensionality but few samples, large measurement uncertainty both in the perturbations and responses, *i.e.* low signal to noise ratio and errors-in-variables, and redundant nearly collinear variables, *i.e.* ill-conditioned response matrices in which most of the variation can be explained by a few linear combinations of the variables (Holter et al., 2000; Alter et al., 2000; Tegnér and Björkegren, 2007; Faith et al., 2007; Cosgrove et al., 2010; Nordling and Jacobsen, 2009a; Wu and Wu, 2010). Biological experiments and data collection are in general so expensive that the available budget typically determines both the number of experiments and collected data points, instead of the information requirements for successful inference. It is therefore essential to obtain as much information as possible from the experiments, but perturbation of the genes one-by-one without any optimisation is still standard (see *e.g.* Lorenz et al., 2009; Cantone et al., 2009; Gardner et al., 2003). Improved design of the experiments is clearly needed, but has, with a few notable exceptions that are discussed later on, largely been neglected in the literature on inference of biological networks.

Design of experiments (DoE) is used in statistics to obtain more relevant information in fewer experiments by tailoring the experiments to the purpose of the study (Pronzato, 2008; Chaloner and Verdinelli, 1995; Steinberg and Hunter, 1984; Ljung, 1999; Fisher, 1935). The idea is to plan the experiment such that when

a mathematical model is fitted to or trained on the recorded data it provides a sufficiently good approximation of the properties of interest, while minimising the experimental effort. Model construction can in general be divided into two steps: selection of the model structure, and estimation of parameters (Ljung, 1999). From a modelling perspective network inference is model selection, since the primal concern is which interactions to include in the model, while determination of the strength of the interactions is parameter estimation. It is tradition to differentiate between DoE for parameter estimation, which is used to decrease parameter uncertainty for a chosen model structure, and model/variable selection, which is used to distinguish among a set of alternative model structures, because the two tasks typically require different data (Schwaab et al., 2008). In particular, experiments designed to minimise parameter covariance should not be used for model selection, since measures of parameter uncertainty are only meaningful for the selected model structure. The recent reviews of DoE related to systems biology (Kreutz and Timmer, 2009; Banga and Balsa-Canto, 2008; Franceschini and Macchietto, 2008) provide a good overview of DoE for parameter estimation and, in particular, model based optimal experimental design (OED) for parameter estimation, but only a brief introduction to DoE for model selection. The two introductions of Schwaab et al. (2008) and Michalik et al. (2010) provide an excellent overview of OED for model selection, which next is complemented by other DoE works related to systems biology that have been published within the last decade. These works are analysed below from a network inference perspective and their benefits and disadvantages are highlighted.

Michalik et al. (2010) derived an objective function based on Akaike weights and prior model probabilities for design of optimal experiments that allow discrimination among a whole set of models in one experiment. Skanda and Lebiedz (2010) developed a method for designing optimal initial conditions and perturbations, as well as the number and location of sampling time points, using a Kullback-Leibler objective function to discriminate between two models, and implemented it in a software package. Mélykúti et al. (2010) considered OED and maximized the  $L_2$  distance between the outputs of two rival models by finding the best initial condition, stimulus profile of unit  $L_2$ -norm, or structural changes. Donckels et al. (2009) modified the objective function proposed by Buzzi Ferraris et al. (1984) to include the expected information of the designed experiment and designed optimal pulse inputs for maximizing the weighted difference between outputs. Schwaab and coworkers (Schwaab et al., 2008, 2006) proposed inclusion of model probabilities and the posterior covariance matrix of the difference between model predictions in an objective function, originating from Buzzi Ferraris et al. (1984), which they used for iterative design of optimal experiments to discriminate among a set of alternative models. Kremling et al. (2004) compared three different OED approaches for discriminating between two models on a test case: step inputs for maximizing the difference of the largest outputs, sinusoidal inputs for maximizing the difference between phase shifts of linear approximations of the two models, and input profiles for bringing the weighted states as apart as possible. The third approach is a simplification of the optimal control formulation proposed by Chen and Asprey (2003). Maiwald et al. (2007)

used statistical hypothesis testing to generate a list of required experiments for discrimination between two alternative models by simulating the response to each feasible experiment on one of the models, calculating the chi-square goodness of fit value of the other model, and iteratively removing the experiment with the highest value from the feasible set to the list. All these works are concerned with optimal discrimination of alternative nonlinear dynamical models, *i.e.* hypothesis testing, and require *a priori* knowledge or data both for constructing the alternative models and estimation of their parameters. This implies that the proposed methods are unsuitable for generation of interaction hypotheses, which network inference so far mainly has been used for. In addition, the model and parameter errors may have a large impact on the designed experiments and discriminatory power provided by the obtained data. The current research theme is to account for model and parameter uncertainty through modified objective functions—an approach which further increases the demand on *a priori* information and data.

Ideker et al. (2000) proposed a perturbation “chooser” for maximizing the number of models that can be discriminated based on the fraction of network models reaching the same Boolean state for each possible perturbation, *i.e.* high or low level of expression of any gene. Their approach is limited to Boolean network representations of the system, is computationally expensive if the sets of alternative models and possible perturbations are large, and does not account for noise in the data, but it is suitable for generating and testing interaction hypotheses. Steinke et al. (2007) derived an approximation of the Bayesian posterior distribution over all linear network models based on expectation propagation and used it to calculate the maximum information gain over a set of possible step perturbations. They demonstrated the method by iteratively designing single gene perturbation experiments and recovering all significant interactions of a *Drosophila* segment polarity network *in silico*. Their method was developed for decreasing the uncertainty of inferred interactions in reverse engineering of GRNs in consecutive perturbation experiments, starting from no knowledge about the network except the sparsity prior, so it is suitable both for hypothesis generation and testing. They however assumed noise-free measurements of the steady-state response, while the applied perturbations were assumed to be corrupted by white noise. The former assumption is questionable considering that the noise level of current gene expression data sets is high. Yoo and Cooper (2003) developed a computer system called GEEVE for recommending which microarray experiment to conduct. GEEVE infers a causal Bayesian network and generates a decision tree where each possible experiment is scored. GEEVE is limited to causal Bayesian network representations and pair-wise relationships between genes. It involves heuristic methods and approximations to decrease the computational complexity, which are not described in detail, but generation of interaction hypotheses should be possible. Tegnér et al. (2003) suggested two heuristic rules for selecting which gene to perturb using a genetic toggle switch: perturb without repetition the gene whose activity has changed the least in all previous experiments until all genes achieve an activity threshold, then perturb without repetition the gene with most uncertain connections. They considered the set of linear mappings in which

each gene is influenced by a maximum of  $k_{\max}$  other genes and used two *in silico* examples to demonstrate iterative design of perturbations until only one model was consistent with recorded data. Experiments generated by their rules are suitable for both hypothesis generation and testing, but it may be computationally expensive to determine all models consistent with data, as needed to apply the second rule. Feng and co-workers developed a technique called optimal identification to recover the full distribution of parameters of a nonlinear model consistent with data (Feng et al., 2004; Feng and Rabitz, 2004; Feng et al., 2006). Their technique combines DoE with parameter identification and can, in principle at least, be used for model selection, because the design is based on a set of alternative models consistent with data. However, they employ genetic algorithms for parameter identification and design of perturbations, which is computationally expensive and does not guarantee an optimal solution. They also require initial estimates of the model, steady-state concentrations, and dynamic rate of each rate constant, which makes the method unsuitable for generation of interaction hypotheses. Apgar et al. (2008) used model based controllers to design dynamical stimuli for driving candidate models through a target trajectory so that the models can be distinguished. They selected the target trajectory based on suitability for the available measurement methods and did not explicitly optimize the discriminatory power. Their method is made for discrimination among a few candidate models and it therefore requires *a priori* knowledge or data, implying that it is not suitable for generation of interaction hypotheses. It is important to note that the authors of the last four design approaches (Yoo and Cooper, 2003; Tegnér et al., 2003; Feng et al., 2006; Apgar et al., 2008) have not provided any proof that their design yields sufficiently informative data for distinguishing among all models in the considered sets. This in contrast to the other designs, which per construction should yield informative data or indicate that the models are indistinguishable for all allowed perturbations. Roberts et al. (2009) tried to distinguish between two alternative chemotaxis models for *R. sphaeroides* based on a sinusoidal ligand stimuli with frequency selected based on linearisation of the models fitted to wild type time series data, but simulations proved it infeasible, so in the end they used over-expression of *CheY4* to successfully discriminate between the models.

Considering the typical characteristics of biological data sets—high dimensionality combined with relatively few samples, low SNR, errors-in-variables, and redundant nearly collinear variables (Holter et al., 2000; Alter et al., 2000; Tegnér and Björkegren, 2007; Faith et al., 2007; Cosgrove et al., 2010; Nordling and Jacobsen, 2009a; Wu and Wu, 2010)—none of the proposed methods for DoE is suitable for inference of biological networks. Moreover, it is known in process control that correct identification of the weak gains of the system, corresponding to signal attenuation, is essential for capturing the multivariable properties of the system (Bruwer and MacGregor, 2006; Jacobsen, 1994), but none of the methods targets the weak gain, which is needed to avoid near collinearity in the data.

To summarize, lack of a design of experiments that targets the weak gain currently hampers network inference. A strategy for iterative design of perturbation experi-

ments for network inference based on available data with the typical characteristics of current data sets is proposed in Chapter 7.

## 1.4 Properties of biological systems

Knowledge of generic system properties is important when considering modelling, analysis and synthetic construction of biological networks. A fact clearly visible already in the classical book *General system theory* (von Bertalanffy, 2006) from the 1960s, where von Bertalanffy defines and discusses a number of general system properties. One recently introduced system property is scaling in random networks (Barabási and Albert, 1999), which already has had a profound impact on the way complex networks are viewed, modeled, analysed and constructed (Boccaletti, 2006). This property has for instance helped to explain why human immunodeficiency virus (HIV) is spreading so fast and how to slow down the acquired immunodeficiency syndrome (AIDS) pandemic, since the distribution of the number of different sexual partners in one year decays as a scale-free power law (Liljeros et al., 2001). Another property that has been observed in many different biological systems is robustness; many biological functions have been shown to be robust even to removal of components participating in the generation of the function (Kitano, 2004; Barkai and Leibler, 1997; Little et al., 1999; Alon et al., 1999; Eissing et al., 2005; Kim et al., 2006; Kitano, 2007b). Robustness of biological function has *e.g.* been used to explain why certain drugs have proven to be ineffective or have unexpected side-effects (Kitano, 2007a). Together these two properties illustrate the importance of searching and accounting for system properties.

Several authors have noted that the variation seen in gene expression data often is confined to a subspace of lower dimension than the gene space and that most of the variation can be explained by a relatively small number of variables (Holter et al., 2000; Alter et al., 2000; Kuruvilla et al., 2002; Wu and Dewey, 2006). A few orthogonal directions, in other words, capture almost all of the biological signal, even though hundreds of arrays have been recorded, each measuring the expression change of thousands of genes. In short, the corresponding data matrix is ill-conditioned. The observation that microarray data is ill-conditioned has led others to introduce linear combinations of observed genes, so called “characteristic modes” (Holter et al., 2000, 2001) and “eigengenes” (Alter et al., 2000; Nielsen et al., 2002; Alter and Golub, 2006; Omberg et al., 2007), in an attempt to explain this observation and connect it to properties of the system. They have shown that the dominating modes contain essentially all information and thus capture the essential features of the expression data. The data set can therefore be compressed or reduced into a few simple patterns given by the dominating modes. These modes have further been connected to biological signals such as cell cycle oscillations and release of pheromone  $\alpha$ -factor synchronisation (Alter, 2007). On the other hand, both the characteristic modes and eigengenes are based on a singular value decomposition which is guided by an *a priori* assumption of the latent factors being orthogonal, rather than any

physical property of the system. The resulting modes or latent factors are therefore features of the data set and do not necessarily reflect any biological property of the system; a fact that *e.g.* is pointed out by [Liao et al. \(2003\)](#). Further investigation is needed.

From a systems perspective the characteristic modes only reflect one side of the coin, *i.e.* signals that are amplified by the system. The other side of the coin, *i.e.* signals that are attenuated by the system, is neglected. Since most biological networks should amplify some variations, *e.g.* in signal transduction, while attenuating other variations, *e.g.* in order to maintain homeostasis, it is trivial to realize that both are equally important for the system. One example of the importance of weak modes is seen in bacterial chemotaxis. Perfect adaptation is used in bacterial chemotaxis to reset the response to a change in the nutrition concentration to its initial value, so that the bacterium can maintain its sensitivity to concentration changes and track the nutrition gradient ([Yi et al., 2000](#)). In other words, attenuation of the signal generated by a prolonged change in the nutrition concentration enables strong amplification of small changes. If we only looked at the dominating modes, then we would miss the core mechanism—the integral feedback that enables the perfect adaptation. Attenuation of signals correspond to weak gains and correct identification of them has in process control been shown to be essential for capturing the multivariable properties of the system ([Bruwer and MacGregor, 2006](#); [Jacobsen, 1994](#)). Ill-conditioned data with correlated regressors is known in statistics to cause instability of estimators, large uncertainty in the estimated parameters, sign errors even in statistically significant parameters, and good fit to data despite statistically insignificant parameters (see *e.g.* [Belsley, 1991](#); [Rao and Toutenburg, 1999](#); [Larose, 2005](#)). Ill-conditioning of current data sets is, in other words, hampering network inference and to avoid it, *e.g.* through design of perturbation experiments, one first needs to understand its cause. It is therefore essential to find the underlying system property.

To summarize, an unknown system property needs to be accounted for in experiment design to obtain informative data for network inference. Interampattiness is introduced in [Chapter 3](#), shown to explain previous observations, and postulated as the system property that causes the ill-conditioning in existing gene expression data sets.

## 1.5 Publications

Parts of the work described in this thesis has previously been presented in the following publications and at the following international conferences.

T. E. M. Nordling and E. W. Jacobsen. Interampattiness—a generic property of biochemical networks. *IET Syst Biol*, 3(5): 388–403 (2009a)

In this article a generic system property—interampattiness—is defined. It explains previous observations of the variation in expression data sets being concentrated

in a few “characteristic modes” or “eigengenes”, and highlights their previously neglected counterpart, the “weak modes”. An interampatte network is characterised by strong INTERactions enabling simultaneous AMPLification and ATTEnuation of different signals. It is postulated that bio-networks are interampatte, based on published experimental data and theoretical considerations. Existence of multiple time-scales and feedback loops are shown to increase the degree of interampattiness. Interampattiness is also shown to have strong implications for the dynamics and hamper reverse engineering of the network. Chapter 3 is with the exception of some minor corrections and modifications identical to this article. Parts of these results have also been presented at the following conferences:

T. E. M. Nordling and E. W. Jacobsen. Invalidating models of gene regulatory networks - the implications of characteristic and weak modes for network inference. Engineering Principles in Biological systems conference, Hinxtun (UK) (2009b)

T. E. M. Nordling and E. W. Jacobsen. Inference of interampatte gene regulatory networks - with application to apoptosis signalling. The 9th International Conference on Systems Biology (ICSB-2008), Gothenburg (Sweden) (2008b)

T. E. M. Nordling and E. W. Jacobsen. Ill-conditioning - a property of bio-networks. The 2nd annual q-bio conference on cellular information, Santa Fe (U.S.A.) (2008a)

T. E. M. Nordling and E. W. Jacobsen. On Sparsity As a Criterion in Reconstructing Biochemical Networks. In B. Sergio, A. Cenedese, and S. Zampieri, editors, *Proceedings of the 18th International Federation of Automatic Control (IFAC) World Congress, 2011*, 11672–11678. The International Federation of Automatic Control, Milano, Italy (2011)

This article begins with a brief review of some network reconstruction algorithms based on subset selection and regularization techniques, and a discussion on their suitability for inferring the structure of bio-networks. A particular problem is the fact that these methods provide little or no information on the uncertainty of individual edges, combined with the fact that the identified networks usually have a large fraction of false positives and negatives. To partly overcome these problems, conditions are considered that can be used to classify edges into those that can be uniquely determined based on a given incomplete data set, those that cannot be uniquely determined due to collinearity in the data, and those for which no information is available. This classification can be used to improve the reconstruction by employing standard unbiased identification methods to the identifiable edges while employing sparse approximation methods for the remaining network. An effective rank based method is demonstrated through application to a synthetic network in yeast which has recently been proposed for *in vivo* assessment of network identification methods. The robust variable selection and network inference in Chapter 4 and 5 stem from the classification of noise-free regressors that was introduced in this article. This classification has also been presented at ICSB-2010.

T. E. M. Nordling and E. W. Jacobsen. Sparsity is a means and not an aim in inference of gene regulatory networks. In *The 11th International Conference on Systems Biology (ICSB-2010) in Edinburgh (UK): Abstract book*. Edinburgh, UK (2010b)

T. E. M. Nordling and E. W. Jacobsen. Experiment Design for Proper Excitation of Gene Regulatory Networks. In *Proceedings of Foundations of Systems Biology in Engineering (FOSBE), 2nd Conference*. Fraunhofer IRB Verlag, Postfach 800469, 70504 Stuttgart, Germany (2007)

Ill-conditioned data in general hampers network inference. In this article a systematic iterative experiment design that ensures sufficient excitations of all network interactions is proposed. The design counteracts the intrinsic signal attenuation of the system via correlated perturbations that lift the small singular values of the response matrix. The method leads to combinatorial perturbation experiments, in which a number of genes are perturbed simultaneously. The effectiveness of the method is demonstrated by application to an *in silico* gene regulatory network. The iterative design proposed in Chapter 7 is based on ideas originating from this article and also utilizes correlated perturbations to lift singular values. Parts of this work has also been presented at the following conferences:

T. E. M. Nordling and E. W. Jacobsen. Design of perturbations is the key to inference of tumour specific gene regulation. MGH-KI-Cell press Days of Molecular Medicine, Stockholm (Sweden) (2010a)

T. E. M. Nordling and E. W. Jacobsen. Experiment design for optimal excitation of gene regulatory networks. In *The 7th International Conference on Systems Biology (ICSB-2006) in Yokohama (Japan): Abstract book*. The Systems Biology Institute (SBI) (2006a)

T. E. M. Nordling and E. W. Jacobsen. Experiment design for systematic excitation of gene regulatory networks. International Workshop on Systems Biology in Maynooth (Ireland) (2006b)

E. W. Jacobsen and T. E. M. Nordling. On identification of genetic and metabolic networks. 15th ERNSI Workshop on System Identification in Linköping (Sweden) (2006)

### 1.5.1 Other publications

I have as a part of my work scientific training at KTH Royal Institute of Technology also contributed to the following publications and conference presentations, which are not part of this thesis.

A. Tjärnberg, T. E. M. Nordling, M. Studham, and E. L. L. Sonnhammer. Optimal sparsity criteria for network inference. *Journal of Computational Biology*, 20(5) (2013)

R. Jörnsten, T. Abenius, T. Kling, L. Schmidt, E. Johansson, T. E. M. Nordling, B. Nordlander, C. Sander, P. Gennemark, K. Funa, B. Nilsson, L. Lindahl, and S. Nelander. Network modeling of the transcriptional effects of copy number aberrations in glioblastoma. *Molecular systems biology*, 7(1): 486 (2011)

M. Hellgren, T. E. M. Nordling, E. W. Jacobsen, and J. O. Höög. Multi-level modelling of the parallel metabolism of ethanol and retinol, with implications for foetal alcohol syndrome. In *The 9th International Conference on Systems Biology (ICSB-2008) in Gothenburg (Sweden): Abstract book*. University Of Gothenburg, Curran Associates, Inc., Gothenburg, Sweden (2008)

## 1.6 Outline

I have made the presentation in each chapter independent so that the chapters can be read in any order. The remainder of this thesis contains work done under the supervision of Prof. Elling Jacobsen and is organized in the following way.

We give an introduction to the prerequisites and objectives of inference/reverse engineering of gene regulatory networks in Chapter 2. The idea of the chapter is to introduce familiar knowledge, such as a network of well studied genes in *S. cerevisiae* for molecular biologists and linear ODE models for control engineers, and relate it to the other parts required for modelling of gene regulatory networks. The chapter contains a presentation of an *S. cerevisiae* network, perturbation experiments for its inference, how to measure gene expression changes by quantitative real-time RT-PCR, how GRNs can be described as a system of linear ordinary differential equations, and its relation to the network inference problem. The yeast network is used as an example in the following chapters.

In Chapter 3 we introduce a system property—interampattiness. It explains why the variation in existing gene expression data is concentrated to a few “characteristic modes” or “eigengenes”, and why previously inferred models have a large number of false positive and negative links. An interampatte system is characterized by strong INTERactions enabling simultaneous AMPLification and ATTENUation of different signals. We show that perturbation of individual state variables typically yield ill-conditioned data with both characteristic and weak modes that hamper network inference.

The network inference problem is a variable/model selection problem. In Section 4, we propose a solution to it by considering all possible models of a specified class that can explain the data at a desired significance level and classifying only the links present in all of these models as existing. This is done in order to get a robust

solution and assign confidence to individual interactions. We find these links without estimating a single parameter or fitting any model to the data by reformulating the variable selection problem as a rank problem, which we solve by using the structured singular value. The proposed method is demonstrated by reverse engineering of a synthetic GRN in *S. cerevisiae* which has been proposed for *in vivo* assessment of network identification methods. We also show that when the number of variables exceeds the number of observations alternative models with different structure that explain noisy data always exist. Consequently no reliable measure of confidence of individual interactions exists.

In Chapter 5 we establish a theory for robust variable selection and network inference based on introduction of uncertainty sets and extension of standard concepts from linear algebra. It enables us to state both sufficient and necessary data conditions both for when a variable must be selected to explain data, and for when a variable always can be excluded while explaining data. The former implies that the corresponding interaction exists and the latter that the corresponding interaction is non-existing. In layman terms all variables and interactions are classified as present/existing, absent/non-existing, alternative, or non-evidential, based on the information available in the data. We also prove that the rank problem and solution by the structured singular value are correct and that the selected interactions are true positives under mild assumptions.

We analyse how unperturbed and unobserved latent state variables that exist in the real system affect the inferred subnetwork in Chapter 6. The existence of latent states implies that a subnetwork of pseudo-direct causal influences, accounting for all environmental effects, is in general inferred. We show that it is necessary to know a linearly independent subset of as many relations between the states of interest and responses or perturbations as there are states of interest. In steady-state experiments it is necessary to both observe and perturb each state of interest independently, which implies that at least as many experiments as variables of interest are needed. We also show that bounds on the number of latent states as well as the number of different paths between the states of interest, in principle, can be determined from time-series data. However, the physical quantities corresponding to the latent states cannot be determined from perturbation experiments alone, unless they are directly perturbed or observed.

Iterative design of correlated multi-gene perturbation experiments that counteract the intrinsic signal attenuation of the system is shown in Chapter 7 to be necessary to obtain informative data for network inference. We also propose and numerically demonstrate a practical design of steady-state experiments that requires no *a priori* information, but explores the gains of the system and provides informative data about the Snf1 signalling pathway in *S. cerevisiae*.

In Chapter 8 all results are summarized and connected to give a picture of our contribution to inference of GRNs. We also make recommendations for future work.

Additional material that is necessary to prove or reproduce our results is included in Appendices A-C. References are given throughout this thesis in order of importance, instead of by increasing or decreasing year of publication.

# From data to linear models of gene regulatory networks

---

*“I think that it is a relatively good approximation to truth  
–which is much too complicated to allow anything but  
approximations–that mathematical ideas originate in empirics”*

John von Neumann, The Mathematician,  
The works of the mind, 1947.

To provide a context for our scientific contribution in Chapter 3-7 and to explain its role in modelling of gene regulatory networks (GRNs), we here give an introduction to the prerequisites and objectives of inference or reverse engineering of GRNs. We first present a network of well studied genes in yeast together with perturbation experiments for its inference in Section 2.1. Then we describe how to measure gene expression changes using quantitative real-time RT-PCR in Section 2.2. Finally, we show how GRNs can be approximated as a system of linear ordinary differential equations (ODEs) and relate this model formalism to the basic network inference problem in Section 2.3. The idea is to include knowledge familiar to the reader, such as the network of well studied genes in yeast for molecular biologists and linear ODE models for control engineers, and to relate it to the other parts required for modelling of gene regulatory networks. To construct a dynamic model of a specific GRN based on quantitative data one, in general, needs to perform the following tasks:

1. Perform perturbation experiments in which the expression of the genes are altered.
2. Collect quantitative data by at least measuring the change in gene expression resulting from each perturbation.
3. Specify a data model, including an error model, that connects the data to the interaction matrix of the GRN.

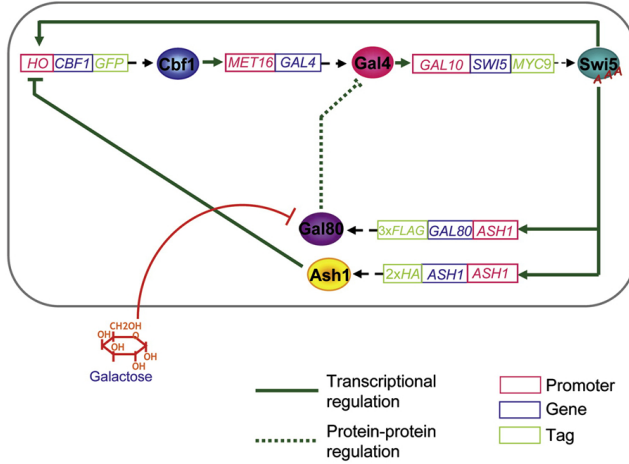
4. Infer the structure of the interaction matrix from the collected data.
5. Estimate the nonzero parameters of the model.

To set the stage for our contribution we address in this chapter the prerequisites and objectives of network inference (Tasks 1-3). How to infer the structure of the interaction matrix (Task 4), *i.e.* the causal interactions that exist among the genes in the network, is the main topic of this thesis and it is addressed in Chapter 4-6. Once the structure of the network is determined, then existing statistically and computationally efficient parameter estimators can be used to estimate the nonzero parameters of the model structure employed here (Task 5). We *e.g.* later in Appendix A-C estimate the strength of inferred interactions when needed by using convex optimization to minimize the weighted residual sum of squares subject to structural constraints.

## 2.1 A gene regulatory network—the IRMA example

As an example of a gene regulatory network and perturbation experiments for its inference, we present here the synthetic GRN engineered in *S. cerevisiae* by Cantone *et al.* (2009) for the purpose of *in vivo* reverse-engineering and modelling assessment (IRMA). This network was created specifically for evaluation of network inference algorithms. It is also small enough to be comprehended by the human mind, and is per design negligibly affected by endogenous genes. Cantone *et al.* performed *in vivo* perturbation experiments for inference of this network and made their steady-state and time-series data publicly available (Cantone *et al.*, 2009). Taken together these facts make it suitable for our illustrations. We next briefly explain how Cantone and co-workers engineered the synthetic network, what the engineered network looks like, how they performed the steady-state perturbation experiments, and measured the resulting expression changes.

The IRMA network consists of the following five genes: *CBF1*, *GAL4*, *SWI5*, *ASH1*, and *GAL80*. Cantone *et al.* selected these genes because they are well characterized and can be knocked out viably in wild type yeast. They used the *swi5<sub>AAA</sub>* mutant of Swi5 with the three phosphorylated Serine residues (Ser-522, Ser-646, and Ser-664) substituted by Alanines to avoid cell cycle control mediated by Cdk8 phosphorylation of Swi5. To obtain the desired structure of their synthetic network, Cantone *et al.* placed different promoters upstream of the genes in vectors containing different markers, which were inserted in the genome of *gal4 $\Delta$ 542 gal80 $\Delta$ 538* yeast strain YM4271, whose *GAL4* and *GAL80* loci were deleted. The *SHE2*, *ACE2*, and endogenous counterparts of the selected genes were also deleted. They cloned a green fluorescence protein (*GFP*) tag at the 3' end of the *CBF1* open reading frame, in order to in the microscope visually verify the operation of the network. This resulted in the following transcriptional units: *HO* promoter/*CBF1-GFP*, *MET16* promoter/*GAL4*, *GAL1-10* promoter/*SWI5-MYC9*, *ASH1* promoter/*ASH1-2XHA*, and *ASH1* promoter/*GAL80-3XFLAG*. For further details, see their original article with supplemental. The



**Figure 2.1: The engineered IRMA network.** The structure of the synthetic gene regulatory network created by the designed transcriptional units (rectangles), each consisting of a promoter (red) and gene (blue) tagged with the specified sequences (green). An arrow shaped head on the link indicates up-regulation or activation, while a bar shaped head indicates down-regulation or repression. Reprinted from Cantone *et al.* (2009) with permission from Elsevier.

structure of the the IRMA network together with the constructed transcriptional units is shown in Figure 2.1. The interaction matrix, representing the designed regulation between the genes of the engineered IRMA network, is

$$\tilde{\mathbf{A}} = \begin{bmatrix} -1 & 0 & 1 & -1 & 0 \\ 1 & -1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & -1 & 1 & 0 & -1 \end{bmatrix}. \quad (2.1.1)$$

More precisely, this is the signed adjacency matrix in which the unknown strength of each interaction is replaced by one, which we discuss in depth in Section 2.3.2. It is an approximation of the system in which only the mRNA abundance of each gene is explicitly included as state variables and we will in Section 2.3.1 make the connection to a system of linear ODEs. This network per design contains both positive and negative feedback loops and one known protein-protein interaction between *GAL4* and *GAL80*.

Transcription of *GAL4* depends on the growth medium of the cells. In absence of Galactose the Gal80 repressor prevents transcription which takes place in Galactose. The IRMA system therefore has two different dynamical modes, termed off and on by Cantone *et al.*, and the expression of the genes in the network depends on

whether the cells are grown in a medium containing Glucose (off) or a mixture of Galactose and Raffinose (on). We therefore refer to the mode and data recorded with the network in a particular mode as Glucose or Galactose for simplicity. Cantone *et al.* perturbed the system by over-expressing each of the five genes one at a time in separate experiments by inserting a plasmid containing a copy of the gene under control of the strong constitutive *GPD* promoter. This is in other words an example of the use of exogenous perturbations that alter the copy number of a gene. They grew the transformed cells at 30°C in synthetic complete medium lacking Histidine with 2% Glucose or 2% Galactose and 2% Raffinose. They then measured the change in expression of each gene when the network reached steady-state, *i.e.* after a sufficiently long time so that the expression of each gene had stabilized, relative to unperturbed control experiments, *i.e.* transformation by an empty vector. We later use their steady-state perturbation data to demonstrate our method for robust network inference in Section 4.6 and Section A.6. The gene perturbed in each experiment is known and we assume that the strength of the perturbation is equal to one in each experiment, matching the approach taken by Cantone *et al.* when they inferred the network using the NIR algorithm. This however implies that the interaction strengths that we estimate in Sections A.6 and B.1 are relative to the actual strength of the perturbations.

To measure the change in mRNA abundance of the five genes in the perturbation experiments, Cantone *et al.* used quantitative real-time reverse transcription polymerase chain reaction (qRT-PCR), which we discuss in more detail in Section 2.2. They used Platinum SYBR Green qPCR SuperMix-UDG with ROX (Invitrogen) for the qRT-PCR reactions. The cycle threshold ( $C_T$ ) values, *i.e.* the fractional number of cycles at which the fluorescence intensity threshold is reached, were obtained from the Applied Biosystems SDS software version 1.2.3. The average  $C_T$  values and standard errors reported by Cantone *et al.* for each measured transcript, including the housekeeping gene *ACT1*, are based on two technical replicates of each reverse transcription polymerase chain reaction (RT-PCR) reaction (Cantone *et al.*, 2009, supplemental). They then used the comparative  $C_T$  method to obtain the relative expression change of each gene; see Cikos and Koppel (2009) for a review of how fluorescence readings from RT-PCR are converted into gene expression and Schmittgen and Livak (2008) for a current comparative  $C_T$  protocol. More precisely, they use the  $2^{-\Delta\Delta C_T}$  method, since the amplification efficiency is assumed to be optimal, *i.e.* a doubling of the complementary DNA (cDNA) in each cycle. Their reported  $\Delta C_T$  values are the difference between the average  $C_T$  of the measured gene of interest and the average  $C_T$  of the housekeeping gene *ACT1* for all clones. And their  $\Delta\Delta C_T$ , which we later denote by  $y_i$  in (2.2.4), is the logarithm of the fold change of the expression of the gene of interest in the perturbed sample relative to the unperturbed control sample.

## 2.2 Gene expression data—real-time RT-PCR for mRNA quantification

The IRMA data used in this work has been generated by quantitative real-time RT-PCR, which has certain characteristics that one needs to know when analyzing or working with such data. We therefore here give a short introduction to quantitative real-time RT-PCR data. When using the data for network inference, a data model, including an error model, must be assumed and we therefore focus on how to model the data generating process. In general, a data model describes our assumptions about how the measurements are obtained, *i.e.* the process generating the data. It also specifies how the observed quantity is related to quantities of interest and the errors or noise, which is assumed to have certain properties specified in an error model. The importance of using an appropriate data and error model was recently demonstrated for immunoblotting (Kreutz et al., 2007), where the signal to noise ratio (SNR) was improved by more than a factor 10 relative to raw background subtracted intensities.

Invention of the polymerase chain reaction (PCR) in 1983, for which Kary Mullis was awarded the Nobel prize in Chemistry 1993, enabled cyclic amplification of essentially any sequence of nucleic acids (Saiki et al., 1985; Bartlett and Stirling, 2003). Real-time PCR is a refinement of the original PCR, first published in 1992 (Higuchi et al., 1992), in which the amount of product formed in each amplification cycle is monitored in order to enable quantification of the initial amount of a sequence of nucleic acids of interest. In qRT-PCR, real-time monitoring of the PCR is combined with initial reverse transcription (RT) of the mRNA, *i.e.* copying of the messenger RNA in the sample to complementary DNA, such that the abundance of an mRNA sequence of interest can be quantified, either relative to an internal standard or in absolute terms. See Kubista et al. (2006) for a general overview of the RT-PCR technique and Schmittgen and Livak (2008) for a current qRT-PCR protocol using the comparative  $C_T$  method. The popularity of real-time RT-PCR for mRNA quantification has increased over the past decade so that in 2008 it was used in approximately 88% of the published studies that quantify mRNA, as opposed to a mere 8% in 1999 (Thellin et al., 2009). To measure the mRNA abundance using qRT-PCR, the following steps are required to get from a sample to data: (i) extraction of RNA from the biological sample, (ii) reverse transcription of the mRNA to generate cDNA, (iii) measurement of fluorescence intensity in each amplification cycle of the cDNA using real-time PCR, (iv) conversion of measured intensities to relative or absolute mRNA quantities. We next focus on the fourth step, because it determines the data model, while merely commenting on errors introduced by the first three steps.

The RT-PCR is in general considered to consist of four phases: a lag phase with exponential amplification but no detectable fluorescence signal, a log phase with exponential amplification and detectable changes in fluorescence between each cycle, a retardation phase when the reaction decelerates, and a stationary phase when no

further amplification takes place (Scheffe et al., 2006). The detected fluorescence intensity after each amplification cycle of the RT-PCR reaction during exponential amplification in the log phase is commonly described by

$$F_k = F_0(1 + E)^k, \quad (2.2.1)$$

with  $F_k$  and  $F_0$  being the fluorescence intensity. The intensity is assumed to be proportional to the amount of DNA amplicons, after  $k$  and 0 cycles, respectively (Scheffe et al., 2006; Cikos and Koppel, 2009). The amount of DNA amplicons at 0 cycles is assumed to be proportional to the abundance of mRNA of the gene of interest in the sample. The amplification efficiency of the reaction is in this model assumed to be constant and described by  $E$ . Alternatively the detected fluorescence intensity can be described by a sigmoidal equation (Rutledge and Stewart, 2008)

$$F_k = \frac{F_{\max}}{1 + \left(\frac{F_{\max}}{F_0} - 1\right) (1 + E_{\max})^k}, \quad (2.2.2)$$

with

$$E_k = \Delta E F_k + E_{\max}, \quad F_{\max} = \frac{E_{\max}}{-\Delta E}. \quad (2.2.3)$$

Here  $F_{\max}$  is the maximal fluorescence intensity,  $E_{\max}$  is the maximal amplification efficiency, *i.e.* the efficiency at the beginning of the first cycle,  $E_k$  is the amplification efficiency in cycle  $k$ , and  $\Delta E$  is the rate of loss in amplification efficiency. The retardation phase of the PCR begins immediately, since the efficiency in cycle  $k$  is assumed to decrease linearly with fluorescence intensity. The sigmoidal equation avoids the issue of determining which cycle points that belong to the log phase, and has recently been shown to improve the accuracy, in particular when determining absolute quantities (Rutledge and Stewart, 2008, 2010). Older comparisons of methods have, on the other hand, shown better accuracy and reproducibility of those based on the exponential model than those on the sigmoidal model (Karlen et al., 2007; Cikos et al., 2007). We however note that the sigmoidal model used in Karlen et al. (2007) and Cikos et al. (2007) differs from that used in Rutledge and Stewart (2008, 2010). In general, no consensus exists in the literature on which method to use for converting fluorescence readings to relative or absolute quantification of gene expression, and essentially all published methods have been shown to have advantages and weaknesses, see *e.g.* Scheffe et al. (2006); Karlen et al. (2007); Cikos et al. (2007); Rutledge and Stewart (2008); Cikos and Koppel (2009). The most common method for relative quantification is the comparative  $C_T$  method (Cikos and Koppel, 2009; Schmittgen and Livak, 2008), which is a benchmark based method according to the classification used by Cikos and Koppel (2009) into (i) benchmark based, (ii) regression based, and (iii) combined methods. It has been shown to yield reproducible albeit fairly inaccurate results (Karlen et al., 2007) and it was used by Cantone et al. (2009) to produce the IRMA data set that we use in this work. More precisely, they used the  $2^{-\Delta\Delta C_T}$  method to obtain the

logarithm of the fold change in expression of the gene of interest

$$\begin{aligned}
 y_{it} &\triangleq \log_2 \left( \frac{x_{it}}{x_{iC}} \right) = -\Delta\Delta C_T \\
 &= \underbrace{C_{iC} - C_{rC}}_{\triangleq \Delta C_{TiC}} - \underbrace{(C_{it} - C_{rt})}_{\triangleq \Delta C_{Tii}}.
 \end{aligned} \tag{2.2.4}$$

Here  $x_{it}$  denotes the abundance of mRNA of gene  $i$  in the experiment marked by  $t$ , with  $C_{it}$  being the mean  $C_T$  value for gene  $i$  in the clones of the experiment in which gene  $t$  is perturbed and  $C_{iC}$  in the unperturbed control sample, and  $C_{rt}$  being the mean  $C_T$  value for the housekeeping gene in the clones of the experiment in which gene  $t$  is perturbed and  $C_{rC}$  in the unperturbed control sample. The abundance of mRNA of gene  $i$  is denoted by  $x_{it}$  for the experiment in which gene  $t$  is perturbed and  $x_{iC}$  in the unperturbed control experiment. Cantone *et al.* assumed the amplification efficiency to be optimal, *i.e.* a doubling of the cDNA in each cycle, and used the housekeeping gene *ACT1* for normalization. We do not have access to the raw fluorescence readings, so we must assume the exponential model that was used by the authors recording the data.

The next question is which error model to assume, *i.e.* how is the error in the estimated expression changes distributed? We have only found two studies focusing specifically on this issue: A theoretical study where the branching process theory was used to compute the probability distribution of the offspring of a single molecule during the PCR (Peccoud and Jacob, 1996). An empirical study where a phenomenological model, distinguishing between RT, PCR, and dilution noise, was constructed based on 24 qRT-PCR replicates of diluting total RNA 4 to 6 times, and 30 replicates of diluting cDNA 5 times and measuring the *Gfap*, *Rps29*, *ChgB*, *Nes*, *Sox2*, and *Ins2* genes in mouse endocrine cells (Bengtsson *et al.*, 2008). Both studies addressed absolute quantification by RT-PCR based on the exponential model and found that the error is largest for genes with few mRNA copies, *i.e.* weakly expressed genes, and then decreases rapidly so that the relative error is around 10-30% at 100 copies depending on the gene and amplification efficiency. The empirical study reports that the error is log-normally distributed, *i.e.* the logarithm of the expression is normally distributed, based on quantile-quantile plots of the  $C_T$  values. These plots in our opinion, however, contain some fairly large outliers, calling for more precise testing of the hypothesis that the errors are normally distributed. The logarithm of the expression is assumed to be normally distributed also in Steibel *et al.* (2009). Steibel *et al.* list a few other works using this assumption, but no investigation of it is performed. The theoretical study by Peccoud and Jacob (1996), on the other hand, reports that the shape of the distribution depends on the amplification efficiency and initial copy number. It has several maxima for high efficiencies and only tends towards a normal distribution as the initial copy number goes to infinity. We tested the null hypothesis that the data comes from a normally distributed population using a Lilliefors test (Lilliefors, 1967) on the fold changes of the IRMA data generated by Cantone *et al.* (2009). The null hypothesis was

rejected with p-values smaller than 0.001 for both the *S. cerevisiae* cells cultured in Glucose and Galactose. For copy numbers  $< 10$ , [Rutledge and Stewart \(2010\)](#) show that the errors can be explained by a Poisson distribution and present a dilution based strategy for accurate quantification of weakly expressed genes, which in our opinion should be used instead of quantification based on the exponential model in order to avoid the poor SNR for low copy numbers. In the empirical study the RT efficiency was found to vary from 99% to below 30% depending on the gene and the RT error to dominate over the PCR error for copy numbers  $> 1000$ , which is in agreement with an earlier study where the RT was found to contribute most to the variation ([Kubista et al., 2006](#)). For copy numbers  $> 1000$ , a single measurement was shown to be accurate to within a factor of two and the total RT, PCR, and dilution errors were negligible compared to the biological variation in the empirical study. In addition to noise stemming from the RT, PCR, and sample dilution, qRT-PCR data suffers from many other error sources, such as pipetting errors, contamination by DNA, varying RNA quality, varying specificity of primers, bad normalization, and varying amplification efficiency. These are in general difficult to quantify and depend on the generating model that is assumed, but the effect of the latter two has been investigated in several studies and we briefly summarize the findings. The use of a single housekeeping gene for normalization may introduce large errors if its expression varies among the samples ([Huggett et al., 2005](#); [Kubista et al., 2006](#); [Vandesompele et al., 2002](#)), but no method for quantifying the actual error has been presented. Use of several housekeeping genes that are validated to be constantly expressed in all experimental conditions of the study is currently the recommended internal standard for quantitative RT-PCR ([Thellin et al., 2009](#)). Investigations of the amplification efficiency  $E$  of the RT-PCR reaction has shown that it is typically between 0.65 and 0.95, but values down to 0.4 and up to 1.15 have been reported ([Rutledge and Stewart, 2008](#); [Karlen et al., 2007](#); [Kubista et al., 2006](#); [Tichopad et al., 2003](#)). It is fairly well established that the efficiency is normally distributed ([Ruijter et al., 2009](#); [Karlen et al., 2007](#)). The efficiency depends on the primer sequence, transcript, sample, and model ([Rutledge and Stewart, 2008](#); [Karlen et al., 2007](#); [Kubista et al., 2006](#); [Suslov and Steindler, 2005](#)), and use of the wrong efficiency leads to large errors in the measurements of relative change of gene expression ([Scheffe et al., 2006](#)). Also the baseline or background fluorescence estimates has been shown to strongly influence the estimated amplification efficiency and thereby the measurements ([Ruijter et al., 2009](#)). In conclusion, we can say for absolute quantification based on the exponential model that the SNR increases with the initial number of mRNA copies and that the PCR error dominates at low copy numbers, but the actual distribution is largely unknown and more research is needed. Based on this the uncertainty of relative quantifications should, in general, also decrease with increasing copy numbers, but it has not been studied systematically. In summary, the potential error sources are many and hard to quantify. We therefore, later, in [Section A.4](#), develop a deterministic error model, based on the largest deviation from the mean for each gene in the IRMA data set, and use this for robust inference of the IRMA network.

## 2.3 Linear models of gene regulatory networks

The primary objective in inference of GRNs is determination of the interactions that exist among a set of genes. The genes are commonly seen as nodes and the interactions as links in a network graph that often is represented as a matrix, which we call the interaction matrix. We here derive the linear data model, used in this thesis to connect the perturbations and responses of the genes to the interaction matrix of interest, based on the assumption that the GRN is a dynamical system that can be described by a system of linear ODEs. This assumption has previously been used when inferring GRNs from quantitative data in *e.g.* Yuan et al. (2011); Nadadoor et al. (2011); Wildenhain and Crampin (2006); Bansal et al. (2006); Gardner et al. (2003); Yeung et al. (2002), just to mention a few. Note, however, that this is only one out of many model formalisms that have been used in network inference and several other formalisms also lead to a linear data model. For an overview of different formalisms and their use in network inference see *e.g.* Hecker et al. (2009); Karlebach and Shamir (2008); de Jong (2002).

### 2.3.1 The linear ODE model of gene regulatory networks

Dynamical models are needed to describe the changes that occur over time in living organisms. We will therefore over the next sections derive a linear data model, which we later in Section 4.6 use for robust inference of the IRMA network, starting from a dynamical model of the system.

Like many others, we assume that the system responsible for the gene regulation can be described by a system of linear ordinary differential equations

$$\frac{d\check{\mathbf{x}}}{dt}(t) = \check{\mathbf{A}}\check{\mathbf{x}}(t) + \check{\mathbf{B}}(\mathbf{p}(t) - \mathbf{f}(t)) \quad (2.3.1a)$$

$$\mathbf{y}(t) = \check{\mathbf{C}}\check{\mathbf{x}}(t) + \mathbf{e}(t). \quad (2.3.1b)$$

This assumption should at least hold for sufficiently small changes in the state variables near the systems native steady-state in a dynamical mode, such as the Glucose or Galactose mode of the IRMA strain. The state vector  $\mathbf{x}(t) \triangleq [x_1(t), \dots, x_n(t)]^T$  is commonly restricted to only include mRNA abundances of the considered genes, and we therefore sometimes use the term gene space to refer to what in dynamic systems theory is known as the state space (Wolkenhauer et al., 2005b). This type of model is in systems theory commonly known as a linear state-space model. The designed input, or perturbation vector,  $\mathbf{p}(t) \triangleq [p_1(t), \dots, p_l(t)]^T$  contains all external factors used to perturb the system by changing the experimental conditions, and is possibly corrupted by unknown perturbations or process errors represented by a random vector  $\mathbf{f}(t) \triangleq [f_1(t), \dots, f_l(t)]^T$ . The observed output or response vector  $\mathbf{y}(t) \triangleq [y_1(t), \dots, y_o(t)]^T$  contains the measurement of the dependent variables, corrupted by random measurement errors  $\mathbf{e}(t) \triangleq [e_1(t), \dots, e_o(t)]^T$ . Mechanistic insight is commonly desired and the primary objective in network inference is therefore to determine which of the interactions among the states, represented in the interaction

matrix  $\check{\mathbf{A}}$ , that exist in the real GRN based on observations of the perturbations  $\mathbf{p}(t)$  and corresponding responses  $\mathbf{y}(t)$ . Here the accent  $\check{\phantom{x}}$  is used to distinguish mean variables or deterministic “true” variables that are free from measurement errors from observed variables, which always contain errors. The  $\check{\mathbf{B}}$  matrix contains the relation between the perturbation and state derivatives. Similarly, the  $\check{\mathbf{C}}$  matrix contains the relation between the state and response vector. This model formalism can be motivated either based on a Taylor approximation of a system of nonlinear ODEs or based on taking the logarithm of an S-system and it is for sufficiently small deviations an acceptable approximation for a biological system in a particular dynamical mode/physiological state (Crampin et al., 2004; de Jong, 2002; Voit, 2000; Brazhnik, 2005; Nordling et al., 2007b). For more information on the use of ODEs in modelling of biological systems see *e.g.* Wolkenhauer et al. (2005b). Note that the real biological system contains many additional state variables, *e.g.* proteins, metabolites, and microRNAs, which at the abstraction level represented by this model are excluded. Their influence is however implicitly included when the model parameters are estimated based on data from perturbation experiments, as we later show in Section 6.6, and the interactions are therefore in general so called regulatory influences (Gardner and Faith, 2005).

If we directly measure the mRNA level of all genes in the network, then the matrix  $\check{\mathbf{C}}$  is diagonal. Similarly, if we directly perturb the rate of transcription of all genes independently, then also the matrix  $\check{\mathbf{B}}$  is diagonal. In this case we can without restrictions scale our perturbations and responses such that  $\check{\mathbf{B}}$  and  $\check{\mathbf{C}}$  are identity matrices  $\mathbf{I}$  and  $o = l = n$ , which enables us to simplify the system model (2.3.1). This scaling leaves the state variables and sought interaction matrix  $\check{\mathbf{A}}$  unaffected.

We next define the objective of network reconstruction and discuss the interaction matrix in general. We then show that the linear ODE model (2.3.1) reduces to a linear mapping in the case of steady-state experiments, which motivates the linear data model that we later assume in our work on robust variable selection and network inference in Chapters 4 and 5.

### 2.3.2 The interaction matrix—the objective of network reconstruction

The primary objective of network inference/reconstruction is to determine the interactions/influences that exist in the network and the secondary objective is to determine the sign of the existing interactions. Estimation of the strength of the interactions, on the other hand, is typically not considered to be part of the network reconstruction, even though some algorithms used for reverse engineering, such as LASSO (Tibshirani, 1996; Bonneau et al., 2006), also estimate these strengths. The goal is, in other words to, based on recorded perturbations and response data, infer the signed structure of the interaction matrix  $\mathbf{A}$ , corresponding to the signed adjacency matrix of the network represented as a directed graph, with each node corresponding to a state variable. We henceforth use the term interaction matrix when the sign and strength of all interactions are of interest, *i.e.*  $a_{ij} \in \mathbb{R}$ , signed

adjacency matrix when the sign of the interactions are of interest,  $a_{ij} \in \{-1, 0, 1\}$ , and adjacency matrix when only the existence of the interactions are of interest, *i.e.*  $a_{ij} \in \{0, 1\}$ , in order to be stringent. We however denote all three by  $\mathbf{A}$  and call it the interaction matrix, since the latter two are simplifications of the interaction matrix and the precise meaning typically is clear from the context. The term adjacency matrix is used to denote the structure, since each state variable, or in general each component of the regulatory system, can be seen as a node or vertex of a directed graph in which the regulatory influences form links or edges (Gardner and Faith, 2005) and the structure or topology of the network per definition is given by the adjacency matrix of the graph (Gross and Yellen, 1999). If element  $a_{ij}$  is zero then no direct causal influence exists from node/state  $j$  to  $i$  among the states included in the network. If  $a_{ij} < 0$  then an increase in state variable  $j$  leads to a decrease in state variable  $i$ , *i.e.* gene  $j$  down-regulates or represses/inhibits gene  $i$ . If  $a_{ij} > 0$  then an increase in state variable  $j$  leads to a increase in state variable  $i$ , *i.e.* gene  $j$  up-regulates or activates gene  $i$ . The magnitude of an element,  $|a_{ij}|$ , is the strength of the influence. Here an interaction is direct if it is not mediated by any other node/state that is explicitly included in the network, which means that it still may be mediated by some latent states of the system that are not explicitly included in the network. Note that inferred interactions only are direct physical relations if the two molecules represented by the states bind to each other, implying that inferred interactions are not in general direct physical relations. Instead they are so called regulatory influences that in general are mediated through non-modelled states, as discussed in Gardner and Faith (2005).

### 2.3.3 Steady-state perturbation experiments

Let us now consider the use of steady-state data, *i.e.* data recorded after applying a linear combination of constant step perturbations and measuring the response of the system when it has reached a new steady-state. Note that the following results can easily be extended to the case with time-series data, *e.g.* if one considers a discrete time model as demonstrated *e.g.* in Schmidt et al. (2005), which under certain conditions allow relaxation of the assumption that we directly perturb and measure each gene independently. At steady-state the rate of change is zero  $d\tilde{\mathbf{x}}/dt(t) = 0$ , which together with the scaling of the perturbations and responses, enable us to express the system model (2.3.1) as

$$\tilde{\mathbf{A}}\tilde{\mathbf{x}}_r = -\mathbf{I}(\mathbf{p}_r - \mathbf{f}_r) \quad (2.3.2a)$$

$$\mathbf{y}_r = \mathbf{I}\tilde{\mathbf{x}}_r + \mathbf{e}_r. \quad (2.3.2b)$$

Here we have replaced the time dependence marker by the subindex  $r$  to mark that the state variables represent the change between the new and original steady-state in experiment  $r$ , since it is the only accessible quantity of interest in steady-state measurements, and the error variables denote the aggregated error in experiment  $r$ . Let  $\mathbf{Y} \triangleq [\mathbf{y}_1, \dots, \mathbf{y}_m] \in \mathbb{R}^{n \times m}$  denote the matrix of all measured responses to the  $m$

combinations of perturbations stored in the column vectors of  $\mathbf{P} \triangleq [\mathbf{p}_1, \dots, \mathbf{p}_m] \in \mathbb{R}^{n \times m}$  and similarly  $\mathbf{E}, \mathbf{F} \in \mathbb{R}^{n \times m}$  be the unknown noise realizations, then we obtain the data model

$$\mathbf{Y} = -\check{\mathbf{A}}^{-1} \mathbf{P} + \check{\mathbf{A}}^{-1} \mathbf{F} + \mathbf{E} = \check{\mathbf{G}} \mathbf{P} - \check{\mathbf{G}} \mathbf{F} + \mathbf{E}. \quad (2.3.3)$$

Here  $\check{\mathbf{G}} = -\check{\mathbf{A}}^{-1}$  is the static gain matrix, *i.e.* the matrix of steady-state gains, corresponding to the inverse of the interaction matrix. Near a particular stable physiological state the GRN is hence assumed to act as a linear mapping from the space of all possible perturbations to the corresponding responses of the measured state variables,  $\check{\mathbf{G}}: \mathbb{R}^n \mapsto \mathbb{R}^n$ .

The static gain matrix can be said to reflect the cumulative long-term, low frequency influences, while the interaction matrix contains the direct causal interactions, *i.e.* short-term, high frequency influences. If a directed path exists from node  $j$  to  $i$ , then typically  $g_{ij} \neq 0$ , but the influence mediated by this path may in principle be cancelled by some other path implying that  $g_{ij} = 0$  (Gilbert, 1994; Huckle, 1999). They emphasize different sides of the same coin, because small terms in  $\check{\mathbf{A}}$  are large in  $\check{\mathbf{G}}$  and *vice versa*. This is seen when the singular value decomposition (SVD) (Horn and Johnson, 1990, p. 414-415) is used to express  $\check{\mathbf{A}}$  as a sum of rank one matrices  $\mathbf{u}_k \mathbf{v}_k^T$ , with norm one, in a dyadic expansion

$$\check{\mathbf{A}} = \sum_{k=1}^{\min(m,n)} \sigma_k \mathbf{u}_k \mathbf{v}_k^T \quad (2.3.4)$$

and  $\check{\mathbf{G}}$  is expressed as a sum of these

$$\check{\mathbf{G}} = - \sum_{k=1}^{\min(m,n)} \frac{1}{\sigma_k} \mathbf{v}_k \mathbf{u}_k^T. \quad (2.3.5)$$

A term with a small singular value  $\sigma_k$  in  $\check{\mathbf{A}}$  is large in  $\check{\mathbf{G}}$ , due to the reciprocal  $1/\sigma_k$  and *vice versa*. Even though one is the matrix inverse of the other, their estimates are affected differently by noise, in particular if  $\mathbf{Y}$  or  $\mathbf{P}$  is ill-conditioned, *i.e.* have singular values of different orders of magnitude. This is seen *e.g.* by considering the ordinary least-squares estimates

$$\hat{\mathbf{A}} = -\mathbf{P}\mathbf{Y}^\dagger \quad \text{and} \quad (2.3.6)$$

$$\hat{\mathbf{G}} = \mathbf{Y}\mathbf{P}^\dagger, \quad (2.3.7)$$

expressed using the Moore-Penrose generalized inverse of the matrix (Friedberg et al., 2003, p. 398-400), which based on the SVD of  $\mathbf{Y}$  is defined as

$$\mathbf{Y}^\dagger \triangleq \sum_{k=1}^{\min(m,n)} \frac{1}{\sigma_k} \mathbf{v}_k \mathbf{u}_k^T. \quad (2.3.8)$$

The smallest singular value of  $\mathbf{Y}$  will have a large impact on the estimate of the interaction matrix, since the reciprocal of the smallest singular values has the largest impact on the generalized inverse  $\mathbf{Y}^\dagger$ , but minor impact on the estimate of the static gain matrix. The opposite holds for the smallest singular value of  $\mathbf{P}$ . The effect of errors is typically largest on the smallest singular value, so the output error  $\mathbf{E}$  has in general a larger impact on  $\hat{\mathbf{A}}$ , while the input error  $\mathbf{F}$  has a larger impact on  $\hat{\mathbf{G}}$ . If *e.g.* the input error is zero, then a better estimate of  $\mathbf{G}$  than  $\mathbf{A}$  is in general obtained. The estimator typically filters out noise by averaging or cancelling small terms and  $\hat{\mathbf{G}}$  will thus for certain estimators be different from the inverse of  $\hat{\mathbf{A}}$ . The inverse of a good estimate of  $\mathbf{G}$  can therefore yield an even poorer estimate of  $\mathbf{A}$  than the poor estimate obtained by using the estimator directly and *vice versa*. Depending on the purpose of the study and the error model one or the other is thus more suitable and should then be inferred directly from data, for an example see [Jörnsten et al. \(2011\)](#). The static gain  $\mathbf{G}$  is needed for prediction of the response of the system, while the interaction matrix  $\mathbf{A}$  reveals the interactions among the states. In the following presentation one just needs to swap place between the perturbations and responses and correct the sign if the static gain matrix is sought instead of the interaction matrix.

The data model (2.3.3), which we derived above based on assuming a linear ODE model of the system and steady-state perturbation experiments, is one representative example of when the system can be described as a linear map. In general the data model of a linear mapping can on matrix form be expressed as

$$\mathbf{Y} = \check{\mathbf{Y}} + \mathbf{E}, \quad \mathbf{P} = \check{\mathbf{P}} + \mathbf{F}, \quad \check{\mathbf{A}}\check{\mathbf{Y}} = -\check{\mathbf{P}}, \quad (2.3.9)$$

where the latent variables  $\check{\mathbf{Y}}, \check{\mathbf{P}}$  correspond to the mean or deterministic “true” signal of the responses and perturbations that can be explained by the linear map  $\check{\mathbf{A}} \in \mathbb{R}^{n \times n}$ , which one seeks to infer. From a parameter estimation perspective, this is a linear regression problem in which the parameters of the linear system of equations is sought. By taking the transpose of the variables and “true” network model, and introducing the notation used for regressors  $\check{\mathbf{\Phi}} \triangleq [\phi_1, \dots, \phi_j, \dots, \phi_n] = \mathbf{Y}^T$ , regressands  $\check{\mathbf{\Xi}} \triangleq [\xi_1, \dots, \xi_i, \dots, \xi_n] = -\mathbf{P}^T$ , regressor errors  $\check{\mathbf{\Upsilon}} \triangleq [\nu_1, \dots, \nu_j, \dots, \nu_n] = \mathbf{E}^T$ , and regressand errors  $\check{\mathbf{\Pi}} \triangleq [\epsilon_1, \dots, \epsilon_i, \dots, \epsilon_n] = -\mathbf{F}^T$ , we obtain the matrix form of the standard linear data model used in errors-in-variables regression problems

$$\check{\mathbf{\Phi}} = \check{\check{\mathbf{\Phi}}} + \check{\mathbf{\Upsilon}}, \quad \check{\mathbf{\Xi}} = \check{\check{\mathbf{\Xi}}} + \check{\mathbf{\Pi}}, \quad (2.3.10a)$$

$$\check{\check{\mathbf{A}}}\check{\check{\mathbf{\Phi}}}^T = \check{\check{\mathbf{\Xi}}} \quad \check{\check{\mathbf{\Phi}}}, \check{\check{\mathbf{\Xi}}} \in \mathbb{R}^{m \times n}. \quad (2.3.10b)$$

Here  $m$  is the number of experiments, *i.e.* data points, and  $n$  is the number of states/nodes, *i.e.* variables. We later use this data model as the starting point for solving the robust variable selection and network inference problem in (5.3.2).

Real measurement data is corrupted by noise, but it is nonetheless worth noting that if the data was noise-free, then the “true” network could be obtained by simply solving the linear system of equations specified by (2.3.10b). Current gene expression

data sets are characterized by following properties (see *e.g.* [Holter et al., 2000](#); [Alter et al., 2000](#); [Tegnér and Björkegren, 2007](#); [Faith et al., 2007](#); [Cosgrove et al., 2010](#); [Nordling and Jacobsen, 2009a](#); [Wu and Wu, 2010](#)): They contain few data points compared to the high number of genes and possible interactions, *i.e.* they are of high dimensionality but contain few samples. The measurement uncertainty is large both in the perturbations and responses, *i.e.* the signal to noise ratio is low and errors-in-variables are present. The variables are redundant and nearly collinear, *i.e.* the response matrices are ill-conditioned and most of the variation can be explained by a few linear combinations of the variables. We will in the following chapters of this thesis study how to perform robust inference based on data with these characteristics and how to design additional perturbation experiments when the data is poor.

---

## Interampattiness—a property of biological networks

---

*“If you’re not prepared to be wrong,  
you’ll never come up with anything original.”*

Sir Kenneth Robinson, in his TED talk: Do schools kill creativity?,  
Monterey, California, 2006.

Analysis of gene expression data sets reveals that the variation in expression is concentrated to significantly fewer “characteristic modes” or “eigengenes” than the number of both recorded assays and measured genes. Previous works have stressed the importance of these characteristic modes, but neglected the equally important weak modes. Herein a generic system property—interampattiness—is defined that explains the previous feature, and assigns equal weight to the characteristic and weak modes. An interampatte network is characterised by strong INTERactions enabling simultaneous AMPlification and ATTENUation of different signals. It is postulated that biological networks are interampatte, based on published experimental data and theoretical considerations. Existence of multiple time-scales and feedback loops is shown to increase the degree of interampattiness. Interampattiness has strong implications for the dynamics and reverse engineering of the network. One consequence is highly correlated changes in gene expression in response to external perturbations, even in the absence of common transcription factors, implying that interampatte gene regulatory networks may be erroneously assumed to have co-expressed/co-regulated genes. Data compression or reduction of the system dimensionality using clustering, singular value decomposition, principal component analysis or some other data mining technique results in a loss of information that will obstruct reconstruction of the underlying network.

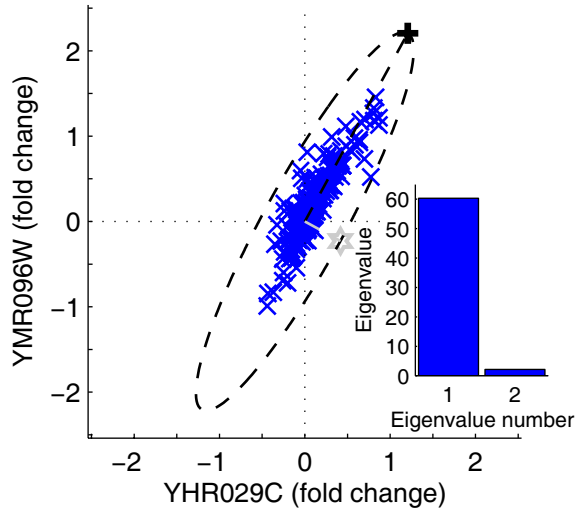
### 3.1 Introduction

Genes interact within complex networks to provide robust biological function. As a consequence of these interactions genes respond to internal and external cues in a correlated fashion, implying that the variation in gene expression is often confined to a subspace of lower dimension than the gene space. Indeed, several authors (Holter et al., 2000; Alter et al., 2000; Kuruvilla et al., 2002; Wu and Dewey, 2006) have observed that most of the variation in the measured response matrix of gene expression changes can be explained by a relatively small number of variables. A few orthogonal directions capture almost all of the biological signal, even though hundreds of arrays have been recorded, each measuring the expression change of thousands of genes. In short, the data matrix is ill-conditioned. A simple example of an ill-conditioned gene response matrix taken from the literature is given in Figure 3.1. The observation that gene response data are ill-conditioned is common, suggesting that it reflects some fundamental property of biological networks.

The observation that microarray data are ill-conditioned has led others to introduce “characteristic modes” (Holter et al., 2000, 2001) and “eigengenes” (Alter et al., 2000; Nielsen et al., 2002; Alter and Golub, 2006; Omberg et al., 2007) in an attempt to explain the observation and connect it to properties of the system. It has been shown that only the dominating modes are required to capture the essential features of the expression data, since they contain essentially all information. In other words, the data set can be compressed or reduced into a few simple patterns given by the dominating modes. These modes have further been connected to biological signals such as cell cycle oscillations and release of pheromone  $\alpha$ -factor synchronisation (Alter, 2007). The collectively conveyed message is that the characteristic modes capture all essential patterns produced by the biological system.

From a systems perspective, however, the characteristic modes only reflect one side of the coin, *i.e.* signals that are amplified by the system. The other side of the coin, *i.e.* signals that are attenuated by the system, is neglected. Since most biological networks should amplify some variations, *e.g.* in signal transduction, while attenuating other variations, *e.g.* in order to maintain homeostasis, it is trivial to realize that both are equally important for the system. For example, in bacterial chemotaxis perfect adaptation is used to reset the response to a change in the nutrition concentration to its initial value, so that the bacterium can maintain its sensitivity to concentration changes and track the nutrition gradient (Yi et al., 2000). In other words, attenuation of the signal generated by a prolonged change in the nutrition concentration enables strong amplification of small changes. If we only looked at the dominating modes then we would miss the core mechanism—the integral feedback that enables the perfect adaptation. Another example is disturbance rejection, or attenuation, of internal build-up of toxic chemicals.

We here introduce a generic system property—interampattiness—that captures both signal amplification and attenuation in gene regulatory networks. It is a property with a simple quantitative measure; essentially the ratio between the ability of the system to amplify and attenuate different signals. The term “interampatte”



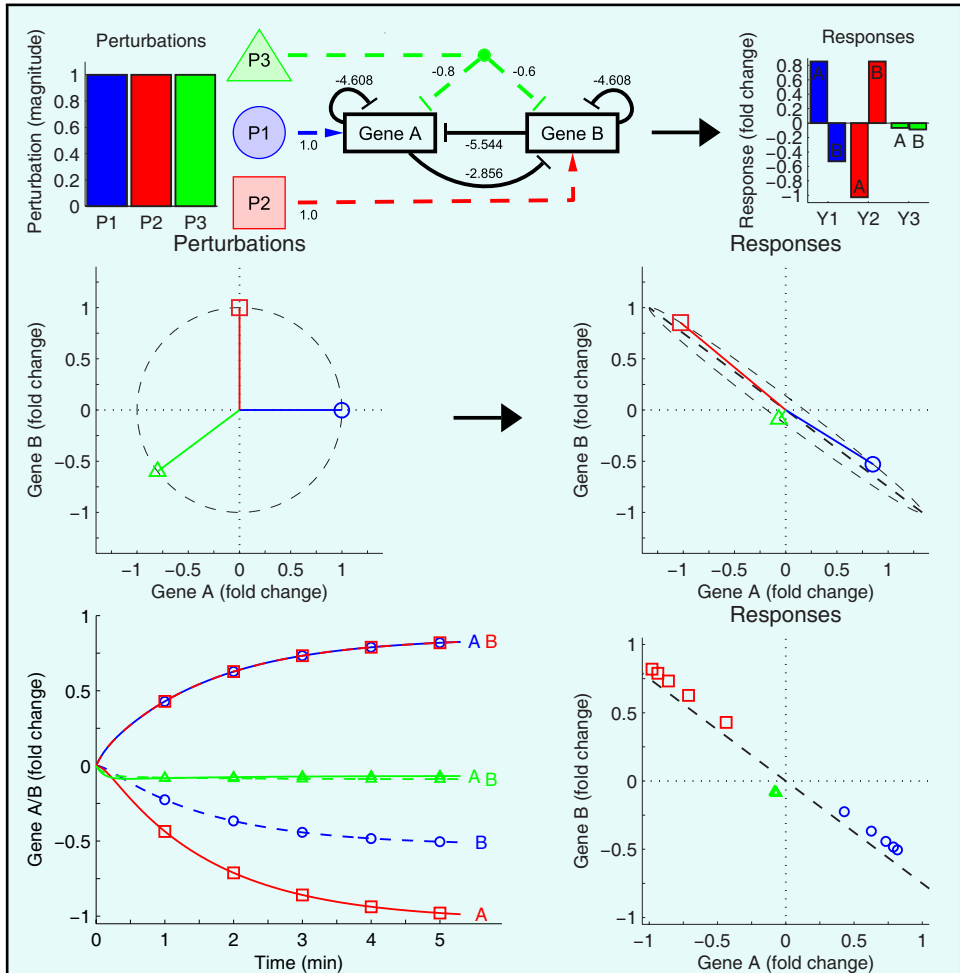
**Figure 3.1:** *Ill-conditioned gene response data.* Fold change in two selected genes from 300 cDNA microarray assays recorded for diverse mutations and chemical treatments in *S. cerevisiae* Hughes et al. (2000). This example illustrates that the data matrix,  $\mathbf{Y} \in \mathbb{R}^{2 \times 300}$ , is ill-conditioned and that the data samples are correlated, *i.e.* all the data points fall more or less on the dashed line ending with a plus. All data points fit within an ellipsoid with a tall and a short principal axis. In this example, the variance in the direction of the tall principal axis is 24 times higher than the variance in the direction of the short principal axis. A more common way to illustrate that the data is ill-conditioned is to show that a few eigenvalues of  $\mathbf{Y}^T \mathbf{Y}$  are significantly larger than the rest, since it also works for data of higher dimension than two. This is illustrated in the right corner of the figure and in Figure 3.8. The eigenvalues reflect the squared spread of the data points, measured from the origin.

(*intər'æmpə,te*) is derived from INTERactions enabling AMPlification and ATTEnuation of signals. An introductory *in silico* example of an interampatte system, illustrating interampattiness and some of its biologically important consequences, is given in Box 1. As we show in this work, interampattiness is a property resulting from long cascades of amplification steps or from positive and negative feedback loops. It is thus partly hard wired in the structure of the network. We also show that any network containing subprocesses operating at different time-scales will be interampatte. Interampattiness is a system property that explains why expression data typically only contain a few dominating modes; the system is designed to amplify some signals while attenuating others. Finally, we show that interampattiness makes reverse engineering of the network a challenging task since the attenuated signals are prone to be hidden in measurement noise.

Knowledge of generic properties of networks is of outmost importance in mod-

elling, analysis and synthetic construction of biological networks. A fact clearly visible in, for example, the classical book *General system theory* (von Bertalanffy, 2006), where von Bertalanffy defines and discusses a number of general system properties. One recently introduced property is scaling in random networks (Barabási and Albert, 1999), which already has had a profound impact on the way we view, model, analyse and construct complex networks (Boccaletti, 2006). We will here show that interampatteness is another network property of significant importance for how biological networks should be conceived. Furthermore, we postulate that interampatteness is a generic property of biological networks, based on both published experimental data and theoretical considerations.

Throughout this work three levels of network abstraction will be employed: (i) a general systems perspective characterized by considering the network as a mapping between two signal spaces, (ii) an intermediate level in which we assume that the mapping can be represented by a set of linear ordinary differential equations (ODEs), and (iii) specific *in vivo* and *in silico* examples to illustrate, motivate or support our claims. In this way we hope to convey the results to readers with any of the diverse backgrounds seen in systems biology. The outline of this chapter is as follows. Interampatteness and how it relates to correlation and common transcription factors (TFs) is illustrated in Box 1, in a setting that we believe most readers with a biological background will be familiar with. We employ a small gene regulatory network (GRN) involving only two genes to illustrate interampatteness in a comprehensive way. Next, the signals and systems perspective is briefly introduced, enabling us to rigorously define interampatteness as a system property. This is followed by a discussion on key features of interampatte systems, largely based on the results presented in Box 1. We then analyse how the degree of interampatteness depends on three fundamental network motifs: cascade, feedback and feedforward, and on the existence of subprocesses with different time-scales. This is followed by a discussion on the implications of interampatteness for network inference, model reduction and data compression. Finally, we postulate that biological networks are inherently interampatte and present experimental and theoretical evidence in support of this claim.



**Box 1:** *Correlated gene response of an interampatte gene regulatory network.* The degree of interampattiness is a measure of the ability of a system to simultaneously amplify and attenuate different perturbations. To illustrate this, consider a gene regulatory network involving two genes *A* and *B* (top, middle). The interactions are for simplicity here described using a linear model. An arrow shaped head on the link indicates up-regulation or activation, while a bar shaped head indicates down-regulation or repression. If we for example perform three steady-state experiments where we apply the external perturbations P1, P2 and P3 (top, left), respectively, then we obtain the responses shown top, right. All external perturbations are relayed to the two genes through different signal pathways, here denoted by the dashed lines. To concretize, let us assume that genes *A* and *B* regulate two alternative metabolic pathways yielding the same product but based

on different nutrients, then P1 can be viewed as an increase in the concentration of one of the nutrients, which through a signal cascade up-regulates gene *A*. After some time, when the system reaches steady-state, the expression of gene *A* is increased, while that of gene *B* is decreased, and the production has adapted to the change of nutrients. We are not interested in the signal cascades of the perturbations here, so we could instead represent each external perturbation by its direct effect on the two genes (middle, left: circle, square, triangle). Similarly, we can study the response of the system directly in the gene space (middle, right). An advantage of the gene space is that the magnitude and direction of each perturbation and response is directly visible as the length and direction of the corresponding vector. Here perturbation P2 is amplified 1.3 times, such that the expression of gene *B* is increased, while that of gene *A* is decreased. Perturbation P3 is attenuated 9 times, *i.e.* the magnitude of the response is 1/9th of the applied perturbation.

A system with a high degree of interampattiness responds in a correlated fashion to perturbations. If we perform many steady-state perturbation experiments with unit magnitude, denoted by the dashed circle (middle, left), then the response of the system will always lay on the dashed ellipsoid (middle, right). Thus, the response of the system depends on the direction of the perturbation, but they all fall almost on the dashed line with slope -0.75 and are therefore strongly correlated. Similarly, a time-series experiment where we measure the expression once every minute shows that the level of expression of one of the genes increases, while it decreases for the second one (bottom, left). If we plot the five samples in the gene space (bottom, right) then they again roughly fall on the same dashed line and are highly correlated, *i.e.* the data are ill-conditioned. The correlation coefficient between the response in the two genes for all samples is -0.97. The two genes appear to be co-expressed or co-regulated, even though they do not have any transcription factor in common. As a matter of fact, the correlation between the response in the two genes is caused by the positive feedback loop that they form. They are mutually down-regulating each other, thereby creating an effect similar to that seen if gene *A* was up-regulated and gene *B* down-regulated by a common transcription factor that all external perturbations acted upon.

### 3.2 A signals and systems perspective on biological networks

The concepts of characteristic modes and eigengenes are based on data mining procedures involving singular value decomposition (SVD) or principal component analysis (PCA), in which one attempts to describe variations in gene expression with as few variables as possible. A low effective dimensionality of the data set does, however, not necessarily reflect any property of the network, but can be a result of the specific perturbations used in the experiments. In order to be able

to distinguish between data properties, as considered in data mining, and system properties, as considered here, we introduce a systems perspective as employed in systems and control theory.

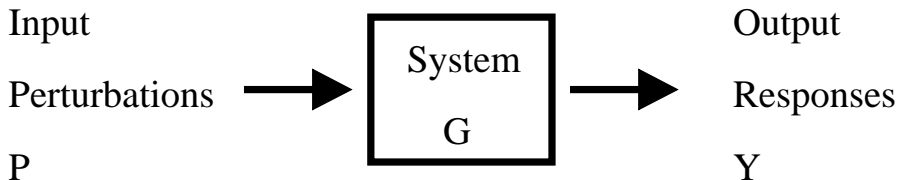
In a systems perspective, a network is simply a system that maps an input signal to an output signal, see Figure 3.2. Here a signal is a function that represents the time variation of a physical variable, or a set of variables arranged into a vector (Lindner, 1999, p. 69). In the case of gene regulatory networks a typical input signal is the applied perturbation, *e.g.* addition of silencing RNA (siRNA), and the output signal is the vector of mRNA levels of a certain set of genes. Both the perturbation and mRNA levels are functions of time, connected through the GRN. A signal space is the set of all possible signals considered, *e.g.* steps, impulses, sinusoids, and a measurement is simply an estimate of a signal at a given time instant. In steady-state experiments a measurement is made only once, when the system has reached its steady-state, while in time-series experiments measurements are collected at several time points.

**Definition 3.2.1.** A system is a mapping between two signal spaces, from the input space to the output space (Green and Limebeer, 1995, p. 79):

$$G : \mathcal{P} \mapsto \mathcal{Y}. \quad (3.2.1)$$

If the perturbation or input is  $\mathbf{p}(t) \in \mathcal{P}$ , then the response or output will be  $\mathbf{y}(t) = G(\mathbf{p}(t)) \in \mathcal{Y}$ . For example, if we are growing *Saccharomyces cerevisiae* (*S. cerevisiae*) in a culture, then the input could be the composition of the growth medium over time, such as galactose and glucose concentration. The output could be fluorescence measurements of Gal3 protein, given that green fluorescent protein has been fused to the *GAL3* gene. Gal3 is involved in the induction of a set of genes that encode galactose metabolizing enzymes and its expression is dependent on the galactose concentration (Suzuki-Fujimoto *et al.*, 1996a). An illustrative example of a signals and systems description of a biological network is the mitogen-activated protein kinase (MAPK) signalling pathway considered in (Wolkenhauer *et al.*, 2005a).

The above definitions of signals and systems are general in the sense that they hold for both deterministic and stochastic systems, *i.e.* they are not restricted



**Figure 3.2: Signals and systems perspective.** A system is a mapping between two signal spaces  $G : \mathcal{P} \mapsto \mathcal{Y}$ . We denote the two signal spaces for the input or perturbation space and output or response space, respectively.

to a certain model formalism. The freedom to choose among commonly used formalisms to describe a given system—graphs, rule-based formalisms, boolean networks, bayesian networks, ordinary and partial differential equations, stochastic master equations (de Jong, 2002)—is essential, since they all have their own advantages, making them better suited for some application.

### 3.2.1 Gene regulatory networks on state-space form

A specific model structure, or formalism, frequently employed for describing gene regulatory networks is the linear state-space model (Gardner et al., 2003; Bansal et al., 2007). The fact that a linear state-space model represents a linear mapping from input signals to output signals simplifies the analysis significantly, and we utilize this when analyzing interampattiness below.

Let us consider a biological system in a particular physiological state (Nordling et al., 2007b), *e.g.* the lytic or lysogenic state of a phage  $\lambda$  infected *Escherichia coli* (*E. coli*). Then it is reasonable to approximate the gene regulatory network using a system of linear first order ODEs, *i.e.* a linear state-space model

$$\frac{d\mathbf{x}}{dt}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{p}(t) \quad (3.2.2a)$$

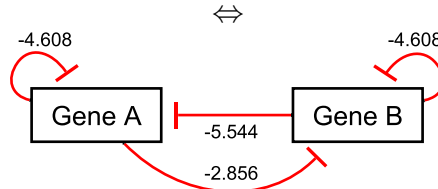
$$\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t). \quad (3.2.2b)$$

The state vector  $\mathbf{x}(t) = [x_1(t), \dots, x_n(t)]^T$  contains the value of each state variable at time  $t$ , defined relative to the corresponding steady-state value, and corresponds to the set of variables needed to completely describe the state of the system at any given time. For a gene regulatory network this would include concentrations of all mRNAs and transcription factors. However, it is common to restrict the state vector to only include mRNA abundances of the considered genes. The perturbation vector  $\mathbf{p}(t) = [p_1(t), \dots, p_m(t)]^T$  contains all external factors of interest that influence the system and that are used to perturb it. It could for example be the Galactose concentration in the growth medium, injection of siRNA fragments (Elbashir et al., 2001) or plasmids containing an extra copy of one or several of the genes. The output  $\mathbf{y}(t) = [y_1(t), \dots, y_o(t)]^T$  contains all the variables that are measured, *i.e.* it may only be possible to measure the abundance change for a subset of the genes. For instance, it is common to only measure the abundance of a few genes with the quantitative real-time polymerase chain reaction (qPCR) technique (Higuchi et al., 1992), while the microarray technology (Scheda et al., 1995) normally is used to measure the expression change of all genes. Note that if we directly measure the mRNA level of all genes in the GRN network of interest then the matrix  $\mathbf{C}$  is diagonal. Similarly, if we directly perturb the rate of transcription, *i.e.* change in mRNA level, of all genes independently in the network then also the matrix  $\mathbf{B}$  is diagonal. In this case we can without restrictions scale our perturbations and responses such that  $\mathbf{B}$  and  $\mathbf{C}$  are equal to the identity matrix,  $\mathbf{I}$ .

The Jacobian matrix  $\mathbf{A}$ , also called interaction, network, connectivity or weighted adjacency matrix, contains the interactions among the states, *i.e.* direct causal

$$\begin{bmatrix} \dot{x}_A(t) \\ \dot{x}_B(t) \end{bmatrix} = \begin{bmatrix} -4.608 & -5.544 \\ -2.856 & -4.608 \end{bmatrix} \begin{bmatrix} x_A(t) \\ x_B(t) \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} p_A(t) \\ p_B(t) \end{bmatrix}$$

$$\begin{bmatrix} y_A(t) \\ y_B(t) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_A(t) \\ x_B(t) \end{bmatrix}$$



**Figure 3.3: Illustration of how a gene regulatory network can be described using a linear state-space model.** This network consists of two genes, *A* and *B*, connected in a feedback loop with mutual down-regulation. Each element of the interaction matrix, *i.e.* weighted adjacency matrix, is indicated on the corresponding edge in the network graph.

influences between the state variables. If element  $a_{ij}$  is zero then no direct causal influence exists from state  $j$  to  $i$ . If  $a_{ij} < 0$  then an increase in state variable  $j$  leads to a decrease in state variable  $i$ , *i.e.* gene  $j$  down-regulates or inhibits gene  $i$ . If  $a_{ij} > 0$  then an increase in state variable  $j$  leads to an increase in state variable  $i$ , *i.e.* gene  $j$  up-regulates or activates gene  $i$ . The size of element  $a_{ij}$  is the strength of the influence. This is illustrated in Figure 3.3. The Jacobian hence contains information about the structure of the network, *i.e.* it is the adjacency matrix of a directed graph with the sign and weight of each edge as its elements (Gross and Yellen, 1999).

For steady-state data, *i.e.* data recorded after applying a linear combination of constant perturbations  $\mathbf{p}_t$  and measuring the response of the system  $\mathbf{y}_t$  when it has reached a new steady-state, equation (3.2.2) gives

$$\mathbf{y}_t = -\mathbf{C}\mathbf{A}^{-1}\mathbf{B}\mathbf{p}_t = \mathbf{G}\mathbf{p}_t. \quad (3.2.3)$$

Let  $\mathbf{Y}[t] \in \mathbb{R}^{o \times t}$  denote the matrix of all measured responses to the  $t$  combinations of perturbations stored in the column vectors of  $\mathbf{P}[t] \in \mathbb{R}^{m \times t}$ , then the data model is

$$\mathbf{Y}[t] = -\mathbf{C}\mathbf{A}^{-1}\mathbf{B}\mathbf{P}[t] = \mathbf{G}\mathbf{P}[t]. \quad (3.2.4)$$

Near a particular stable physiological state the GRN hence acts as a linear mapping from the space of all possible perturbations to corresponding responses of the measured state variables,  $\mathbf{G} : \mathbb{R}^m \mapsto \mathbb{R}^o$ . See Box 1 for an illustration of steady-state experiments.

### 3.2.2 Interampattiness

Having defined a gene regulatory network as a system that maps external perturbations into gene responses, we next consider the relationship between properties of the response data matrix and properties of the system. In particular, we relate the presence of relatively few characteristic modes and eigengenes in the response data to interampattiness of the system, *i.e.* the ability of the system to amplify some signals while attenuating others. In order to do this, we first introduce matrix measures of variation and correlation, which under certain conditions can be applied also to the system. Based on this, we define a general measure of the degree of interampattiness of a system. We thereby highlight the distinction between data properties and system properties.

As stated above, a data matrix with a small number of characteristic modes corresponds to an ill-conditioned matrix. We saw in Figure 3.1 that an ill-conditioned data matrix is characterised by a larger variation in some directions of the gene space and that this can be measured by the eigenvalues of the squared data matrix,  $\lambda_i(\mathbf{Y}^T \mathbf{Y})$ . These eigenvalues measure the squared spread of the data, so if we take the square root of the eigenvalues then we get a measure similar to the standard Euclidean 2-norm,  $\|\mathbf{y}\|_2 = \sqrt{y_1^2 + \dots + y_n^2}$ , called the singular values

$$\sigma_i(\mathbf{Y}) \triangleq \sqrt{\lambda_i(\mathbf{Y}^T \mathbf{Y})}. \quad (3.2.5)$$

The singular value decomposition provides a measure of the variation within the data matrix in each orthogonal direction of the gene space. The maximum variation is given by the largest singular value of  $\mathbf{Y}$  (Horn and Johnson, 1990, p. 420)

$$\bar{\sigma}(\mathbf{Y}) \triangleq \sqrt{\bar{\lambda}(\mathbf{Y}^T \mathbf{Y})}, \quad (3.2.6)$$

where  $\bar{\lambda}$  is the largest eigenvalue. Similarly, the minimum variation is given by the smallest singular value of  $\mathbf{Y}$

$$\underline{\sigma}(\mathbf{Y}) \triangleq \sqrt{\underline{\lambda}(\mathbf{Y}^T \mathbf{Y})}, \quad (3.2.7)$$

where  $\underline{\lambda}$  denotes the smallest eigenvalue. If we take the ratio of the largest and smallest singular values, corresponding to the condition number of the data matrix (Horn and Johnson, 1990, p. 336,442)

$$\kappa(\mathbf{Y}) \triangleq \frac{\bar{\sigma}(\mathbf{Y})}{\underline{\sigma}(\mathbf{Y})} \quad (3.2.8)$$

we obtain a measure of the non-uniformity of the spread of data in the gene space. This was in Figure 3.1 seen as the ratio between the tall and short principal axis of the ellipsoid. A large condition number is an indication of strong correlations within the data set. Correlation is often taken as a token of a physical relationship between variables, *i.e.* a reflection of a system property. Data mining based on SVD tries to

find such underlying physical relationship and reduce the complexity of the data by removing the directions with a small variation, *i.e.* the small singular values, since they contribute little to the gene response data. But as discussed in Box 1, data mining can be misleading and it is therefore important to distinguish between data properties and system properties.

Assume now that the system acts as a linear mapping  $\mathbf{G}$  from the space of all possible perturbations to the corresponding responses in gene expressions, according to (3.2.4). Then the ill-conditioning of the response data matrix  $\mathbf{Y}$  can be due to a non-uniform set of perturbations, reflected by a high condition number of the perturbation matrix  $\mathbf{P}$ , or due to a property of the system  $\mathbf{G}$  that serves to amplify certain perturbations while attenuating others. For the case of steady-state experiments, the linear mapping  $\mathbf{G}$  is a constant matrix and the ratio between maximum amplification and maximum attenuation exerted by the system can be computed using the condition number and we then have (Skogestad and Postlethwaite, 1996, p. 525)

$$\kappa(\mathbf{Y}) \leq \kappa(\mathbf{G})\kappa(\mathbf{P}) \quad (3.2.9)$$

which shows that if  $\mathbf{Y}$  is ill-conditioned, then either  $\mathbf{G}$  must be interampatte or  $\mathbf{P}$  ill-conditioned.

For more general systems, including dynamic and nonlinear systems, it is more instructive to consider the maximum and minimum gains of the system in terms of induced norms. For the case of a linear static mapping  $\mathbf{G}$ , the induced 2-norm happens to be equivalent to the maximum singular value

$$\max_{\|\mathbf{v}\|_2=1} \|\mathbf{G}\mathbf{v}\|_2 = \bar{\sigma}(\mathbf{G}), \quad (3.2.10)$$

and is obtained for the input signal, *i.e.* perturbation, that is amplified the most by the system. Similarly, the minimum gain corresponds to the minimum singular value

$$\min_{\|\mathbf{v}\|_2=1} \|\mathbf{G}\mathbf{v}\|_2 = \underline{\sigma}(\mathbf{G}), \quad (3.2.11)$$

and is obtained for the input signal, *i.e.* perturbation, that is attenuated the most by the system. Here the input signal  $\mathbf{v}$  is a vector of unit length pointing in a certain direction of the gene space, and in principle the induced norm searches through all directions to identify the one amplified the most and least by the system, respectively. Associated with the singular values are hence specific directions of the input vector and corresponding output directions for the obtained responses. These directions can be obtained from the singular value decomposition (Skogestad and Postlethwaite, 1996, p. 76-77)

$$\mathbf{G} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H. \quad (3.2.12)$$

where the columns of the matrices  $\mathbf{U}$  and  $\mathbf{V}$  correspond to the output and input directions, respectively, for each of the singular values in the diagonal matrix  $\mathbf{\Sigma} = \text{diag}(\bar{\sigma}, \dots, \underline{\sigma})$ . If the ratio between the strongest and weakest gain is large, *i.e.* if the condition number of  $\mathbf{G}$  is large, then we say that the system is interampatte.

This is similar to the data case in which we say that the data matrix is ill-conditioned, or that the data are strongly correlated. The reason for introducing the term interampatte is to make a clear distinction between system properties and data properties, and to be able to generalize the degree of interampatteness to more general dynamic and nonlinear systems.

Note again that an interampatte  $\mathbf{G}$  implies that the network will amplify certain directions of the perturbation matrix significantly more than other directions, and hence typically result in an ill-conditioned response matrix  $\mathbf{Y}$ . However, the opposite is in general not true, that is, an ill-conditioned response matrix  $\mathbf{Y}$  may be the result of an ill-conditioned perturbation matrix  $\mathbf{P}$ , *e.g.* perturbations of very different magnitude. The presence of characteristic modes in  $\mathbf{Y}$  may thus be a result of the choice of perturbation matrix  $\mathbf{P}$ , and characteristic modes are hence not properties of the system *per se*.

To generalize the concept of interampatteness to dynamic and nonlinear systems, we employ an induced norm based on the input and output signal spaces defined in (3.2.1)

**Definition 3.2.2.** We define the degree of interampatteness as

$$\kappa(G) \triangleq \sup_{\|\mathbf{r}(t)\|_\alpha = \|\mathbf{s}(t)\|_\alpha} \frac{\|G(\mathbf{r}(t))\|_\beta}{\|G(\mathbf{s}(t))\|_\beta}, \quad \mathbf{r}(t), \mathbf{s}(t) \in \mathcal{P}, \quad G : \mathcal{P} \mapsto \mathcal{Y}. \quad (3.2.13)$$

Here  $\|\cdot\|_\alpha$  denotes some norm on the input space  $\mathcal{P}$  and  $\|\cdot\|_\beta$  denotes some norm on the output space  $\mathcal{Y}$ . Both the input and output spaces are required to be normed vector spaces.

This definition of interampatteness is valid for any system, including nonlinear ones, since it is based directly on the definitions of a signal and a system. We only require the signal spaces to be normed vector spaces, which is a necessary condition to properly quantify the signals. Selection of which norm to use is left to the user, since it will depend on the application and formalism used to model the system. We have already seen that the Euclidean norm works well for systems acting as linear mappings, *e.g.* steady-state experiments, and that the degree of interampatteness then is given by the condition number of  $\mathbf{G}$ . For time-varying signals the  $\mathcal{L}_2$ -norm  $\|\mathbf{y}(t)\|_{\mathcal{L}_2} = \int_{-\infty}^{\infty} \|\mathbf{y}(t)\|_2^2 dt$ , which measures the total energy of a signal, is often a suitable choice.

As discussed above, the degree of interampatteness can be seen as the ratio between the systems ability to amplify and attenuate signals, which was illustrated in Box 1. From an energy perspective it can be interpreted as a measure of the ability of a system to channel energy into different subspaces.

Note that in many cases a complete model of the GRN is not available, and then a lower bound on the degree of interampatteness can be useful

**Definition 3.2.3.** The degree of interampattiness is at least  $\tilde{\kappa}(G)$  for  $G : \mathcal{P} \mapsto \mathcal{Y}$  if

$$\begin{aligned} &\exists \text{ some } \mathbf{r}(t), \mathbf{s}(t) \in \mathcal{P} \setminus \{0\} \text{ such that} \\ &\tilde{\kappa}(G) = \frac{\|G(\mathbf{r}(t))\|_{\beta} \|\mathbf{s}(t)\|_{\alpha}}{\|G(\mathbf{s}(t))\|_{\beta} \|\mathbf{r}(t)\|_{\alpha}}. \end{aligned} \quad (3.2.14)$$

Here  $\|\cdot\|_{\alpha}$  denotes some norm on the input space  $\mathcal{P}$  and  $\|\cdot\|_{\beta}$  denotes some norm on the output space  $\mathcal{Y}$ . Both the input and output spaces are required to be normed vector spaces.

For systems acting as linear mappings, a straight forward lower bound  $\tilde{\kappa}(G)$  based on the Euclidean norm is

$$\tilde{\kappa}(G) = \frac{\kappa(\mathbf{Y}[t])}{\kappa(\mathbf{P}[t])}, \quad (3.2.15)$$

where  $\kappa(\mathbf{Y}[t])$  and  $\kappa(\mathbf{P}[t])$  are the condition numbers of the response and perturbation matrix, respectively. Note that any reasonably smooth nonlinear system can be approximated locally by a linear model and the degree of interampattiness of the linear approximation hence serves as a lower bound for the nonlinear system.

In conclusion, networks with a large  $\kappa(G)$  or  $\tilde{\kappa}(G)$ , are strongly interampatte implying that the network will have a highly uneven amplification of different perturbations. The response data of interampatte systems will in general be characterized by few characteristic modes relative to the number of genes and recorded assays.

### 3.3 Features of strongly interampatte systems

The degree of interampattiness was defined in the previous section for general systems. Here we consider how interampattiness manifests itself in gene response experiments and discuss some important implications. The discussion here further generalises and deepens the picture provided in Box 1 and is hence centred around the same two gene *in silico* GRN.

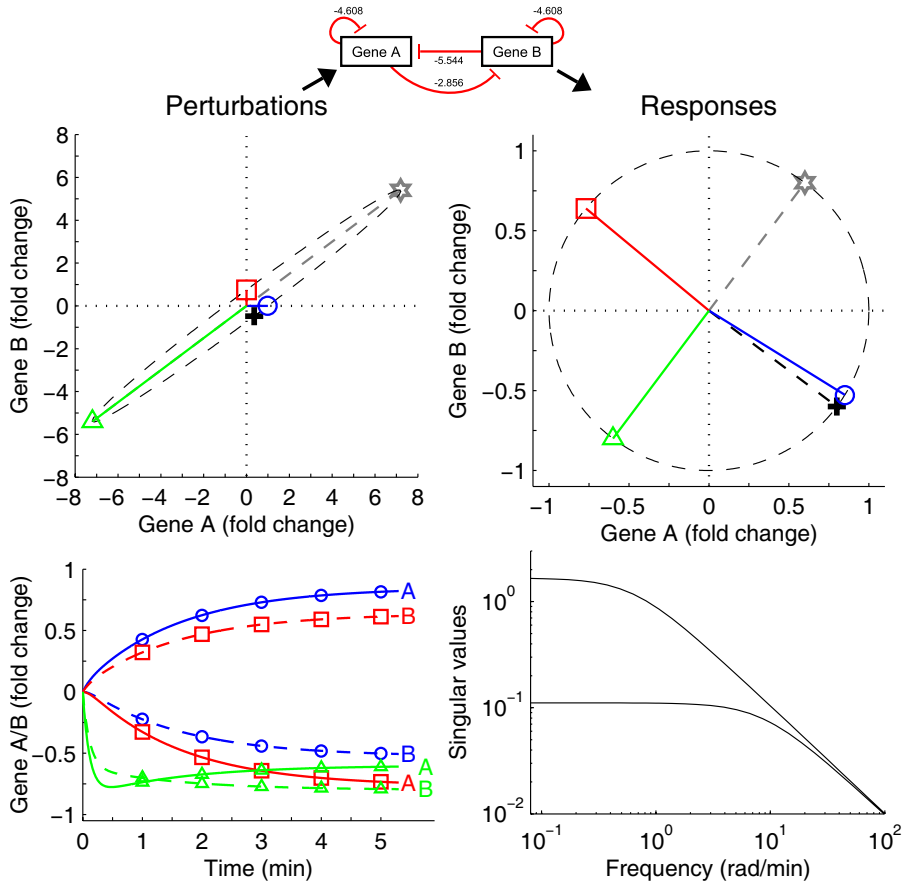
The fact that a system with a high degree of interampattiness will amplify some input signals, while attenuating other signals, implies that random perturbations typically will yield strongly correlated responses. This is illustrated in Box 1 and Figure 3.1. In particular, if the system is linear and the perturbations are uniform on the perturbation space, then the condition number of the steady-state response matrix will be equal to the interampatte number of the system. In the literature, genes displaying a strong correlation in their responses are usually said to be co-expressed or co-regulated and often assumed to have a common TF. However, as shown in Box 1, no common TF is needed for a strongly correlated response. Albeit the response data will be similar for co-regulated genes and interampatte networks, the cause of the correlation is crucial for understanding the functional working of the gene network. Indeed, it is reasonable to assume that strong correlations in gene response data often are caused by network interactions rather than common TFs.

Clustering is a popular method for analysing microarray data (D'haeseleer et al., 2000; Quackenbush, 2001; Slonim, 2002; Knudsen, 2004; Quackenbush, 2006). The

basic idea of clustering is to group things that are close according to some distance metric or similarity measure (Do and Choi, 2008), *i.e.* the correlations within the data set are utilized. In Box 1 the expression of one gene is consistently increased while that of the other is decreased in response to all perturbations, implying that all samples fall approximately on the same line. The responses of the two genes are strongly correlated both in the steady-state and time-series experiments, and the genes would therefore be assigned to the same cluster based on either experiment. For our small example it would imply that the feedback loop responsible for the regulation of the circuit and the interampattiness of the system is neglected. Important mechanistic insight is thereby lost. On the other hand, if the purpose of the model is to predict the response of the system to random perturbations, then the grouping of the genes into a cluster can be beneficial, since the complexity of the model is reduced while the predictive strength is maintained.

As discussed above, correlation within the data set is also the basis for identifying signal patterns such as “characteristic modes” (Holter et al., 2000, 2001) and “eigenmodes” (Alter et al., 2000; Nielsen et al., 2002; Alter and Golub, 2006; Omberg et al., 2007) using clustering, SVD, PCA or some other data mining technique. The characteristic modes and eigenmodes are signal patterns that depend on both the set of perturbations and the system, and they are therefore features of the response data set and do not necessarily reflect any properties of the system. A necessary requirement to connect them to properties of the system is knowledge about the input signals, *i.e.* the magnitude and direction of the perturbations in the case of steady-state experiments. Measuring and reporting the perturbations used in an experiment is thus as important as measuring and reporting the response of the system.

Interampattiness is an inherent property of a biological network, while ill-conditioning of a response matrix also depends on the perturbations used to excite the network. The latter implies that the perturbations in principle can be chosen such that a well conditioned response matrix is obtained even for strongly interampatte systems. For steady-state experiments, the relation between the response, perturbations and system is given by equation (3.2.4). Consider again the two gene example in Box 1 and change the magnitude of the three external perturbations by scaling them by a factor 1, 0.75 and 9, respectively. This yields the perturbations and responses shown in Figure 3.4. As seen, the response in all three perturbation experiments now has magnitude one, *i.e.* we have obtained an uncorrelated response from the same interampatte GRN that in Box 1 gave a strongly correlated response. Instead, the perturbations are now strongly correlated and close to the dashed line with slope 0.75, which corresponds closely to the input direction of the weak mode of the system. We have hence moved the correlation from the output signal to the input signal and demonstrated that either the perturbations or the responses of an interampatte system must be correlated. Note that the strong modes of the system are visible in the response to uniform perturbations, while the weak modes are visible in the perturbations yielding a uniform response, an observation that we will utilize when studying reverse engineering below.



**Figure 3.4: Uncorrelated gene response data.** Correlated perturbations (middle, left) yield a uniform excitation of the gene space (middle, right) for the gene regulatory network in Figure 3.3. The plus denotes the input direction corresponding to maximum amplification, *i.e.* the strong mode of the system. Similarly, the hexagram denotes the direction of maximum attenuation. The bottom left figure shows the corresponding responses observed in the time-series experiments using the same perturbations as above. The bottom right figure shows the maximum and minimum singular values of the two gene network as a function of frequency, illustrating that the amplification and interampattiness of a system depends on the frequency of the perturbation.

In Figure 3.4 the perturbations are chosen such that they counteract the intrinsic interampattiness of the system. The degree of interampattiness of the two gene system is  $\text{cond}(\mathbf{A}) = 15$ , implying that a perturbation in the direction of maximum amplification, *i.e.* exciting the strong mode, will yield a response that is 15 times stronger than that of a perturbation in the direction of maximum attenuation,

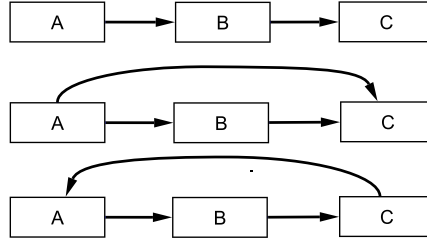
*i.e.* the weak mode. This implies that a uniform excitation of the gene space will require very specific sets of perturbations with widely different magnitudes. Random perturbations will therefore typically yield a strongly correlated response. To connect the characteristic modes seen in a data set to the strong modes of the system it is hence in general sufficient to make sure that the perturbations are reasonably uniform on the perturbation space.

Considering time-series experiments, the sampled expression profiles will typically be almost linearly dependent (Crampin et al., 2004) and hence not provide a more uniform excitation of the gene space compared to steady-state experiments. This is illustrated in Figure 3.4, where the samples in a time-series experiments following a specific perturbation are all in essentially the same direction of the gene space. The linear dependence of samples has indirectly been noted by Faith et al. (2007) who observed that they could obtain the same predictions with 60 well-chosen arrays as with all 445 microarrays. Note that if we instead had sampled the response during the initial transient, *i.e.* the first 30 seconds, then we would have obtained a strong signal in the direction of the weak mode. In Figure 3.4 the response to P3 grows rapidly in magnitude, while the response of P1 and P2 still is small. The weak mode is connected to the fast time-scale of the system, *i.e.* smallest time-constant, while the strong mode is connected to the slow time-scale of the system. In general, dynamic lags in the network connections cause a delay before the interactions take full effect. A response in a direction different from that of the strong mode, which dominates the steady-state response, is therefore likely to be obtained in samples taken during the initial transient of a system. This time dependence is best visualised as a function of frequency, and is illustrated for the two gene network in Figure 3.4. For high frequencies, corresponding to a rapid measurement during the initial transient, all singular values are equal leading to a uniform excitation of the gene space in response to uniform perturbations. However, we also note that all singular values are small at high frequencies, implying strong attenuation independent of the direction of the perturbation.

To conclude this section, the degree of interampattiness is a measure of the ability of a system to amplify some signals while attenuating other signals. The response data depends on both the system and the applied perturbations, but for an interampatte system either the perturbations or the responses need to be correlated. Correlation of the response data enables data compression and reduction of the system complexity using clustering, SVD or some other data mining technique. However, this should be performed with care as it may effectively hide the mechanism generating the correlations.

### 3.4 Interampattiness depends on network motifs

Interampattiness is a system property that results in correlated gene responses to internal and external cues as a result of gene-to-gene interactions within the network. Indeed, we have already seen in the two-gene example in Box 1 how feedback



**Figure 3.5:** Three fundamental network motifs: cascade (top), feedforward (middle) and feedback (bottom).

interactions can cause strongly correlated responses corresponding to a high degree of interampattiness. The purpose of this section is to further explore the impact of fundamental interconnections of genes within networks, *i.e.* basic network motifs, on the degree of interampattiness. We here only consider examples using randomly generated networks. Analytical results on the impact of various fundamental types of interconnections on interampattiness are provided in [Nordling and Jacobsen \(2009a, Supplementary Section S3\)](#).

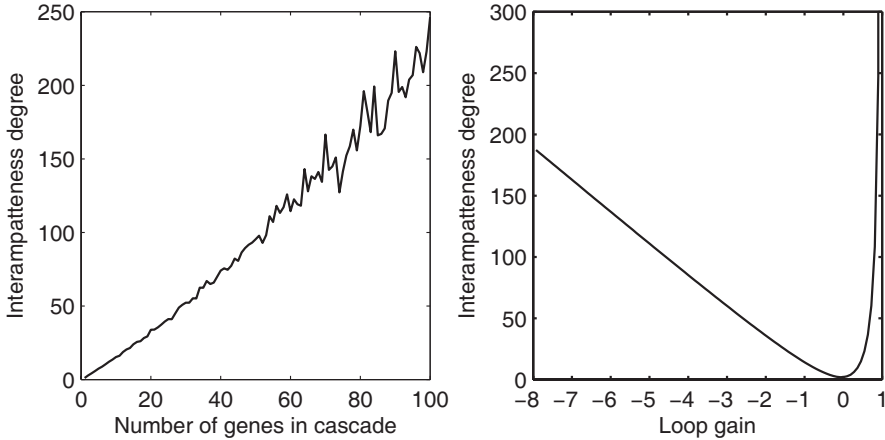
For linear systems, there exist three fundamentally different interconnections of subsystems, or genes; cascade (series), feedforward (parallel) and feedback. These three basic motifs are illustrated in [Figure 3.5](#).

Consider first the case of a cascade of  $n$  genes in which the product of one gene acts as a transcription factor for the next gene. Denote the steady-state amplification of gene  $i$  by  $G_i$ . Then, as shown in [Nordling and Jacobsen \(2009a, Supplementary Section S3\)](#), the following bounds apply for the minimum and maximum steady-state amplification of the overall network

$$\underline{\sigma}(G) \geq \min_i \frac{1}{1 + \frac{1}{|G_i|}} , \quad \bar{\sigma}(G) \leq \max_{i,j} n \prod_{k=i}^j |G_k| \quad (3.4.1)$$

Note that the lower bound for the minimum amplification is related to the amplification of a single gene, while the upper bound is the maximum product of single gene amplifications scaled by the number of genes. These bounds indicate that the interampattiness should scale with the total number of genes  $n$  in the cascade. Indeed, this is supported by the results in [Figure 3.6](#) which show the degree of interampattiness as a function of the number of genes  $n$  in the cascade. For each value of  $n$  we generated 100 random cascades with each gene having normally distributed gains with mean 1 and standard deviation 0.2. As seen, the mean interampattiness number increases more than linearly with the number of genes. This suggests that large networks are likely to have a higher degree of interampattiness than smaller networks.

Consider next a network consisting of  $n$  genes connected in a cascade and with feedback regulation from gene  $n$  to gene 1, corresponding to a single feedback loop.



**Figure 3.6: Degree of interampattness depends on the network motif.** The degree of interampattness  $\kappa$  as a function of the number of genes  $n$  in a cascade pathway (left). For each  $n$ , 100 random networks were generated and the figure shows the mean  $\kappa$ . The degree of interampattness  $\kappa$  as a function of the feedback loop-gain  $L$  in a feedback loop involving 3 genes with feedback regulation of gene 1 by gene 3 (right). A positive  $L$  corresponds to positive feedback while a negative  $L$  corresponds to negative feedback.

Denote the product of the amplification of the individual genes the loop-gain

$$L = G_1 G_2 \dots G_n. \quad (3.4.2)$$

Now  $L > 0$  corresponds to positive feedback,  $L = 0$  to no feedback and  $L < 0$  to negative feedback. As shown in Nordling and Jacobsen (2009a, Supplementary Section S3), negative feedback will make the minimum gain of the network decrease to zero as the loop-gain  $L \rightarrow -\infty$ , while positive feedback will make the maximum gain approach infinity as the loop-gain  $L \rightarrow 1$ . In both cases, the interampattness number  $\kappa$  will approach infinity in the limit. We should therefore expect both positive and negative feedback to cause interampattness which, furthermore, increases with the magnitude of the loop-gain. Note that an infinite negative loop-gain corresponds to integral action, as seen for instance in bacterial chemotaxis, while a positive loop-gain of one corresponds to a singularity in the network, as seen *e.g.* in bistable switches. A positive loop-gain exceeding one makes the system unstable. Figure 3.6 shows the interampattness number as a function of the loop-gain for a feedback network with three genes. The results are based on 500 randomly generated networks with normally distributed gains for the individual genes. The Monte-Carlo simulation supports the analytical results, *i.e.* the degree of interampattness increases as a function of the loop-gain both for positive and negative feedback.

For the case of feedforward, or parallel, interconnections we simply note from the

derivation in Nordling and Jacobsen (2009a, Supplementary Section S3) that they have little or no impact on the interampattiness of the network, and we therefore do not show any results here. In summary, we conclude that long cascades of genes as well as feedback connections tend to increase the interampattiness of GRNs.

### 3.5 Interampattiness related to time-scale separation

Biochemical reactions are known to occur at widely different time-scales (Crampin et al., 2004; Heinrich and Schuster, 1996). Time scale separation has direct implications for the degree of interampattiness of a network. In particular, consider the case with a linear network on the state-space form (3.2.2) with  $\mathbf{B} = \mathbf{C} = \mathbf{I}$ . We then have that the mapping from inputs to outputs is  $\mathbf{G} = -\mathbf{A}^{-1}$  where  $\mathbf{A}$  is the interaction matrix of the network. The time-scales of the network are given by the real part of the eigenvalues of  $\mathbf{A}$ , or more precisely, the time-constants of the system are defined as  $\tau_i = 1/|\Re(\lambda_i(\mathbf{A}))|$ . The smallest and largest eigenvalues of  $\mathbf{A}$  provide upper and lower bounds for the minimum and maximum singular values (Skogestad and Postlethwaite, 1996, p. 522)

$$\underline{\sigma}(\mathbf{A}) \leq |\underline{\lambda}(\mathbf{A})| \ll |\bar{\lambda}(\mathbf{A})| \leq \bar{\sigma}(\mathbf{A}), \quad (3.5.1)$$

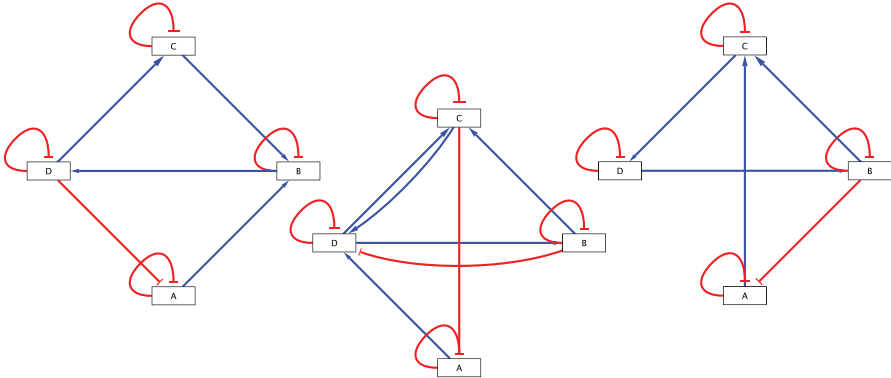
and the degree of interampattiness is given by the condition number of  $\mathbf{G}$

$$\kappa(\mathbf{G}) = \frac{\bar{\sigma}(\mathbf{G})}{\underline{\sigma}(\mathbf{G})} = \frac{\bar{\sigma}(\mathbf{A})}{\underline{\sigma}(\mathbf{A})}. \quad (3.5.2)$$

Thus, an increasing difference between the smallest and largest eigenvalue of  $\mathbf{A}$ , corresponding to increasing time-scale separation, increases the degree of interampattiness of a system.

### 3.6 Implications of interampattiness for reverse engineering of GRNs

Reverse engineering, or network inference, is an important subject that has received considerable attention but shown only limited success in practice (Li et al., 2008b; Cho et al., 2007; Christensen et al., 2007). The aim of network inference is to determine the structure of the network, *i.e.* the signed directed graph, based on measurements of expression changes following known perturbations of the system. Typically several different perturbations are applied to the system and a few microarrays are recorded in each experiment. The presence of weak modes in the network will imply that the collected response matrix can be ill-conditioned, *i.e.* the data set is nearly rank deficient. For strongly interampatte systems this is likely to be the case even after recording significantly more samples than the number of measured genes. As an example we here consider an interampatte four gene network, shown in the middle of Figure 3.7, and perform as many steady-state perturbation experiments as there



**Figure 3.7: Two proposed models (left, right) for the true gene regulatory network shown in the middle.** The steady-state responses of all three networks are almost identical when the genes are perturbed one-by-one using perturbations of unit magnitude, implying that the network structure is practically unidentifiable. This unidentifiability is due to the high degree of interampattiness of the true network, which causes the response matrix to be ill-conditioned with the weak modes hidden in measurement noise. Here every link represents a direct causal effect. An arrow shaped head on the link from  $A$  to  $B$  indicates that gene  $A$  upregulates (activates) gene  $B$ , while a bar (T) shaped head from  $D$  to  $A$  indicates that gene  $D$  downregulates (represses) gene  $A$ . The figure was generated in Cytoscape (Shannon et al., 2003).

are genes in the network. As we show, the ill-conditioning of the response matrix may lead to inference of grossly erroneous network structures. We consider standard steady-state perturbation experiments of the type performed for example in Gardner et al. (2003). We do not have any real biological model system to perform experiments on here, but we simulate the experiments *in silico* using the interampattiness system shown in the middle of Figure 3.7. The four genes are perturbed one-by-one in four separate experiments using perturbations of magnitude one. These represent the direct effect on the genes of exogenous perturbations similar to P1 and P2 in Box 1. The measured response of the system, in terms of change of expression of the four genes, is simulated as

$$\mathbf{Y} = -\mathbf{A}^{-1}\mathbf{P} + \mathbf{E}. \quad (3.6.1)$$

Here  $\mathbf{A}$  is the interaction matrix of the true network, see Nordling and Jacobsen (2009a, Supplementary Table S1),  $\mathbf{P}$  contains the four perturbations, *i.e.*  $\mathbf{P} = \mathbf{I}$  with condition number 1, and  $\mathbf{E}$  represents the measurement noise. In our case,  $\mathbf{E}$  is a random matrix constructed by drawing each element from a normal distribution with zero mean and standard deviation 0.1. We hence assume independent and identically distributed measurement noise,  $\mathbf{E} \sim \mathcal{N}(\mathbf{0}, 0.01\mathbf{I})$ . The selected noise level, *i.e.* standard deviation, is unrealistically low for biological data obtained through microarray or qPCR experiments, but as we will show it is still too high in the

sense that it causes practical unidentifiability of the network structure. The degree of interampattiness of the true GRN is  $\kappa = 319$  and the condition number of the collected response data matrix is approximately 300, *i.e.* slightly lower due to the added measurement noise.

A common way of evaluating the fit of a model to a data set is by calculating the prediction error according to (Ljung, 1999)

$$\varepsilon_i = \left\| -\hat{\mathbf{A}}_i^{-1} \mathbf{P} - \mathbf{Y} \right\|_2. \quad (3.6.2)$$

Here  $\hat{\mathbf{A}}_i$  denotes the interaction matrix of the model being evaluated,  $\mathbf{P}$  is the employed perturbations, and  $\mathbf{Y}$  is the simulated response of the true network. Typically, most inference algorithms aim at finding a model that minimises the prediction error in some sense, given certain constraints, *e.g.* sparsity requirements (Tegnér and Björkegren, 2007). The most widespread method is ordinary least squares. The main purpose here is to illustrate that near rank deficiency of the response matrix leads to practical unidentifiability of the network structure of interampatte systems, and hence it is sufficient to show that there exists some model with an erroneous structure that yields a prediction error comparable to that of the true network. The models in the left and right of Figure 3.7 yield almost the same prediction error: 0.28 and 0.27, respectively. This should be compared to the prediction error of the true network model: 0.31. The prediction errors are similar, so they all predict the network response approximately equally well, which shows that the data set does not contain the necessary information to uniquely determine the network structure. In fact, the prediction error of the two incorrect models is slightly lower, so most inference algorithms would prefer them over the true model. Note that we have only performed four experiments, so the standard regression problem has a unique solution that yields zero prediction error. The unique solution, shown in Nordling and Jacobsen (2009a, Supplementary (S54)), differs significantly from the true network and is overfitted, *i.e.* it compensates for the stochastic measurement noise. The details of the three models and this simulated experiment are given in Nordling and Jacobsen (2009a, Supplementary Section S4).

The structure of the two models and the true network differ significantly, as can be seen from Figure 3.7. As a matter of fact, the structures, *i.e.* signed directed graph, of the left and right models are completely different. They do not have a single link in common, except for the downregulating self-loops, which account for the self-degradation of each mRNA. The left model shares two links with the true network, but one of them is an up-regulation in the left model while it is a down-regulation in the true network. The closest network structures are the right model and the true network, having three links in common. Note that not only the structure differs between the three networks, but also the strength of the links differ, see Nordling and Jacobsen (2009a, Supplementary Tables S1, S2, and S3). In particular, the three links that the right model and the true network share have highly different strengths. Also note that the two erroneous structures are more

sparse than the true network, and hence assuming sparseness as suggested in (Tegnér and Björkegren, 2007) is of no help in this case.

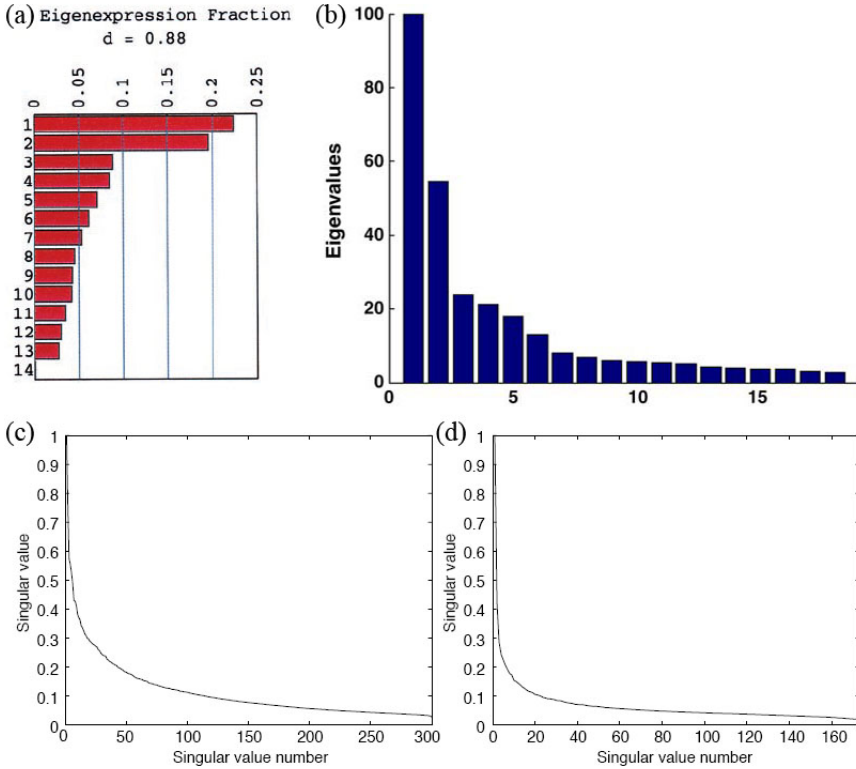
The example considered here serves to illustrate that interampatte systems yield gene response data that are ill-conditioned, implying that the weak modes are largely hidden in the measurement noise, thereby hampering the inference of the correct network structure. On the other hand, if the only objective is to obtain a model that can predict the response of the network, then in fact all three models in Figure 3.7 serve the purpose well, as can be seen by performing validation experiments. For validation purposes, we simulated 16 experiments with random perturbations of an average magnitude of one and tested how well the three models could predict the response data. The relatively small prediction errors of 0.44, 0.40 and 0.42 for the left, right and true models, respectively, show that all three models are good for prediction. The perturbations and responses are presented in Nordling and Jacobsen (2009a, Supplementary Section S4).

The only way to overcome the inference problems related to ill-conditioned response data is to design experiments with the aim of obtaining a more uniform excitation of the gene space. Equation (3.2.4) tells us that the data recorded in  $t$  steady-state perturbation experiments will be ill-conditioned, unless  $\mathbf{P}[t]$  is designed such that it counteracts the interampattiness of  $\mathbf{G}$ . Such a design was illustrated in Figure 3.4. The design of correlated perturbation experiments yielding uniform excitation of the gene space require knowledge of the interaction matrix  $\mathbf{A}$ , hence pointing to the need for designing the perturbation experiments in an iterative fashion.

### 3.7 Evidence of strongly interampatte bio-networks

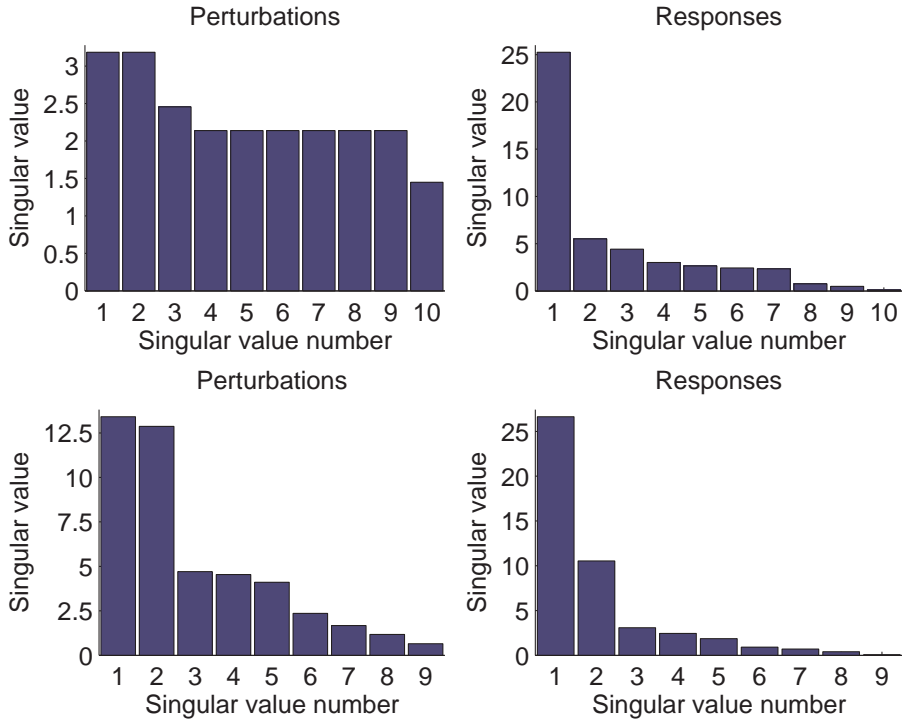
As stated above, there are frequent reports of strong correlation between gene responses in the literature. Some reported data sets that are ill-conditioned are shown in Figure 3.8 showing that essentially all variation is concentrated to a few directions, corresponding to the dominating singular values. Strong correlation between the responses of different genes is often taken as evidence for co-expression or co-regulation, see *e.g.* Eisen et al. (1998); Spellman et al. (1998) and Aoki et al. (2007). However, as we have seen in this paper, strong correlation between genes may also be a consequence of specific network structures such as negative or positive feedback loops. The probability that randomly chosen perturbations of the same magnitude yield a similar response increases with the interampattiness of the system.

To conclude that reported ill-conditioned expression data are caused by interampatte systems requires knowledge of the perturbations applied, something that unfortunately usually is not reported in the literature. We have however identified two biological systems with a high degree of interampattiness: a ten gene network of the Snf1 signalling pathway in *S. cerevisiae* (Lorenz et al., 2009), and a nine gene subnetwork of the SOS pathway in *E. coli* (Gardner et al., 2003). The lower bounds on the degree of interampattiness for the two networks, calculated using equation



**Figure 3.8: Ill-conditioned response matrices, characterised by a few dominating singular values or eigenvalues, have in particular been observed when analysing microarray data.** Here we have reproduced some published data: (a) *S. cerevisiae* cell cycle data synchronised by elutriation (Alter et al., 2000) (Copyright 2000 National Academy of Sciences, U.S.A.), (b)  $\alpha$ -factor synchronised yeast cell-cycle time-series data (Wu and Dewey, 2006) (originally published by Springer), (c) *S. cerevisiae* deletion mutant steady-state data and (d) *S. cerevisiae* perturbed through environmental stress time-series data (Kuruvilla et al., 2002) (originally published by BioMed Central).

(3.2.15), are  $\tilde{\kappa}(\mathbf{G}) = 98$  and  $\tilde{\kappa}(\mathbf{G}) = 19$ , respectively. In this case a better estimate of the degree of interampattiness is however obtained by estimating the interaction matrix  $\mathbf{A}$  using ordinary least squares and then calculating its condition number, which is equal to the degree of interampattiness of the system, yielding  $\kappa(\mathbf{G}) = 253$  and  $\kappa(\mathbf{G}) = 54$ , respectively. This estimate accounts for alignment between the perturbations and responses, and is hence less conservative than the lower bound. The singular values of both the perturbation matrix and response matrix for both systems are shown in Figure 3.9. Both networks have one characteristic mode in



**Figure 3.9: Experimental evidence of interampattene gene regulatory networks.** Two biological systems with a high degree of interampatteness: a ten gene network of the Snf1 signalling pathway in *S. cerevisiae* (Lorenz et al., 2009) (top), and a nine gene subnetwork of the SOS pathway in *E. coli* (Gardner et al., 2003) (bottom). Both systems have one characteristic mode in which the perturbations are strongly amplified.

which the perturbations are amplified strongly. For the network in (Lorenz et al., 2009) the characteristic mode of the system is seen directly in the responses, since the perturbations are orthogonal and of similar magnitude. For the network in (Gardner et al., 2003) the existence of one characteristic mode is verified by performing a singular value decomposition (3.2.12) of the least squares estimate of the system matrix  $\mathbf{G}$ .

Every cell needs to maintain homeostasis, characterised by an internal environment far from thermodynamic equilibrium with the surrounding environment. This requires mechanisms for sensing, signalling, adaptation, active transport, processing, *etc.* Amplification in signalling pathways, like MAPK, epidermal growth factor (EGF), transforming growth factor beta (TGF- $\beta$ ) and Integrin, constitute strong gain directions. Simultaneously, toxic substances must be maintained at low levels, inde-

pendent of environmental perturbations, thus their production must constitute weak gain directions. A good example is bacterial chemotaxis, where perfect adaptation is used to reset the response to a change in the nutrition concentration to its initial value. The bacterium can thereby maintain its sensitivity to concentration changes and track the nutrition gradient (Yi et al., 2000). In other words, attenuation of the signal generated by a prolonged change in the nutrition concentration enables strong amplification of small changes. From our perspective the biochemical control system is operating in closed-loop, containing intrinsic feedback loops designed by the evolutionary pressure. In Csete and Doyle (2004) the large scale organisation of metabolism is described as a bow-tie, where a wide range of nutrients are converted into a large variety of products using relatively few intermediates. In such a system, signals must be amplified and attenuated in a fashion dependent on the current needs of the cell, *i.e.* the system must be strongly interampatte so that it is able to channel energy through the bow-tie.

Based on the discussion above and the presented experimental evidence we postulate that biological networks are likely to have a high degree of interampattiness.

### 3.8 Conclusions

We have defined a generic system property—interampattiness—and postulated that biological networks are inherently interampatte. The degree of interampattiness is a measure of the ability of the system to simultaneously amplify and attenuate different signals. It is primarily caused by intrinsic feedback loops embedded in the network. The existence of biochemical subprocesses with different time-scales also makes a system interampatte. The main motivation for introducing interampattiness as a system property is that it can serve to explain earlier observations by other researchers on characteristic modes in response data from GRNs. Furthermore, as we have shown, it is a system property with significant implications for the inference of the network structure. An iterative approach to experiment design is needed to construct correlated perturbations, such that all modes of the system are sufficiently excited.

It has previously been observed by several authors (Holter et al., 2000; Alter et al., 2000; Kuruvilla et al., 2002; Wu and Dewey, 2006) that most of the variation in the measured response matrix of gene expression changes can be explained by a few variables. Signal patterns such as “characteristic modes” (Holter et al., 2000, 2001) and “eigengenes” (Alter et al., 2000; Nielsen et al., 2002; Alter and Golub, 2006; Omberg et al., 2007) have been introduced to explain the observation and utilize it for data reduction and feature prediction, but their opposite—the weak modes—have been completely neglected. However, both strong and weak modes are of importance for the very function and survival of organisms, and the concept of interampattiness assigns equal weight to all modes. Experimental evidence from a ten gene network of the Snf1 signalling pathway in *S. cerevisiae* (Lorenz et al., 2009), and a nine gene subnetwork of the SOS pathway in *E. coli* (Gardner et al.,

2003), as well as theoretical considerations support the postulation that biological networks are likely to have a high degree of interampattiness.

Interampatte systems will respond in a correlated fashion to a wide variety of perturbations and recorded response data will commonly be nearly rank deficient or ill-conditioned. The notion of co-expressed/co-regulated genes (Eisen et al., 1998; Spellman et al., 1998; Aoki et al., 2007) is therefore potentially also a consequence of interampattiness, and hence not necessarily caused by common transcription factors.

---

# Robust variable selection for network inference

---

*“Tortured data confess to anything.”*

Fredric M. Menger, J. Chem Soc. Perkin II Trans., 1988.

Inference or reverse engineering of gene regulatory networks from expression changes has the potential to reveal thousands of regulatory interactions from a series of perturbation experiments. The methods proposed for this purpose share two essential shortcomings. First, the inferred networks are largely determined by algorithm specific details, resulting in many false positives and negatives. Second, the employed measures of confidence of inferred interactions are in general misleading because they are based on the assumption that the correct network structure has been inferred. Herein we stress that network inference is a variable/model selection problem and prove that no reliable measure of confidence can be assigned to any inferred link for data sets with fewer observations than variables. By considering all possible linear models that can explain available data in the presence of uncertainty, we show that individual interactions can, based on data, be classified as existing, non-existing, non-evidential, and alternative. The data cannot be explained without links that are present in all models within the uncertainty set, so they are under mild assumptions true positives and we achieve truly robust inference by classifying only these as existing. These links can be determined without explicit determination of the model set by reformulating the variable selection as a robust rank problem for a set of uncertain data matrices. We present a method for solving the rank problem that enables us to assign confidence to individual interactions in a reliable way. We demonstrate the proposed method through an *in silico* example and by inferring a synthetic gene regulatory network from published *in vivo* data. In the latter case, the resulting model proves under mild assumptions, that a previously unknown activation of transcription of *SWI5* by *CBF1* exists in yeast. This serves to illustrate how the proposed robust inference method can reveal novel biological knowledge with confidence.

## 4.1 Introduction

Genes, proteins and metabolites do not function in isolation but rather interact within complex networks to generate biological functions and properties such as robustness. The insight that knowledge of the causal interactions that take place among the various molecular components of the cell is fundamental to understand biological functions and malfunctions has spurred the area of systems biology. A key problem in systems biology is the inference of causal interactions based on experimental observations.

Network inference, reverse engineering, or network reconstruction—“a dear child has many names” (Swedish proverb)—has attracted significant interest in the past decade, with more than 800 articles published and indexed in Pubmed ([www.pubmed.org](http://www.pubmed.org)). The invention of measurement techniques, such as real-time RT-PCR ([Higuchi et al., 1992](#)), DNA microarrays ([Schena et al., 1995](#)), chromatin immunoprecipitation (ChIP-chip) technique ([Ren et al., 2000](#); [Iyer et al., 2001](#); [Lieb et al., 2001](#)), and protein arrays ([MacBeath and Schreiber, 2000](#); [Zhu et al., 2001](#)), has enabled high-throughput quantification of gene expression changes and molecular interactions. This offers an effective means to establish causal interactions between a large number of molecules, or rather, to generate hypotheses on regulatory interactions which can be verified in additional perturbation experiments. Indeed, several studies on network inference claim to have demonstrated how the structure of, in particular, gene regulatory networks (GRNs) and transcriptional networks can be inferred based on gene expression data obtained from biological experiments in which the system is systematically perturbed and the resulting expression changes are measured, see *e.g.* [Lorenz et al. \(2009\)](#); [Cantone et al. \(2009\)](#); [Gardner et al. \(2003\)](#). By structure we here refer to the signed adjacency matrix of the graph of the network, *i.e.* a map of the direct causal influences that exist between the considered genes.

Inference problems in biology are often underdetermined due to the fact that the number of available observations typically is less than the number of variables. This again is explained by the large scale of intracellular networks combined with relatively costly experiments, and the problem may thus become less relevant as experimental and measurement techniques become more cost effective. Also, as discussed elsewhere in this thesis, it may often be more relevant to infer interactions between a smaller subset of genes which again reduces the problems related to underdeterminism. Nonetheless, underdetermined inference problems have received significant attention within bioinformatics and systems biology over the past decade. In order to determine a unique solution to a problem with less observations, or samples, than variables, some a priori knowledge or assumptions are required. Since gene regulatory networks typically are known to be sparsely connected, *i.e.* every gene interacts directly only with a few other genes, the focus in this area has been on methods for sparse inference aiming at identifying the biologically most relevant interactions. A large number of inference algorithms, based on ideas or methods from the fields of statistics, machine learning, signal processing, system identification, and related fields, have been proposed and tested on various data ([Emmert-Streib](#)

et al., 2012; Hecker et al., 2009; Gardner and Faith, 2005). Using these methods, a number of authors claim to have successfully inferred the underlying network structure also for underdetermined gene network problems, see *e.g.* Faith et al. (2007); Bonneau et al. (2006); Schäfer and Strimmer (2005) and the recent reviews Emmert-Streib et al. (2012); Hecker et al. (2009); Bonneau (2008); Tegnér and Björkegren (2007); Cho et al. (2007). In addition to considering the case with less observation than variables, several of these works do not utilize information about the applied perturbations. Furthermore, all sparse inference methods rely on regularization, variable selection, or thresholding techniques and they consequently yield different models for different choices of the regularization and thresholding parameters. This implies that, in particular for data sets with fewer observations than variables, the inferred network models largely are determined by algorithmic specific details such as parameter choices. We illustrate this below on an example involving inference of a regulatory network synthesized in yeast. Sensitivity to algorithmic details and noise has also been observed in *e.g.* Marbach et al. (2010); Tjärnberg et al. (2013).

While many authors claim success with their proposed algorithms, benchmarking studies have shown that inference of gene regulatory networks usually results in a large fraction of false positives, *i.e.* interactions identified that do not exist in the underlying network, and false negatives, *i.e.* interactions deemed absent that do exist in the underlying network (Marbach et al., 2010; Penfold and Wild, 2011; Bansal et al., 2007). Thus, how successful inference algorithms are is very much a relative concept. For instance, Lorenz et al. (2009) report successful inference of a gene regulatory network in budding yeast involving 10 genes despite considering 38% of the missing interactions as false negatives and 31% of the inferred interactions as false positives. In addition, 24% of the inferred interactions had the wrong sign, *e.g.* an interaction was identified as repressing while the true interaction was activating. Owing to the high uncertainty typically associated with every identified interaction, it is important to provide some label of confidence together with the inferred interactions, including those determined to be absent. Indeed, ranking of inferred interactions based on confidence measures has for instance been required to participate in the DREAM challenge for benchmarking of network inference methods (Marbach et al., 2010). The main problem with essentially all reported confidence measures, however, are that they are obtained in the context of a fitted model, implying that they are only valid for the selected model structure. If an inconsistent set of variables has been chosen, the confidence measures essentially only reflect the degree to which a specific variable explains the experimental data in the context of the fitted model.

In this work we stress that network inference is a variable selection problem, and that confidence measures should reflect the confidence with which a specific variable, *i.e.* interaction, has been correctly chosen or left out. As we show, no reliable measure of confidence can be assigned to any inferred interactions for data sets with fewer samples than variables. Furthermore, while most proposed algorithms consider the variable selection problem as an implicit part of the parameter estimation problem,

we propose that the two problems should be solved separately. We state the variable selection problem in the framework of solution of systems of linear equations and use algebraic and geometric approaches to derive conditions for when a given interaction should be included. An important feature of the proposed method is that the decision for a given variable is made independent of the decision made for the other variables in the viable set. In principle, the method amounts to searching over all models consistent with the data in the uncertainty set and determining interactions that are present or absent in all feasible models. While this is a deterministic approach to model uncertainty, also probabilistic descriptions can be incorporated by *e.g.* considering all realizations with a given cumulative probability. To select variables in the presence of uncertainty we formulate the problem as a robust rank problem which can be solved efficiently using tools from robust control theory. The size of the uncertainty set for which a variable is robustly selected then also serves as a confidence measure for the corresponding interaction.

We start the chapter by formulating the network inference problem as a system of linear equations and derive necessary and sufficient conditions for determining interactions that must be selected to explain available data in the absence of uncertainty. For illustration, we consider inference of the IRMA network that was recently engineered in *S. cerevisiae* by Cantone *et al.* (2009) for the very purpose of evaluating network inference algorithms. We then motivate why network inference should be considered a variable selection problem and discuss variable selection in the presence of uncertainty. For data sets with fewer observations than variables no reliable measure of confidence can be assigned to any inferred interaction, even though recent results have suggested that correct variable selection, parameter estimation, and network inference is possible in this case (Candès and Plan, 2009; Fan and Lv, 2010; Candès and Wakin, 2008; Zhao and Yu, 2006; Filkov, 2005). We prove this mathematically and also show that existing measures of confidence of inferred interactions are misleading by demonstrating that alternative models, lacking many of the links deemed most significant by Lorenz *et al.* (2009), can explain the available data equally well as the model they inferred. Finally, we present the method for robust variable selection and illustrate it on an *in silico* example and through application to the steady-state data recorded by Cantone *et al.* Unexpectedly, we find that the data cannot be explained without a previously unknown activation of transcription of *SWI5* by *CBF1* in the absence of Galactose. Existence of this interaction implies that the actual network differs from the network they intended to engineer.

The main purpose of this chapter is to present the principle idea of robust variable selection and most proofs are therefore omitted. A more comprehensive presentation, including a theory for robust network inference, is given in Chapter 5.

## 4.2 Problem formulation and introductory example

We here introduce and define the network inference problem, relate it to variable selection problems, and consider classification of interactions as existing, non-existing, alternative, and non-evidential. The 5-gene network called IRMA that was recently engineered in *S. cerevisiae* by Cantone et al. (2009) is used for illustration. The IRMA network was created specifically for evaluation of network inference algorithms and the small scale makes it relatively easy to comprehend.

The primary objective in network inference is to determine the subset of interactions that exists among a set of variables based on observed changes in the variables. In gene regulatory networks each variable corresponds to the abundance of mRNA of a gene of interest and the interactions describe regulatory influences between the genes. The inference problem, for several commonly used model formalisms, boils down to solution of a system of linear equations in close resemblance to linear regression

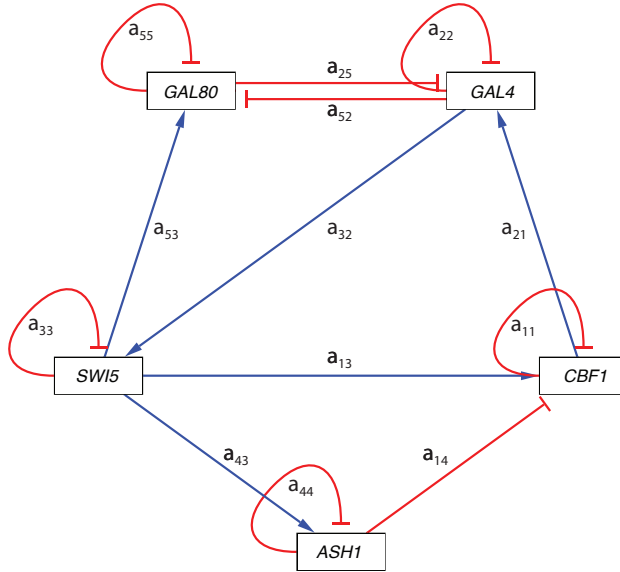
$$\Phi \mathbf{A} = \Xi, \quad (4.2.1)$$

see *e.g.* Bansal et al. (2006); Gardner et al. (2003); Yeung et al. (2002); Hecker et al. (2009); Karlebach and Shamir (2008); de Jong (2002). Here the regressor matrix  $\Phi$  contains observed responses to the perturbations in the regressand matrix  $\Xi$  and the objective boils down to determination of the structure of the interaction matrix  $\mathbf{A}$ . Each element  $a_{ij}$  corresponds to a direct causal influence from gene  $j$  to gene  $i$ . The inference problem corresponds to  $n \times m$  equations in  $n \times n$  unknowns in  $\mathbf{A}$ , and hence we need at least  $m = n$  experiments to have a unique solution. The fact that one often is faced with less data in inference of gene regulatory networks, *i.e.*  $m < n$ , implies that the problem is underdetermined and additional information needs to be added to make the solution unique. As discussed in the introduction, a frequent assumption is that the genes are sparsely connected and based on this various sparse inference algorithms have been employed. However, as demonstrated for the IRMA example below, fulfilling some sparsity conditions is of course no guarantee for inferring the correct network even though the latter in fact is sparse. The aim of the work presented here is therefore to determine what interactions can be determined with confidence from data with limited and uncertain information, rather than focusing on inferring a complete network.

Before proceeding with the IRMA example, note that the overall inference problem in (4.2.1) can be split into  $n$  separate subproblems, one for each row of the interaction matrix  $\mathbf{A}$  to obtain

$$\Phi \theta = \xi \quad (4.2.2)$$

The columns of  $\Phi$  are here called the *regressors*,  $\xi$  the *regressand*, and the inference problem corresponds to finding a linear combination of the former that can explain the latter. For an introduction to linear regression, from a parameter estimation perspective, see *e.g.* Rao and Toutenburg (1999); Bingham and Fry (2010); Jaqaman and Danuser (2006).



**Figure 4.1: The engineered IRMA network.** The structure of the network with the corresponding element of the interaction matrix  $\mathbf{A}$  marked on each edge. An arrow shaped head on the link indicates up-regulation or activation, while a bar shaped head indicates down-regulation or repression. This network graph was generated in Cytoscape (Shannon et al., 2003).

The IRMA network is a 5-gene network engineered in *S. cerevisiae* by Cantone et al. (2009). The engineered structure of the network is

$$\check{\mathbf{A}} = \begin{bmatrix} -1 & 0 & 1 & -1 & 0 \\ 1 & -1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & -1 & 1 & 0 & -1 \end{bmatrix}, \quad (4.2.3)$$

which is illustrated in Figure 4.1. We consider unit strength of all interactions here because no strengths were specified for the engineered structure and it simplifies this introductory example. For the illustration here it is sufficient to consider inference of the active interactions in the third row of the interaction matrix, corresponding to edges affecting the 3rd gene (*SWI5*), based on perturbation of the 2nd, 3rd and 5th gene (*GAL4*, *SWI5*, *GAL80*) in three independent transcriptional perturbation experiments and observed steady-state responses in the expression of all 5 genes. We perform these *in silico* perturbation experiments on the engineered interaction

matrix and obtain the following data set

$$\check{\Phi} = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 \\ 0 & -1 & 0 & 0 & 1 \\ 0 & -1 & -1 & -1 & 1 \end{bmatrix}, \quad \check{\xi}_3 = \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}, \quad (4.2.4)$$

with the first column  $\check{\phi}_1$  corresponding to expression changes in the 1st gene (*CBF1*), the second column  $\check{\phi}_2$  to the 2nd gene (*GAL4*), and so forth, while the first row corresponds to the experiment in which the 2nd gene (*GAL4*) is perturbed, the second row to the experiment in which the 3rd gene (*SWI5*) is perturbed, and the third row to the last experiment with perturbation of the 5th gene (*GAL80*). The rows of  $\check{\xi}_3$  contain the perturbation of the 3rd gene (*SWI5*) in each experiment, but with opposite sign. Note that we here assume noise-free variables. Later on, noisy measurement data recorded by Cantone *et al.* from corresponding *in vivo* experiments will be used. Further details are given in Section 2.3 on modelling of GRNs and in Section 2.1 on the engineered IRMA example.

We next perform network inference based on the regressor matrix and regressand assuming that we do not know the engineered interaction matrix. The data model for the 3rd row of the interaction matrix,

$$\sum_{j=1}^5 \check{\phi}_j a_{3j} = \check{\xi}_3, \quad (4.2.5)$$

connects the regressors to the regressand through the unknown parameters  $a_{3j}$ . As mentioned above, no unique solution exists since  $m < n$  and there are more parameters than equations. Utilizing the knowledge that the network is sparse, *i.e.* that some parameters are identically zero, implies that we should perform variable selection, *i.e.* selecting which parameters to include in the model. Most approaches to sparse inference consider variable selection as an inherent part of the parameter estimation method, and typically employ minimization of some norm of the fitted parameter vector.

Inference of the third row boils down to selection of a subset  $\mathcal{S}$  of the regressors, such that this system of linear equations has a solution, *i.e.*

$$\sum_{j \in \mathcal{S}} \check{\phi}_j a_{3j} = \check{\xi}_3. \quad (4.2.6)$$

If only one such subset exist, then the selection is unique and identical to the “true” selection  $\check{\mathcal{S}} = \{2, 3\}$ . This is however not the case for partly informative data and makes it interesting to ask if any of the regressors must be selected or excluded, as the corresponding interactions are true positives and negatives, respectively. We call a data set partly informative if it contains some information, but is not informative enough for inference of all interactions that exist in the “true” network.

From linear algebra it is known that a system of equations only has a solution if the regressand lies in the subspace spanned by the regressors and that the solution only is unique if the regressors are linearly independent, see *e.g.* Anton and Rorres (2000); Friedberg et al. (2003); Horn and Johnson (1990). Here the regressand lies in the subspace spanned by the regressors, since

$$\check{\xi}_3 \in \text{span } \check{\Phi} = \left\{ \sum_{j=1}^5 \theta_j \check{\phi}_j \mid \theta_j \in \mathbb{R} \right\} = \mathbb{R}^3, \quad (4.2.7)$$

but they are linearly dependent, since

$$\sum_{j=1}^5 \theta_j \check{\phi}_j = \mathbf{0} \quad \text{with some } \theta_j \neq 0. \quad (4.2.8)$$

Hence many solutions of (4.2.5) and alternative selections  $\mathcal{S}$  exist. To answer which variables that must be in all sets, and which are in none, we can make a simple geometric argument. If some regressor  $\check{\phi}_j$  spans a unique direction, *i.e.* can not be expressed as a linear combination of the other regressors, and this direction is present in the regressand  $\check{\xi}_3$  then the regressor  $\check{\phi}_j$  is needed to explain  $\check{\xi}_3$ . Hence, in this case  $a_{3j}$  must be selected, *i.e.* be in all alternative selections  $\mathcal{S}$ . Following the same line of arguments,  $a_{3j}$  must be deselected, *i.e.* not be in any alternative selections  $\mathcal{S}$ , if  $\check{\phi}_j$  spans a unique direction and this direction is not present in  $\check{\xi}_3$ . Based on this we find from the partly informative data for the IRMA example that the 2nd regressor must be included in every set while the 5th regressor must be excluded in all sets. Also, either the 3rd or 4th regressor must be included in every set of regressors, while the 1st regressor is identically zero. In other words,  $\check{\phi}_2$  and  $\check{\phi}_3$  or  $\check{\phi}_4$  must be selected, while  $\check{\phi}_5$  should be excluded and  $\check{\phi}_1$  always can be excluded. We therefore classify the interaction  $a_{32}$  as existing,  $a_{35}$  as non-existing,  $a_{33}$  and  $a_{34}$  as alternative, and  $a_{31}$  as non-evidential. The data is not informative enough for discrimination between models containing  $a_{33}$  and  $a_{34}$ , because the set consisting of only  $\check{\phi}_3$  and  $\check{\phi}_4$  is linearly dependent, *i.e.* these regressors are collinear. Selection of  $\check{\phi}_1$ , on the other hand, never contribute nor harm explanation of the regressand, *i.e.* it does not improve nor decrease the fit; the data therefore contains no evidence about  $a_{31}$  and one should based on the principle of parsimony assume  $a_{31} = 0$ . In summary, one can in general based on partly informative noise-free data classify the possible network interactions into four groups, as shown in Nordling and Jacobsen (2011):

(a) existing edges (true positives)  
 if  $(\mathbf{I} - \mathbf{T}_{j \neq k}) \check{\phi}_k \neq \mathbf{0}$  and  $(\mathbf{I} - \mathbf{T}_{j \neq k}) \check{\xi} \neq \mathbf{0}$ , (4.2.9a)

(b) non-existing edges (true negatives)  
 if  $(\mathbf{I} - \mathbf{T}_{j \neq k}) \check{\phi}_k \neq \mathbf{0}$  and  $(\mathbf{I} - \mathbf{T}_{j \neq k}) \check{\xi} = \mathbf{0}$ , (4.2.9b)

$$(c) \text{ non-evidential edges (no information)} \\ \text{if } \check{\phi}_k = \mathbf{0}. \quad (4.2.9c)$$

$$(d) \text{ alternative edges (partial information)} \\ \text{if } (\mathbf{I} - \mathbf{T}_{j \neq k})\check{\phi}_k = \mathbf{0}, \quad (4.2.9d)$$

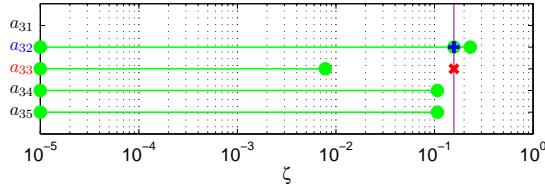
Here the projection matrix  $\mathbf{T}_{j \neq k} = \mathbf{\Phi}_{j \neq k} \mathbf{\Phi}_{j \neq k}^\dagger$  projects any vector in  $\mathbb{R}^m$  onto the subspace spanned by the  $n - 1$  regressors in the matrix  $\mathbf{\Phi}_{j \neq k} = [\check{\phi}_1, \dots, \check{\phi}_{k-1}, \check{\phi}_{k+1}, \dots, \check{\phi}_n]$ , and  $\mathbf{\Phi}_{j \neq k}^\dagger$  denotes the Moore-Penrose generalized inverse of the matrix (Friedberg et al., 2003, p. 398-400). This classification may seem trivial, but is important as it provides knowledge on which interactions that can be inferred with certainty based on partly informative data. It also distinguishes between links that should be left out based on information and lack of information, respectively.

Our *in silico* data above only has one of the two common characteristics of biological data, namely fewer observations than variables, *i.e.*  $m < n$ . To consider the effect of the other characteristic, uncertainty or noise, we next add uncertainty to the IRMA data and consider how this hampers network inference in the sense of variable selection. For illustration, we introduce a simple deterministic error model in the measurement data and consider the observed regressor matrix

$$\mathbf{\Phi} = \begin{bmatrix} 0 & 1 & 1+v & 1 & 0 \\ 0 & -1 & 0 & 0 & 1 \\ 0 & -1 & -1 & -1 & 1 \end{bmatrix}. \quad (4.2.10)$$

Here  $v \neq 0$  denotes a small measurement error in a single element with bounded magnitude  $|v| \leq 0.1$ , *i.e.* all other elements are still assumed noise-free. This error model is deterministic because no probability distribution is given. Any subset containing at least three of the observed regressors 2 to 5 now spans  $\mathbb{R}^3$  and they are therefore classified as alternative based on the rules for noise-free data above. Regressors 2 and 5 hence no longer span unique directions, implying that we cannot say that  $a_{32}$  exists and  $a_{35}$  does not exist. As a matter of fact, selection based on this data set is sensitive to infinitesimal small errors in any element, and so illustrates how uncertainty hampers variable selection for network inference. Note that the corresponding estimation problem is underdetermined and averaging can therefore not be used to reduce the effect of the noise, as customary in statistics, system identification, signal processing, and econometrics.

If we do a brute force search for the sparsest solution, *i.e.* we make the common assumption that the solution with fewest nonzero elements is correct, under the wrong assumption of (4.2.10) being noise-free, then we find  $a_{32} = 1$ ,  $a_{34} = -1$ , and all other elements zero, *i.e.* we obtain the wrong solution and selection. If we instead do the same search under the correct assumption that the 3rd regressor is uncertain and lies in the set  $\mathcal{U}_{\phi_3} = \{\check{\phi}_3 | \check{\phi}_{31} = \phi_{31} - v, \check{\phi}_{32} = \phi_{32}, \check{\phi}_{33} = \phi_{33}, |v| \leq 0.1\}$ , then we find two alternative solutions,  $a_{32} = 1$  and  $a_{34} = -1$ , or  $a_{32} = 1$  and  $a_{33} = -1$ , with all other elements zero. This illustrates that we for partly informative and



**Figure 4.2: Dependence on the regularization coefficient**, illustrated by inference of the 3rd row of the engineered IRMA network in (4.2.3). The solid lines cover the  $\zeta$  range for which the interaction  $a_{3j}$ , marked to the left, is present in the inferred network model, corresponding to selection of regressor  $\phi_j$ . The “true” interactions are marked by + (blue) for up-regulations and x (red) for down-regulations. The vertical solid (purple) line marks one of the inferred networks that are most similar to the “true” one based on the signed adjacency matrix.

uncertain data in general obtain several selections that are indistinguishable based on the data, *i.e.* we need additional data or assumptions.

If we instead employ the ordinary least squares estimator, see *e.g.* Rao and Toutenburg (1999, p. 24-27), then we find a different solution for each  $\tilde{\phi}_3 \in \mathcal{U}_{\phi_3}$ , but most of them correspond to the selection  $\mathcal{S}_{\text{ols}} = \{2, 3, 4, 5\}$ . Note that the ordinary least squares estimator in general does not give sparse solutions, so we also use Glmnet that incorporates an  $l_1$  penalty on the parameter vector as described above (Friedman et al., 2010)

$$a_{ij} = \arg \min_{a_{ij}, a_{i0} \in \mathbb{R}} \frac{1}{2m} \left\| \xi_i - \sum_{j=1}^n \phi_j a_{ij} - a_{i0} \mathbb{1} \right\|_2^2 + \zeta \sum_{j=1}^n |a_{ij}|, \quad (4.2.11)$$

Glmnet is an implementation of the popular LASSO regularization, which penalizes models with small nonzero parameters (Tibshirani, 1996). With Glmnet the resulting variable selection depends on the regularization coefficient  $\zeta$ , as shown in Figure 4.2, as well as the actual noise realization  $v$ . The noise realization, however, essentially only affects the  $\zeta$ -value at which  $\phi_3$  is selected (data not shown). The results in Figure 4.2 serve to show that if we infer a single model from partly informative and uncertain data, then the noise realization together with algorithm specific details determine the variable selection, since as shown above, even for the noise-free data is it impossible to distinguish between  $\mathcal{S}_a = \{2, 3\}$  and  $\mathcal{S}_b = \{2, 4\}$  based on the data as shown above. This simple example thus, in part, explains why inference of gene regulatory networks has been shown to usually result in a large fraction of false positives and false negatives in benchmarking studies, such as the DREAM challenge (Marbach et al., 2010; Stolovitzky et al., 2009).

In conclusion, the introductory example has demonstrated how network interactions can be classified as existing, non-existing, alternative and non-evidential for noise-free data. Real data however contains uncertainty, or noise, which hampers

the inference, and below we therefore establish a method for doing robust variable selection for network inference based on partly informative and uncertain data and derive necessary conditions for assigning confidence to the existence of individual interactions.

### 4.3 Network inference is variable selection

From our introductory example it should be clear that the essence of the network inference problem is the selection of the active interactions based on experimental data, which is a problem known as feature selection, feature reduction, attribute selection, subset selection, variable selection, and model selection in the machine learning, statistics, and system identification literature (Guyon and Elisseeff, 2003; Fan and Lv, 2010; Hara and Sillanp, 2009; George, 2000; Hong et al., 2008; Stoica and Selen, 2004). Here we will mainly use the term variable selection, since our features correspond to observed variables, represented by the regressors in (4.2.5). Many variable selection methods require estimation of model parameters and evaluation of the resulting fit to data, *e.g.* use of the Akaike information criterion and Bayesian information criterion (Cedersund and Roll, 2009; Stoica and Selen, 2004; Akaike, 1973; Schwarz, 1978), and others implicitly select the variables as part of the parameter estimation, *e.g.* LASSO (Tibshirani, 1996; Candès and Plan, 2009). In most algorithms employed for inference of gene regulatory networks variable selection and estimation of model parameters are implicit and combined, see *e.g.* Glmnet (4.2.11). When using such implicit methods, parametric hypothesis tests like the t-test employed in Lorenz et al. (2009), can not be used to determine the significance/confidence level of the inferred interactions although this is often assumed in the literature.

The distinction between variable selection and parameter estimation is usually not clear in the literature on network inference, even though obviously essential when inferring gene regulatory interactions. We believe there are at least three important reasons why network inference should be considered foremost a variable selection problem, while the parameter estimation problem should follow the variable selection. First, parameter estimates as well as the estimates of their variance and significance are only valid for a selected model structure and set of variables (Burnham and Anderson, 2002, p. 14). For instance, the t-test is based on the ratio between estimates of the parameters and their variance and if the erroneous set of variables have been chosen the resulting confidences provide no information on the confidence of the inferred interactions whatsoever. Thus, a correct variable selection must be made before the estimated parameters and confidences make any sense. We demonstrate below the consequences of erroneous variable selection by considering the gene response data recorded by Lorenz et al. (2009). Second, only parameters corresponding to active interactions, *i.e.* interactions that takes place in the biological system when it exhibits the behaviour of interest, are nonzero and need to be estimated. The parameter estimator should have the correct number of degrees of

freedom, *i.e.* number of parameters, in order to efficiently use the data, since *e.g.* the total parameter variance grows with the number of degrees of freedom (Kay, 1993, p. 34,42-43). Third, if the number of parameters exceeds the number of data points, then some parameters are structurally/*a priori* unidentifiable and no unique solution exists to the parameter estimation problem (Ashyraliyev et al., 2009; Fallor et al., 2003; Bellman and Åström, 1970). The distinction between variable selection and parameter estimation is particularly important in gene regulatory network inference, since the networks are assumed to be sparse and the number of possible interactions, *i.e.* unknown parameters, often by far exceeds the number of data points. This renders any formulation as a parameter estimation problem underdetermined, which in general for uncertain data leads to a set of models with different structure that can explain the data, *i.e.* different selections of regressors as demonstrated in the introductory example.

In summary, only interactions corresponding to variables that must be selected to explain available data can be shown to exist at a selected significance level, so for robust network inference it is both necessary and sufficient to do robust variable selection prior to parameter estimation. The remainder of this chapter builds on this observation.

#### 4.4 As many samples as variables are needed to infer sparse networks

We here show that no confidence can be assigned to individual inferred interactions when the available data contains fewer observations than variables by showing that alternative models with a different structure that can explain the data always exist. To demonstrate this, we show that alternative models exist that explain the data recorded by Lorenz et al. (2009) equally well as the model they inferred, despite lacking many of the links found to be the most significant by Lorenz et al.

Consider first the introductory IRMA example and the data in (4.2.4) again and assume that each observation contains uncertainty, measurement errors, or noise of some sort, denoted by  $v_{kj}$  and  $\epsilon_{ki}$  for the responses and perturbations, respectively.

$$\Phi = \begin{bmatrix} 0 + v_{11} & 1 + v_{12} & 1 + v_{13} & 1 + v_{14} & 0 + v_{15} \\ 0 + v_{21} & -1 + v_{22} & 0 + v_{23} & 0 + v_{24} & 1 + v_{25} \\ 0 + v_{31} & -1 + v_{32} & -1 + v_{33} & -1 + v_{34} & 1 + v_{35} \end{bmatrix} \quad \xi_3 = \begin{bmatrix} 0 + \epsilon_{13} \\ -1 + \epsilon_{23} \\ 0 + \epsilon_{33} \end{bmatrix}. \quad (4.4.1)$$

By assigning arbitrarily small values to the coefficients representing uncertainty we can ensure that any subset consisting of three regressors spans the data space, whose dimension is equal to three, and any one of the five regressors can therefore be left out of the model used to explain the data, implying that many alternative models with different structure always exist. Indeed, it is trivial to prove that random errors almost always imply that any combination of  $m$  regressors spans a space of

dimension  $m$  in observed data, since specific combinations of values are required to make a subset of regressors collinear or linearly dependent.

The observation above holds in general, *i.e.* for every possible interaction there always exist an alternative network model excluding it that can explain noisy data if the number of observations  $m$  is smaller than the number of variables  $n$ . This implies that no confidence can be assigned to any inferred interaction, as formally proved in Theorem A.3.2. In other words,  $m \geq n$  is a necessary condition on data to assign any reliable measure of confidence to individual interactions. This should come as no surprise, and simply corresponds to the well known fact that at least  $n$  equations are needed to uniquely determine  $n$  unknown parameters in a linear system of equations. Everyone familiar with recent work on network inference and regularization, however, now probably has two objections in mind which we address next.

First, the observation above appears to be in conflict with recent observations and results suggesting that less than  $n$  observations is sufficient for correct variable selection, estimation of a parameter vector, or inference of a network, see *e.g.* Candès and Plan (2009); Fan and Lv (2010); Candès and Wakin (2008); Zhao and Yu (2006); Filkov (2005). The apparent discrepancy between our observation and these observations is explained by the fact that there is a fundamental difference between actually inferring the correct structure of a network on the one hand, and on the other hand proving that it with some chosen level of confidence is the correct structure. For instance, even results based on the strong irreducibility condition (SIC) only ensures that the LASSO estimator

$$a_{ij} = \arg \min_{a_{ij} \in \mathbb{R}} \left\| \xi_i - \sum_j a_{ij} \phi_j \right\|_2^2 + \zeta \sum_j |a_{ij}| \quad (4.4.2)$$

is sign consistent with a probability that goes to one as the number of samples goes to infinity (Zhao and Yu, 2006). Some of the inferred interactions could thus be false positives, in particular for the low number of samples seen in biological data sets. Considering that from many studies, such as the DREAM challenges (Marbach et al., 2010), aimed at benchmarking different inference algorithms it is evident that inference of GRNs usually result in a large fraction of false positives and negatives (Stolovitzky et al., 2009), we claim that knowledge of the confidence level of each inferred interaction is as important as inferring a model in the first place.

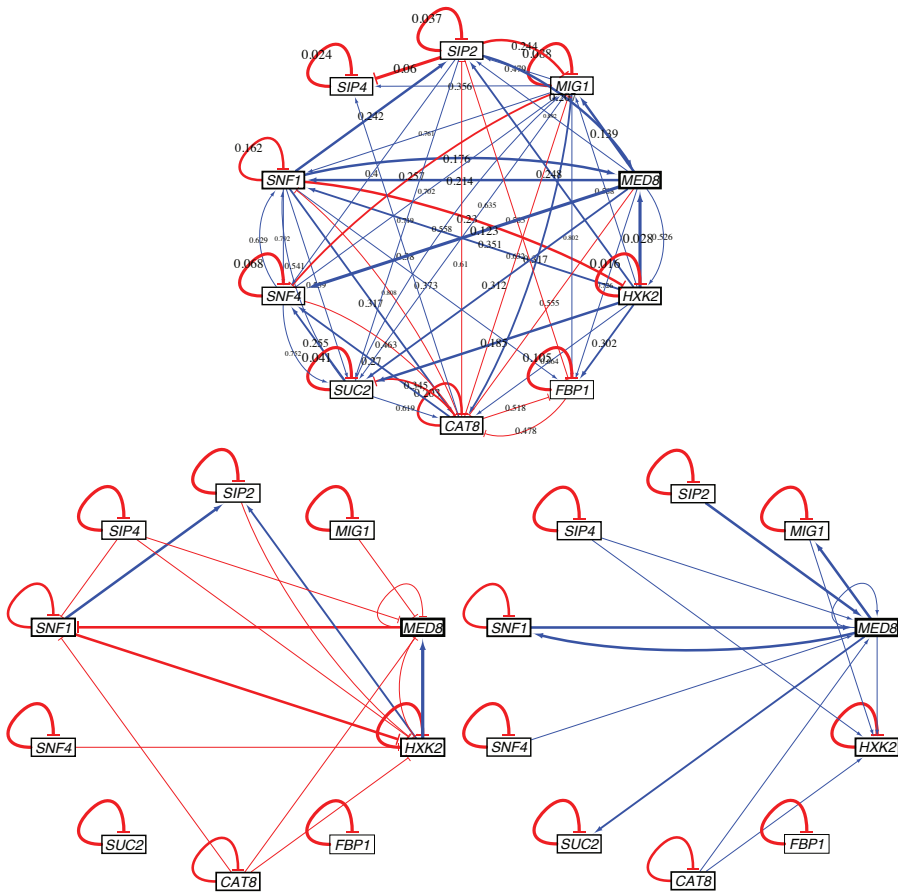
Second, significance levels have indeed been reported for individual interactions (Lorenz et al., 2009) and ranking of the inferred interactions based on confidence has in fact been required to participate in the DREAM challenge (Marbach et al., 2010), so clearly it is possible to assign measures of confidence to individual interactions independent of the number of observations. However, as discussed above, those significance measures and rankings are in general not reliable, simply because they fail to account for alternative selections of the variables, *i.e.* models with alternative structures, that can explain the data. That is, the computed confidence of each chosen variable is based on the assumption that all other variables have

been correctly selected.

To illustrate the consequences of the above discussion, we consider the steady-state data recorded from perturbation experiments in yeast by [Lorenz et al. \(2009\)](#), in which each of the ten *Snf1* signalling pathway genes in *S. cerevisiae* were over-expressed in independent experiments and the resulting expression changes measured, and show that the applied two tailed t-test does not provide reliable estimates of the confidence of the inferred interactions. Since the number of independent experiments and observations in this case equals the number of variables, this example also illustrates that alternative models may exist and confidence tests may be unreliable despite fulfilment of our necessary condition, *i.e.*  $m \geq n$ .

In [Figure 4.3](#) we show the final network model inferred by [Lorenz et al. \(2009\)](#) (top), taken from their supplementary Table S9, and two alternative models (bottom), selected from the large set of models that can explain their data. The two latter models were chosen partly because they are sparse. An interesting observation is that 51 out of the 60 links deemed most significant by [Lorenz et al.](#) are missing or have opposite sign (2 cases) in at least one of the two alternative models. Note that only 3 of the 10 most significant links are missing in the alternative models, since 7 self-loops are among the top 10. Self-loops represent degradation of the mRNA and correspond to parameters on the diagonal of the interaction matrix, which makes them abundant in models and biologically uninteresting. Both alternative models fit the data in Lorenz well when judged based on commonly used measures such as the unweighted and weighted residual sum of squares ([Ashyraliyev et al., 2009](#)), prediction error ([Ljung, 1999](#)), and cannot be rejected based on a  $\chi^2$  goodness of fit test ([Jaqaman and Danuser, 2006](#); [Cedersund and Roll, 2009](#)) at significance level 0.5. By testing at significance level 0.5 we ensure that they cannot be rejected at any commonly used significance level, such as 0.05. For details see [Section C.1](#). Existence of alternative models with different structure is a characteristic of partly informative data and implies that the data does not contain sufficient information to determine if certain interactions exist or not. In this case the lack of information is due to poor excitation of the weak directions of the underlying system, which we earlier have shown to be interampatte, see [Chapter 3](#). The t-test is but one parametric hypothesis test that fails to account for the existence of alternative variable selections, since the parameters and their variance estimates are calculated under the implicit assumption that the variable selection is correct. Any hypothesis test or confidence measure that makes this assumption will be misleading in particular for data with fewer observations than variables. This partly explains why false positives exists among the inferred interactions deemed most significant even by the best performing methods in the DREAM challenge ([Marbach et al., 2010](#)).

The fact that existing confidence measures can be misleading calls for caution when dealing with inferred networks. We next propose a method for determining the interactions that are present in all alternative models that can explain available data at a selected confidence level. The proposed method can also be used to calculate the confidence score of every possible interaction in a manner that correctly accounts for the existence of alternative variable selections.



**Figure 4.3: Alternative models lacking links deemed most significant by Lorenz et al. (2009) that fit well to their *in vivo* data, based on commonly used measures such as the unweighted and weighted residual sum of squares (Ashyraliyev et al., 2009), prediction error (Ljung, 1999), and cannot be rejected based on a  $\chi^2$  goodness of fit test (Jaqaman and Danuser, 2006; Cedersund and Roll, 2009). Here we show the final model inferred by Lorenz et al. (top) and two alternative models (bottom), selected from the large set of models that can explain their data because they are sparse and 51 out of the 60 links inferred by Lorenz et al. are missing or even have opposite sign (2 cases) in at least one of them. Note that these models are not intended to represent current biological knowledge. Here every link represents an inferred causal influence (Gardner and Faith, 2005). An arrow shaped head on the link from *SNF1* to *SIP2* (top) indicates that *SNF1* upregulates (activates) the *SIP2* gene, while a bar (T) shaped head from *SIP2* to *SIP4* indicates that *SIP2* downregulates (represses) the *SIP4* gene. The thickness of each link decreases with increasing p-value as calculated by Lorenz et al. based on a two tailed t-test for the links they inferred, here reported on each link in their model, *i.e.* the thickest link is the most significant one in their model. These network graphs were generated in Cytoscape (Shannon et al., 2003).**

## 4.5 Inference of existing interactions with confidence

Here we propose a method for robust variable selection and network inference, based on reformulating the inference problem as a robust rank problem. We define a measure of robust rank in terms of a distance to rank deficiency under structured uncertainty, show how it can be computed using tools from robust control theory and apply it to inference of the IRMA network considered in the introductory example. Note that the focus in this section mainly is on presenting the principle idea and main results on robust variable selection, while formal proofs together with a more complete theory for robust variable selection are given in Chapter 5.

In the introductory example, we demonstrated that an interaction can be shown to exist based on noise-free data if selection of the corresponding regressor is necessary to explain the regressand in (4.2.6). This is equivalent to requiring that the corresponding regressor  $\phi_k$  and regressand  $\xi_i$  cannot be written as a linear combination of the other regressors, *i.e.* condition (4.2.9a). This corresponds to linear independence of the set of columns in the two matrices

$$\Phi \triangleq [\phi_1, \dots, \phi_k, \dots, \phi_n] ; \quad \Psi_{ki} \triangleq [\phi_1, \dots, \phi_{k-1}, \phi_{k+1}, \dots, \phi_n, \xi_i]. \quad (4.5.1)$$

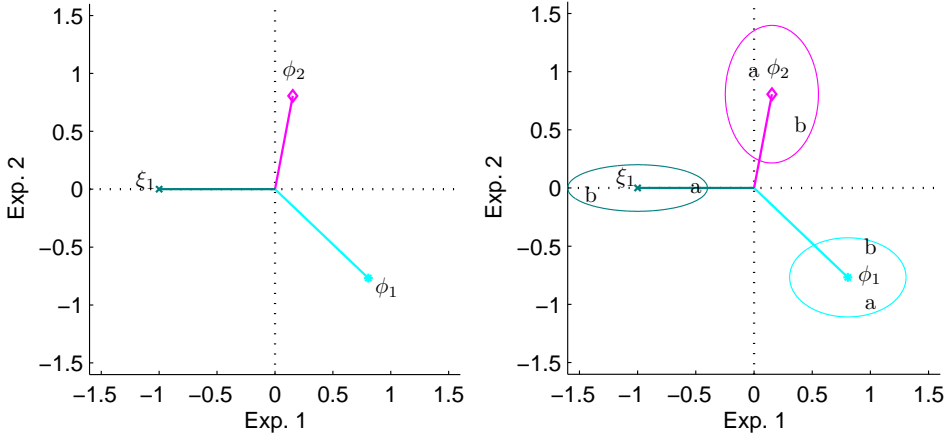
Full column rank of these two matrices is hence sufficient to conclude existence of the  $a_{ik}$  interaction corresponding to regressor  $\phi_k$  when the data are noise-free. Real biological data, however, contains measurement errors and noise that introduce uncertainty in each data point (4.4.1). To describe this uncertainty an error model must be added to the considered data model. Here we model the uncertainty by assuming that the “true” response, or perturbation, lies within a norm-bounded set centred at the nominal measurement. Mathematically, we describe this by adding a norm-bounded perturbation to each regressor and regressand

$$\phi_j = \check{\phi}_j + \mathbf{v}_j, \quad \|\mathbf{v}_j\| \leq \varrho_j \quad (4.5.2a)$$

$$\xi_i = \check{\xi}_i + \epsilon_i, \quad \|\epsilon_i\| < \varpi_i. \quad (4.5.2b)$$

These bounded uncertainty sets are schematically illustrated in Figure 4.4 (right) by ellipses around the nominal regressors and regressand in a two dimensional data space. In higher dimensions, the uncertainty sets correspond to hyperellipsoids around the measurement points in the  $m$ -dimensional data space. Note that (4.5.2) as such give hyperspherical uncertainty sets, and that hyperellipsoids are obtained by scaling the individual measurements points in  $\phi_j$  and  $\xi_i$  accordingly.

Each combination of realizations within the uncertainty sets that fulfils the assumed data model gives rise to a model that cannot be invalidated or rejected based on the available data, implying that we in general have an infinite set of practically indistinguishable models that are all in agreement with the available response data. Thus, the set of data points within the uncertain data model is translated into a set of network models. We have in Figure 4.4 (right) marked two realizations, a and b, that give rise to two different models  $\mathcal{M}_a$  and  $\mathcal{M}_b$ , in this case containing the same interactions. While all models within the uncertainty set



**Figure 4.4: Graphical representation of variables and their uncertainty sets for an example with two regressors in a two dimensional data space.** Observed data is represented by plotting the value in each experiment on a different coordinate axis and viewing these points as vectors in what we call the data space (left). The regressors are depicted by the solid cyan or magenta coloured vectors ending by a star or diamond and tagged by  $\phi_1$  or  $\phi_2$ , while the solid teal coloured vector ending by an x and tagged by  $\xi_1$  depicts the regressand. The uncertainty set of each regressor  $\phi_j$  and regressand  $\xi_1$  is depicted by the solid coloured ellipses drawn around the tip of each vector (right). The three realizations corresponding to the feasible model  $\mathcal{M}_a = \{-0.5\tilde{\phi}_1^a - 0.5\tilde{\phi}_2^a = \tilde{\xi}_1^a\}$  and  $\mathcal{M}_b = \{-\tilde{\phi}_1^b - \tilde{\phi}_2^b = \tilde{\xi}_1^b\}$  are marked by a and b (right), respectively. Robust variable selection is for this case illustrated in Figure 4.5.

are good for prediction of the response to perturbation experiments similar to the ones from which the data were obtained, simply because they cannot be rejected, they may include interactions not present (false positives) or leave out interactions present (false negatives) in the true network as illustrated in Figure 4.3. This again highlights the fact that one in general should distinguish between identification of predictive models and inference of mechanistic models.

If a deterministic uncertainty model is employed, then the uncertainty sets will by assumption contain the “true” realizations, and hence interactions that are present in all models within the set correspond to true positives that must exist in the underlying GRN. This means that robust variable selection and network inference are achieved by determining the regressors that are selected in all of these models and thus necessary to explain the data, as well as the corresponding interactions that are also present in all of these models. Note that bounded uncertainty sets that contain the “true” realizations with a probability greater or equal to  $1 - \alpha$  can in general be generated for probabilistic error models at any selected significance level  $\alpha > 0$ , as we demonstrate in Section 5.4, and so the method proposed here is not

limited to deterministic error models.

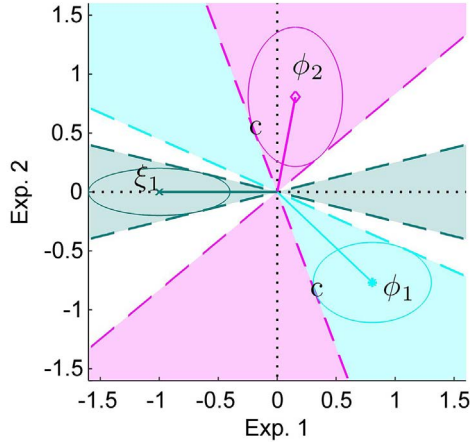
In principle, we could show that an interaction exists based on noisy data by showing that it is present in all models that cannot be rejected based on the data. Similarly, we could show that an interaction is absent if it is not included in any of the models within the set. This however involves checking an infinite set, which is not computationally feasible, so we next reformulate our robust variable selection problem as a robust rank problem, *i.e.* using a measure of distance to rank deficiency under structured uncertainty for the observed  $\Psi_{ki}$  matrix. If the set of columns in  $\Psi_{ki}$  are linearly independent, or equivalently the  $\text{rank}(\Psi_{ki}) = n$ , for all realizations that are consistent with the error model, then the regressor  $\phi_k$  must be selected since none of the regressand realizations can be written as a linear combination of any realization of the other regressors alone. This is a sufficient condition for showing that the regressor  $\phi_k$  must be selected and is the robust analogous to the conditions for variable selection with noise-free data above. The condition on linear independence of the set of columns in  $\Phi$  is not needed, since no solution exists to the equivalence of (4.2.6) for any set of linearly dependent realizations. This is illustrated in Figure 4.5 where both regressors must be selected to explain the regressand. Only if a realization such that  $\Psi_{ki}$  is rank deficient exists within the uncertainty sets is it possible to exclude  $\phi_k$  and find a model in which the regressand is a linear combination of the other regressors. We can hence show that  $\phi_k$  is selected in all practically indistinguishable models by showing that no  $\Psi_{ki}$  within (4.5.2) is rank deficient.

To check whether a rank deficient  $\Psi_{ki}$  exists within the uncertainty set given by (4.5.2), we employ the structured singular value  $\mu$  (Doyle, 1982; Safonov, 1982; Zhou and Doyle, 1998, p. 189). and define the confidence score  $\gamma(\Psi_{ki})$  as the reciprocal of  $\mu$

$$\gamma(\Psi_{ki}) \triangleq \frac{1}{\mu(\Psi_{ki})} \triangleq \min_{\Delta^s} \{ \bar{\sigma}(\Delta^s) \mid \det(\mathbf{I} - \Psi_{ki}^\dagger \mathbf{S}_{ki} \Delta^s) = 0 \text{ for structured } \Delta^s \}. \quad (4.5.3)$$

Here  $\Delta^s$  is an  $mn \times n$  block-diagonal matrix with  $n$  blocks of dimension  $m \times 1$ , representing the uncertainties in the columns of  $\Psi_{ki}$ , and  $\mathbf{S}_{ki} = [\varrho_1 \mathbf{I}_{m \times m}, \dots, \varrho_{k-1} \mathbf{I}_{m \times m}, \varrho_{k+1} \mathbf{I}_{m \times m}, \dots, \varrho_n \mathbf{I}_{m \times m}, \varpi_i \mathbf{I}_{m \times m}]$  is a scaling matrix taking into account the size of the uncertainty of the individual regressors and regressand in  $\Psi_{ki}$ . If the structured singular value  $\mu(\Psi_{ki})$  is larger than 1, corresponding to  $\gamma < 1$ , then there exist some  $\Psi_{ki}$  within the uncertainty set which is rank deficient. Similarly, if  $\mu(\Psi_{ki}) < 1$ , or  $\gamma(\Psi_{ki}) > 1$ , then no  $\Psi_{ki}$  within the uncertainty set is rank deficient and we can conclude that an influence from gene  $k$  to gene  $i$  is included in all models within the uncertainty set. Thus,  $\gamma$  provides a robustness margin, or confidence score.

The structured singular value is by now a classical tool in robust control and efficient methods for calculating upper and lower bounds on  $\mu$  exist, see *e.g.* Skogestad and Postlethwaite (1996); Zhou and Doyle (1998). These bounds are used in practice because exact calculation of the structured singular value in general is NP-hard. In this work we use the `mussv` implementation in Matlab ([www.mathworks.com](http://www.mathworks.com)) to



**Figure 4.5: Schematic illustration of robust variable selection using uncertainty sets for an example with two regressors in a two dimensional data space.** The ellipses around each variable—the same two regressors  $\phi_1$  and  $\phi_2$ , and regressand  $\xi_1$  as in Figure 4.4—depict the uncertainty set of the variable, which contains all realizations that are consistent with the error model, and define the uncertainty cone of each variable. Both regressors are selected in all feasible models, since the uncertainty cone of the regressand only intersects the uncertainty cones of the regressors in the origin. No realization of the regressand can be written as a linear combination of the two realizations marked by  $c$ , since they are parallel.

calculate the structured singular value bounds, and we report only the lower bound on  $\gamma$ , since no rank deficient realization is included in the uncertainty sets that result from scaling by any smaller value than this one, implying that the regressor  $\phi_k$  is selected in all feasible models up to this size of the uncertainty set. The structured singular value method proposed here therefore provides a reliable way to calculate a lower bound on the confidence score of every possible interaction. It ensures that no model with an alternative structure lacking the  $a_{ik}$  interaction exists within the set of all practically indistinguishable models up to the significance level that  $\gamma(\Psi_{ki})$  corresponds to.

## 4.6 Concluding examples

In order to illustrate the proposed method for robust network inference, we first consider simulated data from a simple known *in silico* 4-gene network and then proceed to analyse published response data for the engineered IRMA network in yeast. The first example serves to verify that the method is able to identify existing interactions with confidence, while the second example shows how the method can be used to verify new interaction hypotheses based on available response data.

#### 4.6.1 Example 1: Robust inference of a known 4-gene network

Consider a simple 4-gene network with the interaction matrix

$$\mathbf{A} = \begin{bmatrix} -1 & 1 & 0 & 1 \\ 1 & -1 & 0 & 0 \\ 0 & -1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} \quad (4.6.1)$$

corresponding to a cascade connection through all genes with feedback both from gene 2 and 4 to gene 1. We perform 4 simulations in which every gene in turn is perturbed and measure the steady-state responses of all genes. We assume that both the perturbations and measurements are hampered by uniformly distributed noise in the interval  $[-0.1, 0.1]$ , yielding an signal to noise ratio (SNR) of approximately 9. The resulting regressor matrix is

$$\Psi = \begin{bmatrix} 0.94 & 0.908 & -1.08 & -1.09 \\ -0.0816 & 0.924 & -1.06 & -1.1 \\ 0.99 & 0.994 & -0.0509 & -0.0511 \\ 0.934 & 0.99 & -1.06 & -0.0889 \end{bmatrix} \quad (4.6.2)$$

and the regressand matrix is

$$\mathbb{E} = \begin{bmatrix} -0.952 & -0.06 & 0.0137 & -0.0821 \\ 0.0636 & -0.953 & 0.0709 & 0.0728 \\ -0.0739 & -0.0159 & -1.01 & 0.071 \\ -0.0706 & -0.0244 & 0.0298 & -1 \end{bmatrix}. \quad (4.6.3)$$

Using these matrices, we employ (4.5.3) to compute the confidence score for each individual interaction shown in Table 4.1, allowing for an uncertainty of 0.2 in the 2-norm of each regressor and regressand. This guarantees that each uncertainty set contains the “true” variable even for the worst realization of the noise, *e.g.*  $\mathbf{v}_j = [0.1 \ 0.1 \ 0.1 \ 0.1]^T$ . With this uncertainty, we see that we can guarantee the existence of 5 out of a total of 9 interactions in the network, *i.e.* all interactions with a confidence score  $\gamma > 1$ . We have in Table 4.1 marked these links by blue colour when the interaction is activating and red colour when it is repressing. We also see from Table 4.1 that the confidence score is fairly close to 1 for the other four interactions (marked by yellow colour), while it is significantly smaller for all non-existing interactions. In other words, a clear distinction between existing and non-existing interactions in terms of confidence score exists, which could be utilized in cases when the correct uncertainty model is not known. In general, the confidence scores can be increased by performing additional perturbation experiments that make the data set more informative and increases the SNR.

**Table 4.1: Confidence scores for the 4 gene *in silico* example with simulated data.** More precisely, lower bounds on the confidence score  $\gamma$  for each possible interaction  $a_{ij}$ , calculated by the structured singular value method using the upper bound 0.2 on the radius of the uncertainty sphere of each variable. In practice, the higher the  $\gamma$  value is the more likely is the  $a_{ij}$  interaction to exist.

Gene $i \setminus j$	1	2	3	4
1	1.4	0.99	0.017	0.85
2	1.3	1.2	0.05	0.14
3	0.1	0.88	1.1	0.13
4	0.14	0.19	1.2	0.95

#### 4.6.2 Example 2: robust inference of IRMA from *in vivo* data

Here we perform robust inference of the IRMA network from the *in vivo* steady-state data set recorded by Cantone et al. (2009) when growing the *S. cerevisiae* strain in Galactose free medium, see Section 2.1. Thus, we consider inference of the same network as in the introductory example, but this time from real uncertain *in vivo* data instead of noise-free *in silico* data. The example serves to demonstrate both the feasibility of robust inference as such and more specifically the use of the method for robust inference proposed above.

To perform robust inference one needs a data model, including an error model that contains a description of the measurement errors, noise, and other sources of uncertainty in the data. Following Cantone et al. (2009), we assume that the log-fold changes, obtained by the commonly used  $2^{-\Delta\Delta C_T}$  method (Cikosz and Koppel, 2009; Schmittgen and Livak, 2008), can be explained by a linear data model

$$\sum_{j=1}^n \check{\phi}_j \check{a}_{ij} = \check{\xi}_i, \quad \phi_j = \check{\phi}_j + \mathbf{v}_j, \quad \xi_i = \check{\xi}_i + \epsilon_i. \quad (4.6.4)$$

In other words, this is the same data model as in the introductory example, Section 4.2, but now each observation is assumed to contain uncertainty, represented by  $\mathbf{v}_j$  and  $\epsilon_i$ . Since conflicting results on the error distribution of quantitative real-time RT-PCR (qRT-PCR) have been reported (Peccoud and Jacob, 1996; Bengtsson et al., 2008), we analysed the biological replicates of Cantone et al. (2009) and found that the log-fold change of the data was not normally distributed, see Section A.5. We were unable to estimate the distribution due to the low number of replicates, but note that histograms of the data contains multiple peaks, Figure A.3 and Figure A.4, which indicate a multimodal distribution in agreement with the theoretical study by (Peccoud and Jacob, 1996). To prioritize robustness of the results, we therefore decide to use an additive deterministic error model in which the absolute error of each data point is assumed to be bounded by the maximum absolute error calculated separately for each gene and perturbation experiment in Table A.2. For the pertur-

**Table 4.2: Confidence scores for the IRMA Glucose data.** More precisely, lower bounds on the confidence score  $\gamma$  for each possible interaction  $a_{ij}$ , calculated by the structured singular value method using the upper bounds in Table A.2 and Table A.4 to bound the uncertainty ellipsoid of each variable for the Glucose steady-state data recorded by Cantone et al. (2009). In practice, the higher the  $\gamma$  value is the more likely is the  $a_{ij}$  interaction to exist.

Gene $i \setminus j$	<i>CBF1</i>	<i>GAL4</i>	<i>SWI5</i>	<i>ASH1</i>	<i>GAL80</i>
<i>CBF1</i>	1.6	0.61	0.89	0.14	0.26
<i>GAL4</i>	0.27	1.6	0.11	0.2	0.0083
<i>SWI5</i>	1.2	0.89	1.7	0.039	0.37
<i>ASH1</i>	0.59	0.24	0.23	1.8	0.38
<i>GAL80</i>	0.33	0.14	0.42	0.11	1.5

bations we also assumed an additive deterministic error model with the absolute error of each data point being bounded by 0.1 for unperturbed genes and 0.2 for the perturbed genes, see Table A.4. This error model is in general conservative so we assume that the uncertainty set of each variable is ellipsoidal, with semi-principal axes of length equal to the bounds on the absolute errors, when calculating the confidence score of each interaction, since all errors are unlikely to be equal to the upper bound at the same time.

We use the structured singular value method described in the previous section to calculate the confidence score for each possible interaction up to which it is present in all models explaining the steady-state data recorded by Cantone et al. (2009) when growing their IRMA strain in medium without Galactose. For details see Section A.6, where we also infer a model for IRMA grown in Galactose. Based on the results in Table 4.2, we note that only one interaction,  $a_{31}$  (marked by blue colour), besides the self-loops on the diagonal (marked by red colour) have a  $\gamma$  value above one. The  $\gamma$  value corresponds to the factor by which the uncertainty set of each variable must be scaled in order for the interaction in question to be present in all feasible models. We have here reported it in favor of the significance level, since the uncertainty sets are based on our error model that, as every model, is an approximation that only holds under certain assumptions. It is therefore informative to know by how much we need to have underestimated the actual uncertainty in order for an interaction to not exist in all feasible models. *E.g.* the  $a_{31}$  interaction has  $\gamma = 1.2$ , meaning that it is guaranteed to exist even if the actual uncertainty level was 20% larger than predicted by our error model. This interaction should not exist according to Cantone et al. (2009), since it is not present in the engineered network in Figure 4.1. In fact, this implies that the method invalidates the engineered network based on its inability to explain the recorded data under the error model assumed here. We have not found any evidence for this link in the literature, so it constitutes a previously unknown activation of transcription of *SWI5* by *CBF1* and it illustrates that even the

knowledge of intensively studied genes and organisms is incomplete. The data is not informative enough to show that any interaction with  $\gamma < 1$  exists, but we selected the 11 interactions that are most likely to exist based on our confidence scores and fitted a model with this structure to the data using constrained optimisation to minimise the residuals. The reason for including more than the six interaction with confidence score  $\gamma > 1$  is that our error model is likely to be conservative and the “true” network is unlikely to only have six interactions. However, an important outcome of the method is still that we can provide a confidence level for every fitted interaction which is independent of the other variables chosen and the specific data fit obtained. The resulting model of the IRMA GRN, with interaction matrix

$$\hat{\mathbf{A}}_{\text{Glu}} = \begin{bmatrix} -0.143 & 0.043 & 0.073 & 0 & 0 \\ 0 & -0.183 & 0 & 0 & 0 \\ 0.103 & 0.091 & -0.219 & 0 & 0 \\ 0.036 & 0 & 0 & -0.177 & 0 \\ 0 & 0 & 0.009 & 0 & -0.152 \end{bmatrix}, \quad (4.6.5)$$

explains the data well and is a trade-off between regressor and regressand errors, see Section A.6. This model has 8 links in common with the engineered model, three additional links not found in the engineered one, and it lacks 5 links present in the engineered one. Two of the lacking interactions correspond to the Gal4-Gal80 protein interactions (Bhat and Murthy, 2001; Sellick et al., 2008), which strictly speaking should not have been included in the engineered model in the first place, since these protein interactions do not affect transcription of the *GAL4* or *GAL80* gene. Note that it is only one of several alternative models with different selections of links having  $\gamma < 1$  that can explain the data, see e.g. (A.6.5) and (A.6.4), and additional experiments are needed to discriminate among them. Cantone et al. (2009) also recorded steady-state data for IRMA grown in Galactose and two time-series when changing growth medium, but the system has two different dynamical modes (Nordling et al., 2007b), reflected in different interactions being active in the two growth mediums, so these do not provide the needed additional information.

Finally, we would like to stress that all alternative network models of IRMA grown in Galactose free medium contain the transcriptional activation of *SWI5* by *CBF1*, since we have proven that every linear model that lacks it is rejected based on the error model associated with the data and that a linear model explains the data well. It is worth noting that our regressors are logarithmic functions of the states of the underlying system, because we used the log-fold changes obtained by the  $2^{-\Delta\Delta C_T}$  method. Implying that the underlying biological system is assumed to be nonlinear, even though we use a linear data model. In general, existing feature construction techniques can be used to represent nonlinear functions in the linear framework of our robust inference method. This regulatory interaction/influence is thus hereby proven to exist in the “true” network. It does not necessarily, however, imply that a physical binding of the *CBF1* mRNA or protein to the promoter sequence of *SWI5* occurs, nor that *CBF1* is a transcription factor of *SWI5* in the

traditional sense, because inference based on perturbation experiments only reveals regulatory influences—a type of information flow (see *e.g.* Gardner and Faith, 2005).

## 4.7 Conclusions

Inference of gene regulatory networks from perturbation data is limited partly by the lack of information in the data, partly by the presence of uncertainty that hampers correct information extraction. For the case where the number of variables exceeds the number of observations, corresponding to lack of information, alternative models with different network structures that explain the available data always exist. This implies that commonly used confidence measures for individual interactions are misleading since the measures depend on the specific network model that is chosen. As we showed, many of the inferred interactions in *S. cerevisiae* deemed most significant by Lorenz et al. (2009) are in fact not needed to explain their data. This partly explains why false positives have been observed among the inferred interactions deemed most significant even by the best performing methods in the DREAM challenge (Marbach et al., 2010). Similar problems exist when inferring networks from data that in principle are sufficiently informative, but where there is significant uncertainty in the data.

Considering the fact that algorithmic details of current inference methods together with noise have been observed to largely determine the structure of inferred models, both herein and elsewhere (Marbach et al., 2010; Tjärnberg et al., 2013), caution is advised when dealing with inferred networks. Indeed, benchmarking studies have shown that inference of gene regulatory networks usually result in a large fraction of false positives and false negatives (Marbach et al., 2010; Stolovitzky et al., 2009). The current focus of network inference should therefore be shifted to determining the interactions that can be shown to exist at the significance level desired by the biologist designing the study, instead of merely inferring a network model containing all observed variables as nodes. To foster such a development we proposed a robust network inference method based on classification of all possible interactions as either existing, non-existing, alternative, or non-evidential. The method finds the interactions that are present in all linear models that can explain data at a desired significance level, as we here demonstrated by inferring the IRMA network that recently was engineered in *S. cerevisiae* by Cantone et al. (2009) for evaluation of network inference algorithms. The proposed method enables us to obtain a robust solution and assign confidence to individual interactions, without resorting to any approximation or asymptotic results, or having to tune any regularization coefficients. Applying the method to the *in vivo* steady-state data recorded by Cantone et al. (2009) we found, to our surprise, that the data cannot be explained without a previously unknown activation of transcription of *SWI5* by *CBF1* in the absence of Galactose. Existence of this interaction implies that the “true” network differs from the network they intended to engineer and illustrates that even the accumulated knowledge of well studied genes and organisms is incomplete. In fact, we find that the

---

recorded steady-state data only is informative enough to show that six interactions exist, implying that the “true” network is largely unknown and method assessments that include any other interactions are in general both unfair and misleading. These facts partly explain the multitude of different networks that have been inferred (see *e.g.* [Penfold and Wild, 2011](#); [Cantone et al., 2009](#)) and must be accounted for when the data recorded by Cantone *et al.* is used.



---

# Theory of robust variable selection and network inference

---

*“Whether you can observe a thing or not depends on the theory which you use. It is the theory which decides what can be observed.”*

Albert Einstein, objection to Werner Heisenberg’s talk on quantum mechanics, Berlin, 1926.

The desire to infer existing causal interactions with confidence directly from observed changes in variables within systems biology and bioinformatics calls for a theory for robust variable selection and network inference. Inference of causal interactions from response data is an open variable/feature/subset/model selection problem under active investigation. No current theory can cope with the combination of low signal to noise ratios, errors-in-variables, near collinearity, and few data points—seen in current gene expression data sets. We therefore here develop a theory for robust variable selection that can handle these issues by classifying all variables and interactions as either present/existing, absent/non-existing, non-evidential, or alternative. We define new concepts with necessary and/or sufficient conditions for each case, and use them in tests that only require the observed data as input and implicitly checks all models within a chosen class that cannot be rejected based on the observed data and assumed error model. The conditions enable, with desired confidence/significance, determination of all variables and interactions that must be included in or always can be excluded from all models within the set. Thus, allowing a robust classification of variables and interactions as true positives or true negatives under mild assumptions. This is achieved without estimating any model parameters. Our variable selection is therefore decoupled from parameter estimation, contrary to existing methods, such as the likelihood ratio test, F-test, and Akaike and Bayesian information criteria. The decoupling removes all interdependencies between selection and estimation that, in general, weaken the conclusions that can be drawn.

## 5.1 Introduction

Inference of causal interactions between variables based on quantitative measurements of changes caused by genetic perturbations is an important problem in the post genomic era that has received considerable attention in the systems biology and bioinformatics community (Brazhnik et al., 2002; Wolkenhauer et al., 2009; Hecker et al., 2009). The focus has in particular been on reverse engineering of gene regulatory networks from microarray and quantitative RT-PCR data, see *e.g.* Cantone et al. (2009); Gardner et al. (2003); Faith et al. (2007); Bonneau et al. (2006). Many genes have been shown to be regulated by a small number of other genes (McAdams and Arkin, 1998; Arnone and Davidson, 1997; Jeong et al., 2000; Lima-Mendez and van Helden, 2009), implying that these networks are sparsely connected and one needs to determine which of the possible interactions that actually exist. The observed data, however, is not fully informative and contains measurement uncertainty and can therefore in general be explained by different models containing different interactions—highlighting that network inference is a variable/model selection problem, where one aims to find a model containing the interactions that exist in the real system. To perform robust network inference and assign a reliable measure of confidence to individual interactions, we showed in Chapter 4 that it is necessary to check the set of all models that can explain the data. Only interactions present in all of these models, *i.e.* models consistent with the uncertainty model associated with the data at a desired significance level, can be guaranteed to exist with desired confidence. Checking every model in an infinite set is, of course, computationally infeasible, but we have reformulated the problem to do this by solving a robust rank problem based on the data matrices. Due to the close connection between variable selection and network inference, we here outline the beginning of a theory for both robust variable selection and network inference to formally prove and extend these results. It enables us to determine with desired confidence the variables and interactions that must be included or always can be excluded from the model, in other words, the true positives and negatives under mild assumptions. The data lacks information for determination of the remaining ones. We also clarify the connection and distinction between variable selection, network inference, and parameter identifiability.

Despite the fact that variable and model selection is a prerequisite of mathematical modelling, variable selection is still largely an unsolved problem. Variable selection is also known as feature selection, feature reduction, attribute selection, subset selection, and model selection (Guyon and Elisseeff, 2003; Fan and Lv, 2010; Hara and Sillanp, 2009; George, 2000; Hong et al., 2008; Stoica and Selen, 2004). We here mainly use the term variable selection, since we do not address feature construction, and the features in our examples correspond to observed variables. The increase in the amount of collected data and number of observed variables in many fields makes variable selection crucial for knowledge discovery (Fan and Li, 2006; Kohavi and Provost, 2001; Guyon and Elisseeff, 2003). Many different solutions to the variable selection problem have therefore been and are being developed and used in diverse

fields. Several recent reviews and overviews, however, show that variable selection largely is an unsolved fundamental problem. Let us first note that brute-force search for an optimal subset of variables is an NP-hard problem and thus in general not feasible even when the data is informative enough (Yusta, 2009; Cotta et al., 2004; Kohavi, 1995; Cover and Van Campenhout, 1977). Current selection methods are “either computationally feasible but far from optimal, or they are optimal or almost optimal but cannot cope with the computational complexity of feature selection problems of realistic size” (Yusta, 2009). The most relevant variables are often suboptimal for prediction, while seemingly irrelevant variables taken together can provide good prediction (Fu and Desmarais, 2010; Guyon and Elisseeff, 2003; Kohavi and John, 1997). Fan and Fan (2008) have recently demonstrated that classification using any linear discriminants can perform as poorly as random guessing due to noise accumulation in population centroids estimated in high-dimensional variable spaces, and that selection of a subset of important variables is necessary for correct classification. Many methods for selection of subsets of variables are sensitive to noise or small perturbations in the data (Guyon and Elisseeff, 2003). In particular, low signal to noise ratios, errors-in-variables, near collinearity, and few data points are problematic and no current theory or method copes with the combination of these in a robust manner. We therefore here develop the beginning of a theory of data based robust variable selection that can cope with these characteristics, which are common in gene expression data sets (Holter et al., 2000; Tegnér and Björkegren, 2007; Cosgrove et al., 2010; Nordling and Jacobsen, 2009a). By robust we mean that a desired level of confidence can be assigned to the selection of a given variable, *i.e.* that the selected subset of variables is present in all models that are consistent with a specified uncertainty model at the desired significance level. It is essential to distinguish between robust and non-robust variable selection, since they are conceptually different. The objective of the former is to determine the variables that must be included in or always can be excluded from models explaining data at a desired confidence level, while selection of the variables that optimize some quantity, *e.g.* maximize the likelihood of the observed data or minimize the prediction or classification error, typically is the objective of the latter. Methods for non-robust variable selection strive to finding the best subset of variables in some sense, while variables present in all subsets yielding a non-rejectable model are found by methods performing truly robust variable selection. Philosophically, non-robust methods can only generate alternative hypotheses, while robust methods perform hypothesis testing of implicitly generated hypotheses. The method presented here therefore offers automated discovery, *i.e.* hypothesis generation and testing, of the type sought in systems biology (see *e.g.* Ideker et al., 2001). The term robust is used in the literature in general in many different meanings, see *e.g.* Aelst et al. (2008), while we here strictly use it to mark that a property holds for all models in a set—in accordance with the tradition in robust control. In fact, we have not yet found any other line of works that would accomplish truly robust variable selection except ours. Robust variable selection is important when mechanistic insight and causal relations are sought, since good prediction, in general, does not imply that the “true”

variables or interactions have been found. This fact is illustrated in Section 4.4 by three networks with different structure that explain the data by Lorenz et al. (2009) well, and in model selection by the inability to combine the consistency of the Bayesian information criterion (BIC) with the near optimality of the Akaike information criterion (AIC) in prediction (Yang, 2005, 2007; Grünwald and De Rooij, 2011).

We here restrict ourselves to developing a theory for robust variable selection and network inference for linear data models. We assume, in other words, that the data has been preprocessed/transformed such that some feature is a linear function of the features that we select among, and we do not address this feature construction. From an estimation theoretical perspective our data model corresponds to linear regression. This resemblance of the variable selection problem to linear regression and the solution of systems of linear equations, for which there is a well established theory in statistics and linear algebra, makes it imperative to motivate why we here develop a theory for data based robust variable selection and not merely a method. We therefore next complement the discussion above by pointing out several reasons from a statistical perspective with the characteristics of biological data in mind: few data points, high dimensionality, large uncertainty, errors-in-variables, and ill-conditioned data. In linear regression the focus has traditionally been on finding an efficient minimum variance unbiased estimator under various assumptions—a theme that is recurrent in all general texts on parameter estimation, see *e.g.* Rao and Toutenburg (1999); Bingham and Fry (2010); Kay (1993); Ljung (1999); Casella and Berger (2001); Young and Smith (2005). Linear regression *per se* is not a variable selection method, but it is logical to say that only variables corresponding to nonzero parameters are selected. Typically a large number of data points, additive errors only in the regressand with known probability distribution, and a given model structure with a well conditioned regressor matrix are assumed. These conditions are rarely fulfilled in the biological cases that interests us, so the results are of little use. Strictly speaking only nonzero parameters should be estimated to utilize the data optimally, because model structure selection ideally should precede parameter estimation. Variable selection can be seen as a subproblem in model structure selection, which traditionally in statistics has been addressed though hypothesis testing, such as the F-test or likelihood ratio test (Cedersund and Roll, 2009; Casella and Berger, 2001), or information criteria, such as AIC and BIC (Stoica and Selen, 2004; Akaike, 1973; Schwarz, 1978). During the past decade a theory for regularization of linear regression, which combines variable selection with parameter estimation and deals with low numbers of data points, has emerged. In particular, the penalized likelihood framework (Fan and Lv, 2010; Fan and Li, 2001) is a nice extension of traditional theory, but it is restricted to errors in the regressand and does not provide any reliable measure of confidence for the selection, *i.e.* the selection is not robust. The variance of the parameter estimates cannot be calculated for underdetermined problems, unless one assumes that the selection is fixed. Actually, Leeb and Pötscher have shown that it is impossible to estimate the distribution of post-model-selection estimators (Leeb and Pötscher,

2006, 2008). Moreover, a well conditioned regressor matrix is crucial for correct selection and estimation (Candès and Plan, 2009). Errors-in-variables problems, *i.e.* problems with errors in the regressors, are in general harder to solve than estimation problems with errors only in the regressand and they have been studied, in particular, in econometrics and system identification (Griliches and Hausman, 1986; Söderström, 2007). Established solutions exist for cases with a large number of data points and certain assumptions on the probability distribution of the errors. Of these the total least squares solution, which minimizes the Frobenius norm of small correction matrices applied to both the regressor matrix and regressand (Markovsky and Van Huffel, 2007; de Groen, 1998), is the most well known technique. All methods we are aware of yield a single model without information about alternative models containing different subsets of the variables that can explain data, so they cannot be used for robust variable selection. A solution of a system of equations cannot be unique if the regressors are collinear, *i.e.* linearly dependent (see *e.g.* Anton and Rorres, 2005; Friedberg et al., 2003). For noise free data only exact collinearity matters, but for noisy data ill-conditioning may have adverse effects on the solution. Near collinearity or weak multicollinearity has been introduced to detect and quantify these adverse effects of ill-conditioned data with correlated regressors in estimation, such as instability of estimators, large uncertainty in the estimated parameters, sign errors even in statistically significant parameters, and good fit to data despite statistically insignificant parameters, and if possible to avoid them, see *e.g.* Belsley (1991); Liao (2010); Rao and Toutenburg (1999); Larose (2005). In variable selection collinearity causes selection of non-informative variables, in particular in high dimensions where the number of variables exceeds the number of observations (Fan and Ly, 2010). Many indicators of near collinearity have been developed, *e.g.* Belsley (1991) lists no less than 10 methods for diagnosing collinearity, but even the notion of near collinearity is vaguely defined and rules of thumb are the state of the art. In our opinion this makes collinearity diagnostics an unsolved problem. In theory, standard methods from perturbation theory for calculating and constraining the effect of errors in numerical analysis, see *e.g.* Higham (1996), could be used to combat near collinearity and obtain confidence. However, they are not applicable in practice, because the errors in observed data are orders of magnitude larger than round off errors and these methods in general become too conservative (Stewart, 1987). To summarize, we have been unable to find any theory suitable for robust variable selection for network inference that is applicable to data with the previously mentioned characteristics, based on extensive search of the literature on systems biology, bioinformatics, statistics, system identification, econometrics, machine learning, numerical analysis, robust control, and linear algebra. Linear algebra provides deep-rooted and useful methods for noise free data, but they cannot handle uncertainty. We therefore build our theory on the extension of concepts of linear algebra to uncertainty sets. Introduction of uncertainty sets and the corresponding uncertainty cones enables us to graphically explain how robustness is achieved, which fosters understanding. As a side result we also, to the best of our knowledge, give the first exact solution of the conditioning problem.

The remainder of this chapter is structured in the following way. We start by presenting the variable selection problem in Section 5.2 followed by the network inference problem in Section 5.3. In Section 5.4 we introduce uncertainty sets to describe the errors in the variables, and explain how to construct them. In Section 5.5 we develop tools for the analysis of the robust variable selection and network inference problem that are needed to classify the variables and interactions as present/existing, absent/non-existing, non-evidential, and alternative. Our first goal in this section is to present the ideas and reasoning that lead up to the key concept in robust variable selection, namely practical selectability. A practically selectable variable must be selected to explain the data and the corresponding interaction therefore exists. The second goal is to extend classical concepts of linear algebra to account for errors and uncertainty, which we use to establish our theory. We then in Section 5.6 use these tools to make a general definition of practical selectability and establish both sufficient and necessary conditions on data for practical selectability. Since the sign of the interactions often also are of interest, we also establish data conditions for when the sign can be determined with a desired confidence, corresponding to practical assignability of the regressor. The counter part of practical selectability—practical excludability—is defined in Section 5.7. These variables can always be excluded while explaining available data and the corresponding interactions are non-existing. We establish sufficient and necessary data conditions also for this classification. Definitions and data conditions for the final two cases—non-evidential and alternative variables and interactions—that exist due to lack of information are presented in Section 5.8. All essential definitions and conditions for the classification are summarized in Table 5.1. In Section 5.9 we geometrically interpret the practical independence and collinearity requirements for practical selectability and excludability in terms of uncertainty sets and cones using a two gene example. To perform robust variable selection and network inference in practice one needs to determine if an uncertain matrix has full column rank, so we end by showing how the singular value decomposition and structured singular value can be utilized for this in Section 5.10. Our presentation of the theory is intended for computational biologists or engineers working on network inference or variable selection. The proofs of the presented results are given in Section 5.12, in order to make the presentation of the results more concise.

## 5.2 The variable selection problem

Variable selection has been studied in many fields from various perspectives. We therefore start here by discussing common formulations of the variable selection problem, before introducing our own robust formulation, and end with a discussion on how it fits into the big picture.

Selection of a subset of variables is, of course, at the core of every formulation of the variable selection problem, but the objectives differ so much that it is hard to give a general definition. We therefore refer the interested reader to [Dash and](#)

Liu (1997); Peng et al. (2005); Molina et al. (2002), where no less than six different definitions of variable selection are given. All these formulations concern selection of a subset such that some quantity or measure is constrained or optimized. None of the formulations aim at the determination of the subset that always is present and the subset that always can be excluded in models that are consistent with a specified uncertainty model at a desired significance level, which is our aim with robust variable selection. Thus, none of these six formulations covers robust variable selection, and we therefore give our own definition of robust variable selection.

**Definition 5.2.1. Robust variable selection.** Given a function  $f$  of the  $n$  variables  $\phi_j$ , indexed by  $\mathcal{V}$ , and a target variable  $\xi$ , find the variables that, for all realizations  $\tilde{\xi} \in \mathcal{U}_\xi$ , either must be selected, or always can be excluded, for some realizations  $\tilde{\phi}_j \in \mathcal{U}_{\phi_j}$ , in

$$f(\tilde{\phi}_j, j \in \mathcal{V}) = \tilde{\xi}. \quad (5.2.1)$$

Here  $\mathcal{U}$  for each variable denotes the uncertainty set containing all realizations of it that are consistent with a specified error/noise/uncertainty model at the desired confidence/significance level.

The two main differences between our robust definition and the ones found in the literature are the way in which we account for uncertainty and the fact that we seek two subsets—the variables that must be selected and those which always can be excluded. We explicitly account for existence of many different realizations of each variable that are consistent with a specified uncertainty model at the desired significance level through introduction of the uncertainty set  $\mathcal{U}$  of each variable, which we later explain in detail. All variables/features constructed from measurement data contain measurement errors and one therefore in general has variables that actually represent a set of realizations. These realizations are indistinguishable in the sense that none of them can be rejected based on the data model used to describe the measurement process and error caused by it. The idea behind robust variable selection is to achieve correct selection by checking all realizations, and thereby utilize the fact that the “true” realization belongs to the uncertainty sets when the correct uncertainty model is used. To make the selection robust, one needs to account for the measurement errors and the indistinguishable realizations of each variable that they give rise to. The relation  $f(\tilde{\phi}_j, j \in \mathcal{V}) = \tilde{\xi}$  is in general fulfilled for many different combinations of realizations of the variables. This implies that four possibilities exist for each variable  $\phi_j$ , *i.e.* it is either

- (a) present in every combination,
- (b) absent in some combination of variables for all realizations of the target variable,
- (c) without any impact on the ability to explain the target variable, or

(d) present in some combination and absent in some other combination.

In the first case, the variable must always be selected to explain data, since it is present in every combination, while in the second case it always can be excluded when explaining data, since it is absent in some combination for all realizations of the target variable. In the third case, the variable is non-evidential, since it does not affect the ability to explain data whatsoever. Equation (5.2.1) is for every combination of realizations of the other variables fulfilled independent of if this variable is present or absent, implying that the data does not contain any information about it. In the fourth case, the variable is alternative, since for some combination it is needed to explain data while it is not for some other combination. We will later say that the variable is either practically selectable, excludable, negligible, or alternative depending on which case it belongs to, and give precise definitions of these cases. We presented the interaction analogue of these cases—existing, non-existing, non-evidential, or alternative—already in (4.2.9) in the previous chapter. In particular the first and second cases are of interest, since they correspond to variables that must be selected and always can be excluded, respectively. The former variables are present in the “true” system that generated the data when its relations are approximated using the function  $\mathbf{f}$ , assuming that the “true” and observed variables belongs to the uncertainty sets. In other words, we not only select the correct subset of variables but we also “know” that this subset is correct. This is in our opinion a major advantage over non-robust variable selection methods that give the most likely subset, because then even when the subset is correct we have no way of knowing that it is correct. Due to the existence of the fourth case, the subset of variables that must be selected is in general different from the subset obtained if one *e.g.* would minimize  $J(\phi_j, j \in \mathcal{S}) = \|\mathbf{f}(\tilde{\phi}_j, j \in \mathcal{V}) - \tilde{\xi}\|$  for any realizations  $\tilde{\phi}_j \in \mathcal{U}_{\phi_j}$  and  $\tilde{\xi} \in \mathcal{U}_{\xi}$ , which commonly is done when the most likely selection is sought. Our definition, in other words, provides two subsets—the variables that must be selected and those which always can be excluded—and knowledge that in general cannot be obtained using any non-robust method.

The solution of this robust variable selection problem enables us to gain mechanistic insight, since observed data when accounting for uncertainty and desired confidence only can be explained using the function  $\mathbf{f}$  if the variables that must be selected are used, while the variables that always can be excluded are not needed. Robust variable selection therefore provides a way to do automated systematic data based hypothesis generation and testing, while existing methods only can be used for hypothesis generation. We therefore believe that robust variable selection opens up a new research direction within statistical inference and complements the optimal subsets for prediction, classification, and approximation provided by existing variable selection methods.

### 5.3 The network inference problem

We here first discuss the objective of network inference, based on work within systems biology and bioinformatics, to motivate the commonly used definition and introduce our definition of robust network inference. We then discuss the limitations and assumptions made on the data model and underlying system dynamics. Finally we make the connection between our robust network inference problem and the robust variable selection problem in the previous section, as well as comment on our strategy for solving them.

The objective of network inference is to discover interactions/relations among variables based on observations of the variables. Together these variables and interactions form a network that can be represented by a graph having the variables as nodes and the interactions as edges. The structure/topology of the network is typically inferred from observations of changes in the variables following a set of perturbations of the system, *e.g.* changes in mRNA abundance following insertion of a plasmid with an extra copy of one gene. We therefore favour the name network inference over the alternatives—network reconstruction and reverse engineering—that also are used in the literature.

Many different inference algorithms that are based on different model formalisms and assumptions exist, see *e.g.* the reviews [Emmert-Streib et al. \(2012\)](#); [Hecker et al. \(2009\)](#); [Bonneau \(2008\)](#); [Gardner and Faith \(2005\)](#). Different types of interactions and networks are discovered by using them, implying that different formulations of the network inference problem are used. The recurring objective is, however, to determine the edges that exist between the nodes under some condition of interest based on quantitative observations of variables representing the nodes. All of the published inference algorithms that we have analysed return the most likely network in some algorithm dependent sense, including the works claiming to do robust inference, such as [Yuan et al. \(2011\)](#); [Nadadoor et al. \(2011\)](#). One can thus define network inference as prediction of edges between nodes based on quantitative observations of variables representing the nodes. Many attempts to assign confidence measures to the inferred edges exist, but all measures that we have analysed fail to account for the existence of alternative network models with different structure that can also explain the observed data. We mathematically prove that no reliable measure of confidence can be assigned to an inferred interaction for data sets with fewer observations than variables in Chapter 4.

Biologists often want to test and prove that an interaction of regulatory importance exists between two genes, but current inference methods can in principle only predict interactions, *i.e.* generate hypotheses. We have earlier shown that possible network interactions based on partly informative noise-free data in general can be classified as ([Nordling and Jacobsen, 2011](#), or Section 4.2):

- (a) existing,
- (b) non-existing,

(c) non-evidential, or

(d) alternative.

We here show that this classification also can be done for noisy or uncertain data and later give precise definitions of these four cases. Ideally we would like to have all possible interactions belonging to either of the first two cases, *i.e.* they must either be included or always can be excluded from the network model. We therefore define robust network inference in the following manner.

**Definition 5.3.1. Robust network inference.** Find the interactions  $a_{ij}$  that, for all realizations  $\check{\xi}_i \in \mathcal{U}_{\xi_i}$ , either must exist, *i.e.*  $a_{ij} \neq 0$ , or always can be excluded, *i.e.*  $a_{ij} = 0$ , for some realizations  $\check{\phi}_j \in \mathcal{U}_{\phi_j}$ , such that

$$\sum_{j \in \mathcal{V}} \check{\phi}_j a_{ij} = \check{\xi}_i \quad (5.3.1)$$

holds separately for each  $i \in \mathcal{V}$ . Here  $\mathcal{U}$  for each variable denotes the uncertainty set containing all realizations of it that are consistent with a specified error/noise/uncertainty model at the desired confidence/significance level. Typically, each regressor  $\phi_j$  represents the responses of node  $j$ , while each regressand  $\xi_i$  represents the perturbations of node  $i$ .

By making this definition of robust network inference, we restrict ourselves to inference of a linear map or transformation  $\mathbf{A}$ , consisting of the elements  $a_{ij}$ , and assume that the data has been generated by the following data model

$$\phi_j = \check{\phi}_j + \mathbf{v}_j, \quad \xi_i = \check{\xi}_i + \epsilon_i, \quad \phi_j, \xi_i \in \mathbb{R}^m, \quad (5.3.2a)$$

$$\sum_{j=1}^n \check{\phi}_j \check{a}_{ij} = \check{\xi}_i, \quad \check{a}_{ij} \in \mathbb{R}, \quad \forall i, j \in \mathcal{V} = \{1, 2, \dots, n\}. \quad (5.3.2b)$$

In this data model the observed regressors (regressands)  $\phi_j$  ( $\xi_i$ ) are equal to the latent “true” variables, marked by  $\check{\phantom{x}}$ , plus additive errors  $\mathbf{v}_j$  ( $\epsilon_i$ ), and the “true” regressors and regressands have been generated by a system of linear equations defined by the “true” interaction matrix  $\check{\mathbf{A}}$ , *i.e.* the “true” map representing the network model including the observed states of the system. We here refer to (5.3.2b) as the network model, because it specifies the considered set of models of the network, and (5.3.2a) together with an assumption on  $\mathbf{v}_j$  and  $\epsilon_i$  as the error model. Our assumption on the errors  $\mathbf{v}_j$  and  $\epsilon_i$  is given in the following section, where we discuss construction of the uncertainty sets  $\mathcal{U}$  and give examples both for a deterministic and stochastic error model. From a parameter estimation perspective this is an errors-in-variables regression problem, which is why we henceforth call the variables regressors and regressands. Depending on the context and field, the regressors are also known as independent variables, explanatory variables, predictor variables, controlled variables, manipulated variables, exposure variables, or input

variables, while the corresponding term for the regressands are dependent variables, responding variables, response variables, measured variables, observed variables, explained variables, or output variables. By using the pair regressor-regressand we also hope to avoid confusion with the system concept of input-output variables and common assumption of noise-free explanatory variables. Strictly speaking the regressors and regressands are often features, because they typically are functions of the state variables of the system, *e.g.* logarithmic transformations as used in [Jörnsten et al. \(2011\)](#). The term “feature” is used to distinguish functions of observed variables from the observed variables (see *e.g.* [Guyon and Elisseeff, 2003](#)), but we do not address feature construction here. However, it implies that the dynamics of the underlying system may be nonlinear and use of our definition of robust network inference is thus not limited to linear systems, even though it may seem so at first sight. Many model formalisms for representing biological networks lead to inference of linear maps, and inference of linear maps is therefore an important problem tackled in numerous studies, such as [Gardner et al. \(2003\)](#); [Cosentino et al. \(2007\)](#); [Cosgrove et al. \(2008\)](#); [Julius et al. \(2009\)](#); [Lähdesmäki and Shmulevich \(2008\)](#). One such formalism is the linear ODE model of GRNs, for which we have included a detailed derivation in Section 2.3. It is also worth noting that the variables observed in gene expression data sets typically are logarithmic transformations of the state variables of the underlying system, see *e.g.* (2.2.4) in Section 2.2, which originally motivated us to not make any distinction between features and variables.

The connection between the robust network inference problem in Definition 5.3.1 and the robust variable selection problem in Definition 5.2.1 may at first sight not be obvious and we therefore explain it. First, note that, for each  $i$ , *i.e.* row of the interaction matrix, we have a separate problem and we therefore henceforth study

$$\underbrace{\sum_{j \in \mathcal{V}} \theta_j \tilde{\phi}_j}_{=f(\tilde{\phi}_j, j \in \mathcal{V})} = \tilde{\xi}, \quad (5.3.3)$$

where we have renamed the parameter  $a_{ij}$  as  $\theta_j$  and dropped the index of the regressand  $\tilde{\xi}_i$ . On the left hand side we recognize  $f$ , a parametrized function of the variables  $\phi_j$ . This equation is identical to (5.2.1), except for the fact that we instead of one function have a class of functions—all linear functions of  $\phi_j$ . In a system identification context this class of functions would be called a model set, *i.e.* the set of all linear models of  $\phi_j$ , see *e.g.* [Ljung \(1999, p. 107\)](#). Note that this is a system of linear equations for each combination of realizations  $\tilde{\phi}_j$  and  $\tilde{\xi}$ . Second, when an interaction can be excluded, then  $\theta_j = 0$  and the corresponding regressor  $\phi_j$  is excluded. On the contrary, when an interaction exists, then  $\theta_j \neq 0$  and the corresponding regressor  $\phi_j$  is selected. In other words, each robust network inference problem consists of  $n$  robust variable selection problems, and the ability to solve the latter for all linear functions of  $\phi_j$  is both sufficient and necessary to solve the former. Variable selection is at large an unsolved problem and solution of the robust variable selection problem for all linear functions is a prerequisite to

solve the nonlinear case. Moreover, all continuous nonlinear functions are locally approximately linear, as long as the linear term of the Taylor expansion dominates, and solution of linear problems are in general easier than nonlinear problems. Each regressor and regressand is in general a feature, which can be a nonlinear function of a state of the underlying system, *e.g.* an expression given by the reaction kinetics. In other words, assumption of the linear data model (5.3.2) does not imply that the underlying system is assumed to be linear. We in the remainder of this work restrict ourselves to showing how to solve the robust variable selection problem for linear functions.

We next lay the foundation for a theory of robust variable selection by developing concepts and building tools. The data enters the variable selection through the system of linear equations in (5.3.3) and it is clear from the noise free case with only one realization of each variable that one should select a set of regressors that is necessary for spanning the subspace of the regressand, as demonstrated in Section 4.2. Standard concepts of linear algebra that have been used for analysis of systems of linear equations, such as subspace, span, linear independence, and linear dependence, are independent of the number of data points, incorporate collinearity, and are essential for understanding, explaining, and performing variable selection. But these concepts must be extended to the case with uncertain data. We therefore next introduce uncertainty sets and define extensions of these concepts based on ideas picked from robust control, statistics, set membership identification, and optimization theory, among others. The idea is to introduce the necessary definitions and theorems so that we can establish necessary conditions on data for robust variable selection and determine what can be inferred from a given data set both qualitatively and quantitatively.

## 5.4 Introduction of uncertainty sets

To represent errors, noise, and uncertainty of observed data, we introduced uncertainty sets, denoted by  $\mathcal{U}$ , in the definition of our robust variable selection problem. Despite being standard in robust control, see *e.g.* Skogestad and Postlethwaite (1996), uncertainty sets used in few other fields, so we here first motivate why we use them and the idea behind them. We then state the assumptions that we make on the errors by using the uncertainty sets and explain our notation. Finally we explain how to construct uncertainty sets from both commonly used stochastic and deterministic error models through a few examples.

We decided to use uncertainty sets to represent the error, noise, and uncertainty present in observed variables, because they capture the essential consequence of these errors, *i.e.* that more than one realization exists for each variable, in a simple and robust way. Today different error models are used and selection methods are often tailored to specific models. We need a flexible way to describe the uncertainty, which enables us to extend the concepts of linear algebra to uncertain data in a way that is independent of the assumed error model and that can be used to find

the variables that must be selected or always can be excluded. Uncertainty sets can in many cases even be constructed for a mixture of error models, which allows us to deal with a combination of both stochastic noise, *i.e.* random variables with known probability distributions, and unknown deterministic errors bounded by some known function. The former uncertainty description is standard in statistics, signal processing and system identification (Casella and Berger, 2001; Kay, 1993; Ljung, 1999), while the latter is used in numerical analysis, robust control, and bounded error/set membership identification (Higham, 1996; Skogestad and Postlethwaite, 1996; Reinelt et al., 2002). For a discussion on relative merits of stochastic versus deterministic error models we refer to Ninness and Goodwin (1995); Reinelt et al. (2002). By introducing flexible uncertainty sets, we hope to refrain users from neglecting error sources. Currently, the error model is too often selected to fit the inference method, instead of being based on the measurement technique and experiment set-up. In the extreme case, no error is assumed; *e.g.* Steinke et al. (2007) assumes no errors in the measured gene expression profiles. Traditionally, normally distributed errors are assumed, but the number of data points is typically low, so the central limit theorem, see *e.g.* Casella and Berger (2001, p. 236-238), cannot be used to motivate this assumption in general. We have taken the concept of uncertainty sets from robust control, where sets are used to describe uncertainty about the model and analyse how it affects stability and performance of a system, as well as design controllers that provide robust stability and performance despite the uncertainty, see *e.g.* Skogestad and Postlethwaite (1996).

Philosophically the introduction of uncertainty sets can be motivated from either a model invalidation or robustness perspective; the former being standard in statistics and descriptive science, the latter in robust control and design science. To make a long story short one can say that models in descriptive science are used to describe and test inferred relations, while they in design science are used to design, construct or modify systems. Invalidation is in the former case used to reject false models based on data, while robustness in the latter is used to ensure functioning despite errors in the model, *i.e.* obtain confidence. Today biological networks are mainly inferred to gain insight but we expect the models to soon be used for design of drugs and synthetic pathways, so both perspectives are relevant in biology. By constructing the uncertainty sets such that they contain all realizations of the variables that are consistent, *i.e.* in agreement with the assumed error model at a desired significance level, the type of robustness that we seek is guaranteed. Any model predicting or containing a realization that does not belong to the uncertainty set of the corresponding variable is rejected from an invalidation perspective, as well as any a model for which functioning is not guaranteed from a robustness perspective. All models that cannot be rejected and for which functioning needs to be guaranteed, on the other hand, predict or contain realizations that belong to the uncertainty set of the corresponding variable. In principle each combination of one realization from the uncertainty set of each variable gives rise to one model, so by taking all consistent combinations we get the set of models that cannot be rejected based on the data. By considering the set of all non-rejectable models, robustness against errors and desired

confidence is achieved. The error model itself is often uncertain and estimated from data and it is therefore of interest to assess how sensitive the inferred relations are to changes in the size of the uncertainty sets. This can be done either by using different significance levels  $\alpha$  to generate uncertainty sets, such that the cumulative probability of observing a combination of variable realizations belonging to the uncertainty sets is  $1 - \alpha$  based on the assumed error model, or different scaling of the uncertainty sets, which can be interpreted as different significance levels. From an invalidation perspective the significance level determines the probability of type I errors (false positives), corresponding to rejection of a model that is true. The required quality of the model, *i.e.* level of approximation, is determined by the application the network model is constructed for, and the significance level can also be seen as a measure of how poor an approximation of the underlying biological network that the set of non-rejectable models contain, implying that it should be selected based on the application. We recommend that the variable selection is evaluated for different choices of the significance level, in particular if the error model is poorly motivated or the significance level required by the application is unclear. Alternatively, the theory that we develop later on can be used to assess the level of uncertainty at which a specific variable and interaction can be classified as existing or non-existing.

Having motivated the choice of uncertainty sets, it is time to specify the assumptions that we make on the additive errors  $\mathbf{v}_j$  and  $\epsilon_i$  of the data model in (5.3.2). We henceforth assume that the errors belong to uncertainty sets, *i.e.*  $\mathbf{v}_j \in \mathcal{U}_{\mathbf{v}_j}$  and  $\epsilon_i \in \mathcal{U}_{\epsilon_i}$ , and that  $\mathbf{v}_j = \epsilon_i = \mathbf{0}$  is included in these uncertainty sets. The shape and size of the uncertainty sets depend on the measurement technique, the system under study, *etc.*, and they should therefore be specified by the scientists conducting the study. We will later use  $\tilde{\phi}_j \in \mathcal{U}_{\phi_j}$  as shorthand notation both for  $\tilde{\phi}_j = \check{\phi}_j + \mathbf{v}_j$  and  $\tilde{\phi}_j = \phi_j - \mathbf{v}_j$ , with  $\mathbf{v}_j \in \mathcal{U}_{\mathbf{v}_j}$ . The former case is used for analysis when the “true” regressor  $\check{\phi}_j$  is known, *e.g.* when analyzing or comparing selection algorithms based on numerical examples, while the latter is used when the observed regressor  $\phi_j$  is obtained from data, *e.g.* to analyse how informative the data is or to classify the interactions. The theory that we develop later on is useful in both cases and it is usually obvious if the former or latter case is in question, so this ambiguity should not cause any confusion. Note that  $\phi_j$  is used to denote both the observed regressor and the variable in general, while  $\check{\phi}_j$  always denote the “true” regressor, and  $\tilde{\phi}_j$  a realization of the variable.

Traditionally a single stochastic or deterministic error model is used and we therefore next explain how to construct uncertainty sets from these, including an illustrative example in the next paragraph. The basis for construction of uncertainty sets based on traditional stochastic or deterministic error models is the selection of an application specific significance level  $\alpha$  or confidence level  $1 - \alpha$ . For each significance level  $\alpha$  one can based on the assumed error model calculate uncertainty sets  $\mathcal{U}_{\mathbf{v}_j}^\alpha$  and  $\mathcal{U}_{\epsilon_i}^\alpha$  that contain the error realizations with cumulative probability  $1 - \alpha$  in order of decreasing probability density. Many different sets with the same

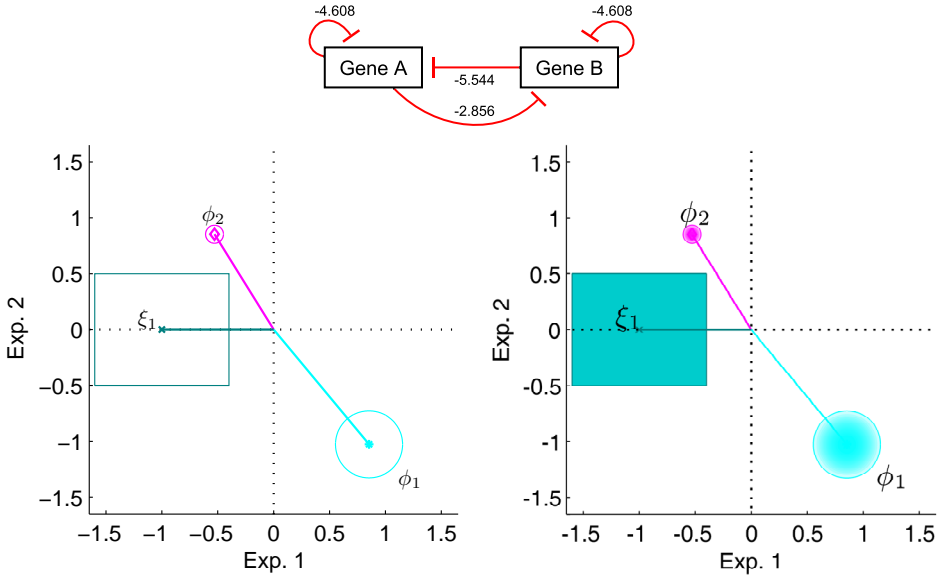
cumulative probability exist, but in general different content of error realizations and varying size. Sets with different content in general provide different power when used in hypothesis tests, see *e.g.* Casella and Berger (2001, p. 387-402); Young and Smith (2005, p. 67-68). By constructing the uncertainty sets such that they contain realizations in order of decreasing probability density, we minimize the size of the sets and maximize the power from a model rejection perspective. We therefore henceforth assume that the power is maximized and that the size of the uncertainty sets increases with decreasing  $\alpha$ . In order for the uncertainty sets to contain exactly the error realizations that are consistent with the assumed error model for all types of error models, we can in general only allow consistent combinations of the regressors and regressand errors, which for some uncertainty descriptions are a subset of all possible combinations. Henceforth, when we use the term “consistent” in conjunction with the uncertainty sets then it implies that we only consider combinations of the regressors and regressand errors that are in agreement with the assumed error model when the power of a test is maximised. If the assumed error model is stated in terms of the uncertainty sets or is deterministic, then all possible combinations are in agreement with the error model and consequently considered, but for certain stochastic error models only a subset of the possible combinations are in agreement with the error model and hence considered.

The uncertainty sets are typically easy to calculate for deterministic error models, while for stochastic error models we have to in general resort to Monte Carlo simulations and conservative uncertainty sets. To demonstrate how the uncertainty sets are calculated we provide two examples, using a deterministic and a stochastic error model, both illustrated in Figure 5.1. The illustration is done using data from the TwO gene ExampLe (TOEL) network, which is described in detail in Section A.7. The example and values are selected to get a simple and clear illustration, which is biologically plausible, despite being numerical. In the first example we assume the deterministic error model  $\|\mathbf{v}_1\|_2 \leq 0.3$  for  $\phi_1$ , which is the regressor corresponding to expression changes in gene *A*, and  $\|\mathbf{v}_2\|_2 \leq 0.08$  for  $\phi_2$ , which is the regressor corresponding to expression changes in gene *B*. In other words, the sum of the errors in the measurements of each gene is bounded and the accuracy of the measurements of each gene is the same in each experiment but it differs for each gene. We call this column uncertainty. For the errors in the first regressand we assume  $|\epsilon_{11}| \leq 0.6$  and  $|\epsilon_{21}| \leq 0.5$ , and for the second regressand  $|\epsilon_{12}| \leq 0.5$  and  $|\epsilon_{22}| \leq 0.6$ , *i.e.* element-wise errors where the uncertainty of the perturbation is larger in the experiment when the corresponding gene is perturbed. The nominal uncertainty sets corresponding to significance level  $\alpha = 0$  are

$$\mathcal{U}_{\mathbf{v}_j} = \{\mathbf{v}_j \mid \|\mathbf{v}_j\|_2 \leq \varrho_j\}, \quad (5.4.1)$$

with  $\varrho_1 = 0.3$  and  $\varrho_2 = 0.08$ , and

$$\begin{aligned} \mathcal{U}_{\boldsymbol{\epsilon}_i} &= \{\boldsymbol{\epsilon}_i \mid \boldsymbol{\epsilon}_i \triangleq [\epsilon_{1i}, \dots, \epsilon_{ki}, \dots, \epsilon_{mi}]^T, \\ &\text{with } |\epsilon_{ki}| \leq \varpi_{ki}, k \in \{1, 2, \dots, m\}\}, \end{aligned} \quad (5.4.2)$$



**Figure 5.1: Illustration of uncertainty sets derived from a deterministic error model (left) and a stochastic error model (right).** The two steady-state perturbation experiments (P1 and P2) that yielded the data for the TOEL network (top) are illustrated in Figure A.7. The regressors are depicted by the solid cyan or magenta coloured vectors ending by a star or diamond and tagged by  $\phi_1$  or  $\phi_2$ , while the solid teal coloured vector ending by an x and tagged by  $\xi_1$  depicts the regressand. The uncertainty set of each regressor  $\phi_j$  and regressand  $\xi_1$  is depicted by the solid coloured circles or rectangles drawn around the tip of the respective vector. In the deterministic case (left), every vector within the uncertainty set is an equally likely realization of the corresponding regressor or regressand. In the stochastic case (right), the likelihood and probability density decreases with the distance from the observed regressor. The gradual decrease in intensity of the colour in the uncertainty circles is proportional to the probability density, until the circle, which marks the set containing all realizations that cannot be rejected based on the assumed normal distribution of the errors in each regressor at significance  $\alpha = 0.05$ . On the other hand, the probability density is constant within the rectangle based on the assumed uniform distribution. The rectangle marks the set containing all realizations that cannot be rejected based on the distribution at significance  $\alpha = 0$ . In both cases, we have a set of vectors—realizations of each variable—and properties that hold for all vectors are robust at the selected significance level. This example is further used to illustrate robust variable selection in Figures 5.2 and A.8.

with  $\varpi_{11} = \varpi_{22} = 0.6$  and  $\varpi_{21} = \varpi_{12} = 0.5$ . The two regressors with their uncertainty circles and the first regressand with its uncertainty rectangle are illustrated in Figure 5.1 (left). In this case all possible combinations are consistent with the error model, meaning that any combination of one vector picked out of each uncertainty set agrees with the error model.

In the second example we instead assume a stochastic error model, in which the regressor errors are independent of the regressand errors. The regressor errors are distributed according to a multivariate normal distribution as  $\tilde{\mathbf{Y}} \sim N(\mathbf{0}, \mathbf{\Lambda}_{\mathbf{Y}})$ , with  $\mathbf{Y} \triangleq [\mathbf{v}_1, \dots, \mathbf{v}_n]$ ,  $\tilde{\cdot}$  denoting vectorization column by column, *i.e.*  $\tilde{\mathbf{Y}} = [\mathbf{v}_1^T, \dots, \mathbf{v}_n^T]^T$ , and  $\mathbf{\Lambda}_{\mathbf{Y}}$  being the covariance matrix. The regressand errors are distributed according to a multivariate uniform distribution as  $\tilde{\mathbf{\Pi}} \sim U(\tilde{\mathbf{Q}}_s, \tilde{\mathbf{Q}}_e)$ , with  $\mathbf{\Pi} \triangleq [\epsilon_1, \dots, \epsilon_n]$ ,  $\mathbf{Q}_s \triangleq [\mathbf{q}_1^s, \dots, \mathbf{q}_n^s]$  being a matrix with the start values and  $\mathbf{Q}_e \triangleq [\mathbf{q}_1^e, \dots, \mathbf{q}_n^e]$  a matrix with the end values of the interval in which each realization lies. The uncertainty sets are in this case

$$\begin{aligned} \mathcal{U}_{\mathbf{v}_j}^\alpha &= \{ \mathbf{v}_j \mid \tilde{\mathbf{Y}}^T \mathbf{\Lambda}_{\mathbf{Y}}^{-1} \tilde{\mathbf{Y}} \leq \chi^{-2}(\alpha, nm), \\ &\quad \text{with } \mathbf{Y} = [\mathbf{v}_1, \dots, \mathbf{v}_j, \dots, \mathbf{v}_n] \\ &\quad \text{for some } \mathbf{v}_k \in \mathbb{R}^m \\ &\quad \forall k \in \{1, \dots, j-1, j+1, \dots, n\} \} \end{aligned} \quad (5.4.3)$$

and

$$\mathcal{U}_{\epsilon_i} = \{ \epsilon_i \mid \mathbf{q}_i^s \leq \epsilon_i \leq \mathbf{q}_i^e \}. \quad (5.4.4)$$

Here  $\chi^{-2}(\alpha, nm)$  is the inverse of the chi-square cumulative distribution with  $nm$  degrees of freedom, such that  $P[\chi^2(nm) > \chi^{-2}(\alpha, nm)] = \alpha$ , see *e.g.* Chew (1966). We here use the covariance matrix of the regressors  $\mathbf{\Lambda}_{\mathbf{Y}} = \text{diag}([0.0095, 0.0095, 0.00067, 0.00067])$ , such that the uncertainty sets, which contains all error vectors  $\mathbf{v}_j$  that cannot be rejected by a  $\chi^2$  hypothesis test at significance level  $\alpha = 0.05$ , are almost identical to the ones in the deterministic case above. Note that we selected  $\alpha = 0.05$ , since it is the most commonly used significance level in statistical textbooks and the uncertainty sets in this case would contain the whole space if we had used the unreasonable significance level  $\alpha = 0$ . We also use the start and end values of the multivariate uniform distribution

$$\mathbf{Q}_s = \begin{bmatrix} -0.6 & -0.5 \\ -0.5 & -0.6 \end{bmatrix} \quad \mathbf{Q}_e = \begin{bmatrix} 0.6 & 0.5 \\ 0.5 & 0.6 \end{bmatrix}, \quad (5.4.5)$$

such that the regressand uncertainty sets at significance level  $\alpha = 0$  are identical in size and shape to the deterministic example above. This selection is made to make the similarities and differences between uncertainty sets derived from deterministic and stochastic error models clear. The two regressors with their uncertainty circles and the first regressand with its uncertainty rectangle are illustrated in Figure 5.1 (right). The only difference, besides the significance level, is that not all possible

combinations of regressor realizations are consistent with the error model in the stochastic case, while they are in the deterministic case. In the stochastic case the uncertainty sets in (5.4.3) must be complemented with the additional constraint that only combinations of realizations  $\tilde{\mathbf{Y}}$  such that

$$\tilde{\mathbf{Y}}^T \mathbf{\Lambda}_{\tilde{\mathbf{Y}}}^{-1} \tilde{\mathbf{Y}} \leq \chi^2(\alpha, nm) \quad (5.4.6)$$

are consistent. This implies that the consistent realizations of  $\mathbf{v}_j$  depend on the realization of the other regressors. We later on use the word “consistent” as short hand notation for any additional constraints. The uncertainty sets are in other words conservative for this stochastic error model. The conditions that we derive later on, *e.g.* for proving existence of an interaction, are hence sufficient if we for this error model consider all possible combinations, but necessary only if we restrict ourselves to the consistent combinations within the uncertainty sets. The uncertainty sets contain combinations that are rejected based on the stochastic error model that we assumed due to  $\tilde{\mathbf{Y}}^T \mathbf{\Lambda}_{\tilde{\mathbf{Y}}}^{-1} \tilde{\mathbf{Y}} > \chi^2(\alpha, nm)$ . One such example is

$$\tilde{\Phi} = \begin{bmatrix} 1 & -0.5 \\ -1.2 & 0.9 \end{bmatrix}, \quad (5.4.7)$$

for which  $\tilde{\mathbf{Y}}^T \mathbf{\Lambda}_{\tilde{\mathbf{Y}}}^{-1} \tilde{\mathbf{Y}} = 9.9 > 9.5 = \chi^2(0.05, 4)$ , which should be rejected but belongs to the uncertainty circles in Figure 5.1 (right). At least for jointly normally distributed errors with the same variance in each regressor for independent experiments it is possible to account for the constraint in (5.4.6) and obtain necessary conditions, as we show in Section A.8. The essential thing is however, that uncertainty in our framework implies that we have for each variable a set of vectors instead of just the observed one, independent of how it is derived. We henceforth use the deterministic example to illustrate our results.

## 5.5 Robust tools for analysis of variable selection

Here tools for analysis of the robust variable selection problem (Definition 5.2.1) and network inference problem (Definition 5.3.1) are introduced. These tools are needed to classify the variables and interactions and thereby solve the two problems, as well as to prove that the solution is correct. We recommend impatient readers to have a look at Table 5.1, where the essential definitions and conditions for the classification are summarized. Our first goal in this section is to present the ideas and reasoning that lead up to the key concept in robust variable selection—practical selectability. A practically selectable variable must be selected to explain the data and the corresponding interaction therefore exists. The second goal is to extend classical concepts of linear algebra that are needed later on. We start from a solution of systems of linear equations perspective and make connections to optimisation theory, parameter identifiability, parameter estimation, model selection, and analysis of multicollinearity.

To solve the robust variable selection and network inference problems we need, based on (5.3.3), to study the following systems of linear equations

$$\sum_{j \in \mathcal{V}} \theta_j \tilde{\phi}_j = \tilde{\xi}, \quad \forall \text{ consistent } \tilde{\phi}_j \in \mathcal{U}_{\tilde{\phi}_j}^\alpha \subseteq \mathbb{R}^m$$

$$\text{and } \tilde{\xi} \in \mathcal{U}_{\tilde{\xi}}^\alpha \subseteq \mathbb{R}^m. \quad (5.5.1)$$

In principle, we get one system of linear equations for each consistent combination of realizations of the variables that cannot be rejected based on the observed data and assumed error model. We therefore need to determine which regressors and interactions are present in all of these systems, absent for all realizations of the regressand, without any impact on the ability to explain the regressand, and present in some system while absent in some other system. For continuous variables the number of systems is always infinite and checking all of them is computationally infeasible. We therefore next develop tools that allow us to do this classification of regressors and interactions without checking or knowing the solution to any of these systems of linear equations. These tools form the beginning of a theory of robust variable selection and network inference.

Robust variable selection and network inference are, based on (5.5.1), a question of solution of systems of linear equations and we therefore begin our analysis of the problem from this perspective. In general, a linear system of equations has no, one, or infinitely many solutions, see *e.g.* Anton and Rorres (2000); Friedberg et al. (2003); Horn and Johnson (1990). The system is said to be consistent if a solution exists and if it is the only solution, then it is unique and the selection of regressors needed for it is unique. Noise however complicates the situation, as illustrated in Section 4.2. One can say that for noisy data we have one system of equations for each consistent combination of regressors and regressand, and each of these systems has no, one, or infinitely many solutions. Note that we above used the term “consistent” in two different ways. When used about a system of equations, then it implies that at least one solution exist, as customary in linear algebra, see *e.g.* Anton and Rorres (2000); Friedberg et al. (2003). When used about uncertainty it implies that the combination of regressors and regressand agrees with the assumed error model. For noisy data, common numbers of data points in published biological data sets, and commonly used significance levels there is almost never only one unique solution that cannot be rejected; at the same time one unique solution is not necessary for variable selection, so we next introduce practical uniqueness. The term “practical” is borrowed from identifiability analysis, where it is used to indicate that the noise/error model is accounted for and we use it in the same manner—for reasons that we explain in detail later. On the other hand, it is rare that the observed regressand lies in the subspace spanned by the observed regressors for noisy data when the number of data points significantly exceeds the number of variables, so we also introduce practical existence.

In parameter estimation the least squares solution, which minimizes the sum of squared residuals, or some other solution based on minimizing or constraining

the residuals is typically used when a system of equations lack a solution, see *e.g.* Rao and Toutenburg (1999); Bingham and Fry (2010); Kay (1993); Ljung (1999). However, we do not consider solutions with nonzero residual, since each variable is represented by the set of realizations that are consistent with the error model and all parameter vectors except exact solutions to some combination of realizations are rejected based on the error model. Instead of searching for one solution, we are interested in the set of all solutions corresponding to a consistent combination of realizations, which we term “the set of feasible solutions”. The term “feasible” is borrowed from optimization theory, where it is used to denote all points that fulfil the constraints of the optimization problem, see *e.g.* Nash and Sofer (1996, p. 16). Assuming that both the observed and “true” variables are in the uncertainty sets, variables and interactions present in all models derived from the set of feasible solutions exist and are correct. Therefore we here formally introduce the notion of a feasible parameter vector, *i.e.* solution, of systems of linear equations with uncertain regressors and regressand and the set of feasible parameters/solutions for further analysis below.

**Definition 5.5.1. Feasible parameters/solutions.**

A parameter vector  $\boldsymbol{\theta} \triangleq [\theta_1, \dots, \theta_j, \dots, \theta_n]^T$  is feasible if

$$\begin{aligned} \sum_{j \in \mathcal{V}} \theta_j \tilde{\phi}_j = \tilde{\boldsymbol{\xi}} & \quad \text{for some consistent } \tilde{\phi}_j \in \mathcal{U}_{\phi_j}^\alpha \subseteq \mathbb{R}^m \\ \text{and } \tilde{\boldsymbol{\xi}} \in \mathcal{U}_{\boldsymbol{\xi}}^\alpha & \subseteq \mathbb{R}^m, \end{aligned} \quad (5.5.2)$$

*i.e.* if it is a solution to any of the systems of equations formed by the variables in the considered set  $\mathcal{V} = \{1, 2, \dots, n\}$  that are consistent with the error model at a selected significance level  $\alpha$ . The set of feasible parameters/solutions  $\mathcal{F}$  of (5.5.2) consists of all feasible parameter vectors  $\boldsymbol{\theta}$  at a selected significance level  $\alpha$ .

If no feasible solution exists, then the data model (5.3.2) is rejected based on the observed regressors and regressand at the selected significance level  $\alpha$ , implying that either the demanded significance level is too low ( $\alpha$  too large), or the assumed error model, *i.e.* (5.3.2a) combined with the uncertainty sets, or network model, *i.e.* (5.3.2b), is wrong. The set of feasible parameters specifies the set of models that cannot be rejected

$$\mathcal{M}_C \triangleq \left\{ \sum_{j \in \mathcal{V}} \theta_j \phi_j = \boldsymbol{\xi} \mid \boldsymbol{\theta} \in \mathcal{F} \right\}. \quad (5.5.3)$$

We say that these models are practically indistinguishable, since they generate input-output data, *i.e.* a regressor matrix and regressand, that is consistent with the error model. Classically two models are indistinguishable if the output is identical for every input, see *e.g.* Vajda (1981); Zhang et al. (1991); Chapman and Godfrey (1996) and references within. The purpose of the following paragraphs is to introduce concepts that enables us to determine if a variable must be selected and the corresponding interaction exists from the set of feasible solutions.

Existence of a feasible solution is clearly a necessary requirement for variable selection and network inference to be meaningful, since the data model otherwise is rejected, but the actual system of equations given by the observed regressors and regressand does not necessarily have a solution. We therefore introduce practical existence.

**Definition 5.5.2. Practical existence.**

A solution of (5.5.1) practically exists if  $\mathcal{F} \neq \emptyset$ , *i.e.* if some feasible parameter vector exists.

When a feasible solution exists then we also say that the system of equations is practically consistent, since at least one consistent system of equations that consists of a combination of regressors and regressand in agreement with the assumed error model exists. Practical existence therefore implies that the observed data can be explained within the considered model set, which in our case is the set of all linear models of the considered variables.

Let us next consider an extension of the uniqueness concept. If any system of equations has infinitely many solutions, then at least one parameter can take any value, including zero in some solution. This hampers variable selection and network inference, since no unique selection of regressors exists. We desire a unique selection, since it is equal to the “true” selection when both the observed and “true” variables are in the uncertainty sets. In general a unique selection of all regressors can only exist if each consistent system of equations has exactly one solution. Some parameter may however take the same value in all of the infinitely many solutions of a system of equations and hence be unique, despite the fact that some other parameter can take any value. Selection of a regressor corresponding to a unique parameter may be necessary to explain data even though a consistent system of equations with infinitely many solutions exists, so it is from a variable selection perspective beneficial to speak about uniqueness of parameters. We therefore next introduce practical uniqueness on a parameter basis and define a solution as practically unique if all parameters are practically unique. We here give the least restrictive definition that we can think of based on the feasible set of solutions, which ensures that a parameter cannot take every possible value.

**Definition 5.5.3. Practical uniqueness.**

A parameter  $\theta_k$  is with  $\mathcal{F} \neq \emptyset$  practically unique if there exists a constant  $c \in \mathbb{R}$ , such that  $\{\boldsymbol{\theta} \mid \boldsymbol{\theta} = [\theta_1, \dots, \theta_k, \dots, \theta_n]^T, \theta_k = c\} \cap \mathcal{F} = \emptyset$ , *i.e.* if no parameter vector  $\boldsymbol{\theta}$  with element  $k$  equal to  $c$  exists in a non-empty set of feasible solutions. A solution of (5.5.1) is practically unique if all parameters in it are practically unique.

This definition has the advantage that it is based directly on the parameter and set of feasible solutions, which makes it conceptually simple, but restricts its use to cases with a non-empty set of feasible solutions. We later give a more general definition after having developed necessary concepts. Practical uniqueness of a solution is equivalent to each consistent system of equations generated by the

uncertain regressors and regressand having exactly one solution, since if any system has infinitely many solutions then at least one of its parameters can take every possible value. Some parameter may, however, be practically unique even when the solution is not.

Practical uniqueness is closely related to practical, or *a posteriori*, identifiability, which previously has been used in parameter estimation to indicate that the uncertainty of the estimated parameters is within an application specific bound, see *e.g.* Ashyraliyev et al. (2009); Cedersund and Roll (2009); Gadkar et al. (2005); Faller et al. (2003); Zak et al. (2003), and we therefore include a definition of practical identifiability. Practical identifiability is a case specific noisy data analog to the more well known and more precisely defined structural, or *a priori*, identifiability defined in *e.g.* Ljung (1999); Bellman and Åström (1970). In our setting practical identifiability is defined based on the following bounds on the set of feasible solutions.

**Definition 5.5.4. Practical identifiability.**

A parameter  $\theta_k$  is for cases when  $\mathcal{F} \neq \emptyset$  practically identifiable if  $|\theta_k - \tilde{\theta}_k| < r_k \forall \theta, \tilde{\theta} \in \mathcal{F}$ , with  $\theta = [\theta_1, \dots, \theta_k, \dots, \theta_n]^T$ , and  $0 < r_k < \infty$  being an application specific constant, otherwise it is practically unidentifiable. The parameter vector  $\theta$  in (5.5.1) is practically identifiable if all parameters in it are practically identifiable.

When all parameters are practically identifiable then the set of feasible solutions is bounded and it is customary to say that the model is practically identifiable. For more information on bounded sets, see *e.g.* Khuri (2003, p. 9). The concepts that we develop later on could be used to give a more general definition, but practical identifiability is not needed for variable selection and we therefore leave it to future work on parameter estimation. Clearly practical uniqueness is necessary, but in general insufficient, for practical identifiability. The relation between practical uniqueness and identifiability is, however, the reason for why we have borrowed the term “practical”, and one may therefore use the term *a posteriori* instead of practical in any of our definitions. From a robust control perspective the term robust could be used instead of practical, since our usage is similar to that of robust stability and robust performance in *e.g.* Skogestad and Postlethwaite (1996).

Unfortunately neither practical uniqueness nor practical identifiability of systems of linear equations with uncertain regressors and regressand do in general imply that only one unique selection of regressors exist. This is easily understood if we, *e.g.*, assume that one parameter is zero in the “true” system model, *i.e.* some interaction does not exist in the “true” network, then the feasible set contain solutions with both positive, zero, and negative values for this parameter due to the uncertainty—given that the correct data model is assumed such that both the observed and “true” variables are in the uncertainty sets. Biological networks are sparse, *i.e.* lack some interactions, so actually this example implies that we in general cannot find a unique selection of the regressors or network model. From an invalidation perspective, it is for a finite number of experiments impossible to reject the existence of a very weak interaction. On the other hand, if all models lacking a certain interaction are

rejected, then the interaction is sufficiently strong to be inferred based on the given data set and the corresponding regressor is selected in all subsets of regressors for which the system of equations has a feasible solution. To distinguish between the regressors that must be selected for a system of equations with uncertain regressors and regressand to explain data and regressors that are not required, we introduce practical selectability.

**Definition 5.5.5. Practical selectability.**

A regressor  $\phi_k$  is for cases when  $\mathcal{F} \neq \emptyset$  practically selectable if  $\theta_k \neq 0$  in  $\boldsymbol{\theta} = [\theta_1, \dots, \theta_k, \dots, \theta_n]^T \forall \boldsymbol{\theta} \in \mathcal{F}$ , *i.e.* the corresponding parameter  $\theta_k$  is nonzero in every feasible solution, otherwise it is not practically selectable. The set of regressors in (5.5.1) is practically selectable if each regressor is practically selectable.

We later give a more general definition, since this one only can be used when the set of feasible solutions is non-empty, but we start with this one because it relates practical selectability to the values of the corresponding parameter in all feasible solutions and it can be stated based on existing concepts. We classify interactions corresponding to practically selectable regressors as existing if a solution practically exists, since they are present in all models that cannot be rejected and therefore exist at the selected confidence level.

**Definition 5.5.6. Existing interaction.**

An interaction exists if the corresponding regressor is practically selectable (Definition 5.5.5 or 5.6.1) and a solution of (5.5.1) practically exists (Definition 5.5.2).

If a regressor is practically selectable, then the corresponding parameter is clearly practically unique, since no parameter vector with  $\theta_k = 0$  belongs to the set of feasible solutions, but it is possible that the parameter is not practically identifiable. On the other hand, the corresponding parameter must clearly be practically unique for the regressor to be practically selectable, but this is not in general sufficient. We can for the assumed data model (5.3.2) at the selected significance level only be certain that interactions with practically selectable regressors exist. Practical selectability is thus the key to robust variable selection and robust network inference. In addition to characterizing parameters based on uniqueness, we also introduce the following classification. Note that the first class corresponds to practically selectable regressors and contains the second and third classes.

**Definition 5.5.7. Parameter classification.**

A parameter  $\theta_k$  in  $\boldsymbol{\theta} = [\theta_1, \dots, \theta_k, \dots, \theta_n]^T$  is for cases when  $\mathcal{F} \neq \emptyset$

- (a) practically nonzero if  $\theta_k \neq 0 \forall \boldsymbol{\theta} \in \mathcal{F}$ ,
- (b) practically positive if  $\theta_k > 0 \forall \boldsymbol{\theta} \in \mathcal{F}$ ,
- (c) practically negative if  $\theta_k < 0 \forall \boldsymbol{\theta} \in \mathcal{F}$ ,
- (d) practically zero if  $\theta_k = 0$  for some  $\boldsymbol{\theta} \in \mathcal{F}$ .

Parameters corresponding to practically selectable regressors are practically nonzero and the interaction is classified as existing if a solution of (5.5.1) practically exists, but it is often also of interest in network inference to determine if the interaction is activating or repressing, *i.e.* the signed structure of the network. This can only be done for cases when the parameter is practically positive or negative. To distinguish the regressors for which the sign of the corresponding parameter can be inferred from those for which it cannot we introduce practical assignability.

**Definition 5.5.8. Practical assignability.**

A regressor  $\phi_k$  is for cases when  $\mathcal{F} \neq \emptyset$  practically assignable if either  $\theta_k > 0 \forall \theta \in \mathcal{F}$  or  $\theta_k < 0 \forall \theta \in \mathcal{F}$ , with  $\theta = [\theta_1, \dots, \theta_k, \dots, \theta_n]^T$ , *i.e.* the corresponding parameter  $\theta_k$  is either practically positive or negative. The set of regressors in (5.5.1) is practically assignable if each regressor is practically assignable.

Every regressor that is practically assignable is also practically selectable, but the opposite is in general not true. If and only if a regressor is practically assignable, then we can classify the corresponding interaction as activating or repressing based on the sign of the corresponding parameter in any solution. In other words, if  $\theta_j > 0 \forall \theta \in \mathcal{F}$ , then the parameter is practically positive and we classify the corresponding interaction as activating, else if  $\theta_j < 0 \forall \theta \in \mathcal{F}$ , then the parameter is practically negative and we classify the corresponding interaction as repressing.

**Definition 5.5.9. Activating interaction.**

An interaction is activating if the corresponding regressor is practically assignable (Definition 5.5.8) and the parameter is shown to be positive in at least one feasible solution of (5.5.1) (Definition 5.5.1), *i.e.* the corresponding parameter is practically positive (Definition 5.5.7b).

**Definition 5.5.10. Repressing interaction.**

An interaction is repressing if the corresponding regressor is practically assignable (Definition 5.5.8) and the parameter is shown to be negative in at least one feasible solution of (5.5.1) (Definition 5.5.1), *i.e.* the corresponding parameter is practically negative (Definition 5.5.7c).

We say that the parameter is practically zero if it is zero in some solution, *i.e.* if neither of the first three conditions in Definition 5.5.7 apply. It might at first thought seem strange to say that a parameter is practically zero even though it *e.g.* may be positive in almost all solutions and even take relatively large values, but from a model rejection and robustness perspective the essential issue is that it is zero in one solution. Zero in one solution immediately implies that we cannot reject all network models lacking the corresponding interaction and hence not prove that it exists at the selected significance level. The parameter corresponding to a regressor that is not practically selectable is thus practically zero; we will later look at classification of these as practically excludable, alternative, or negligible and the corresponding interactions as non-existing, alternative or non-evidential. We will also give general definitions, that allow the parameters and regressors to be

classified as practically unique and selectable, respectively, even if no solution of (5.5.1) practically exists, in order to facilitate comparison of different data models at different significance levels. Practical assignability, on the other hand, is tightly linked to the sign of the corresponding parameter, so conditions that are both necessary and sufficient for practical assignability cannot be stated for all kinds of uncertainty sets unless a feasible solution exists and no extension of Definition 5.5.8 is attempted. To state these general definitions and establish conditions on the data for classification we first need to extend the notions of span, linear independence, and rank to uncertain data, which we do in the following section. These extensions will also enable us to rephrase the above definitions in terms of data conditions that actually can be calculated or checked. All definitions in this section are namely based on the set of feasible solutions, which in general cannot be calculated since it would require the solution of an infinite number of systems of equations. Use of the set of feasible solutions in these definitions however makes the concepts well defined, conceptually simple and easy to understand, and it connects them to established theory, in particular within parameter estimation. We next also derive conditions for practical existence and uniqueness, while conditions for practical selectability and assignability are left to the following section on practically selectable regressors (Section 5.6), and the treatment of regressors that are not practically selectable to the sections on practically excludable, negligible, and alternative regressors (Sections 5.7 and 5.8).

### 5.5.1 Extension of concepts from linear algebra

In the noise free case a solution to the linear system of equations in (5.5.1) only exists if the regressand lies in the subspace spanned by the regressors, and it is only unique if the regressors are linearly independent, see *e.g.* Anton and *Rorres* (2000); *Friedberg et al.* (2003); *Horn and Johnson* (1990). The concepts of span and linear independence provide insight and allow us to use simple tools such as rank and dimension to establish existence and uniqueness of a solution in the noise free case, so we extend them to noisy and uncertain data in this and the following three paragraphs. Several extensions, such as different measures of near collinearity (*Belsley*, 1991) and effective rank (*Konstantinides and Yao*, 1988; *Roy and Vetterli*, 2007; *Aksasse et al.*, 2006), already exist in the literature. We relate to them, but they only apply to certain error models and they are inadequate for robust variable selection and network inference, see Section A.1. We typically analyse a set of regressors and have therefore used  $\phi$  for vectors or  $\Phi$  for matrices in the following formulas, even when the definition or theorem applies to vectors in general. When the term regressor or regressand is used or we refer to (5.5.1) then the definition or result applies to them only. To simplify the notation and introduce the extended concepts, let us first define the set of all combinations of realizations of vectors that are consistent with the error model

$$\mathcal{V}_C \triangleq \left\{ \tilde{\Phi} \mid \tilde{\Phi} \triangleq [\tilde{\phi}_1, \dots, \tilde{\phi}_j, \dots, \tilde{\phi}_n], \text{ for consistent } \tilde{\phi}_j \in \mathcal{U}_{\phi_j}^\alpha \subseteq \mathbb{R}^m \right\}. \quad (5.5.4)$$

The span of a set of vectors in the matrix  $\tilde{\Phi}$  is defined as (Friedberg et al., 2003, p. 30)

$$\text{span } \tilde{\Phi} \triangleq \left\{ \sum_{j=1}^n \theta_j \tilde{\phi}_j \mid \theta_j \in \mathbb{R}, \tilde{\phi}_j \in \tilde{\Phi} \right\}. \quad (5.5.5)$$

Existence of a solution to at least one of the systems of equations in (5.5.1) that are consistent with the error model is necessary and sufficient for practical existence, so we introduce practical span in the following way to make its implications as analogous to ordinary span as possible. Practical span enables us to express exact conditions in a concise manner below and we use it extensively in our proofs later on.

**Definition 5.5.11. Practical span.**

The practical span of the set of uncertain vectors in the matrix  $\Phi = [\phi_1, \dots, \phi_j, \dots, \phi_n]$  is

$$\text{pspan } \Phi \triangleq \left\{ \sum_{j=1}^n \theta_j \tilde{\phi}_j \mid \theta_j \in \mathbb{R}, \tilde{\phi}_j \in \tilde{\Phi}, \tilde{\Phi} \in \mathcal{V}_C \right\}, \quad (5.5.6)$$

*i.e.* the set consisting of all linear combinations of all consistent combinations of vectors in the uncertainty sets.

This definition of practical span is equivalent to the union of the span for all realizations.

**Corollary 5.5.1.**

$$\text{pspan } \Phi = \bigcup_{\tilde{\Phi} \in \mathcal{V}_C} \text{span } \tilde{\Phi}. \quad (5.5.7)$$

Note that  $\text{pspan}(\Phi)$  in general is not a subspace of  $\mathbb{R}^m$  nor a vector space, even though  $\text{span}(\tilde{\Phi})$  is a subspace of  $\mathbb{R}^m$  for each  $\tilde{\Phi} \in \mathcal{V}_C$ , because  $\text{pspan}(\Phi)$  is a union of subspaces, which in general is not closed under addition, see *e.g.* Friedberg et al. (2003, p. 19). Many standard concepts of linear algebra are only defined for vector spaces, and  $\text{pspan}(\Phi)$  does *e.g.* not have a dimension in the ordinary sense. Despite this we can now state the condition for practical existence in terms of the practical span of the regressors in the matrix  $\Phi$ .

**Theorem 5.5.2.** *A solution of (5.5.1) practically exist (Definition 5.5.2) if and only if*

$$\tilde{\xi} \in \text{pspan } \Phi \quad \text{for some consistent } \tilde{\xi} \in \mathcal{U}_{\xi}^{\alpha} \subseteq \mathbb{R}^m. \quad (5.5.8)$$

Some consistent regressand must, in other words, lie in the subspace spanned by some consistent combination of regressors in order for a solution of (5.5.1) to

practically exist, but the observed regressand need not lie in the subspace of the observed regressors. The essence of this condition is, as we expected, the same as the previously mentioned condition for the noise free case. Practical span is also the concept we need in order to make a general definition of practical uniqueness, which is equivalent to Definition 5.5.3 for the cases covered by the latter, but it also applies to cases when no feasible solution exists.

**Definition 5.5.12. Practical uniqueness.**

A parameter  $\theta_k$  corresponding to regressor  $\phi_k$  in (5.5.1) is practically unique if,

$$\forall \text{ consistent } \tilde{\xi} \in \mathcal{U}_{\xi}^{\alpha} \subseteq \mathbb{R}^m \text{ such that } \tilde{\xi} \in \text{pspan } \Phi_{j \neq k}, \quad (5.5.9)$$

$$\tilde{\phi}_k \notin \text{pspan } \Phi_{j \neq k} \quad \forall \text{ consistent } \tilde{\phi}_k \in \mathcal{U}_{\phi_k}^{\alpha} \subseteq \mathbb{R}^m, \quad (5.5.10)$$

with  $\Phi_{j \neq k} = [\phi_1, \dots, \phi_{k-1}, \phi_{k+1}, \dots, \phi_n]$ .

**Theorem 5.5.3.** Definition 5.5.12 is equivalent to Definition 5.5.3 when a solution of (5.5.1) practically exists (Definition 5.5.2) and a constant  $0 < r_{\xi} < \infty$  exists, such that

$$\|\dot{\xi} - \hat{\xi}\| \leq r_{\xi} \quad \forall \text{ consistent } \dot{\xi}, \hat{\xi} \in \mathcal{U}_{\xi}^{\alpha} \subseteq \mathbb{R}^m. \quad (5.5.11)$$

Definition 5.5.12 guarantees that regressor  $\phi_k$  is not, for any of the realizations that are consistent with the error model and span a subspace that contains a realization of the regressand, a linear combination of the other regressors. Not even when no feasible solution of (5.5.1) exists. Because  $\phi_k$  is not a linear combination of the other regressors, none of them can replace it and the corresponding parameter  $\theta_k$  is unique for all realizations such that the system of equations has a solution. On the other hand, it allows  $\phi_k$  to be a linear combination of the other regressors for all cases when no solution can exist, so the condition is necessary for what we understand as uniqueness of a parameter in a solution of a system of equations. This definition disconnects practical existence from practical uniqueness and has the advantage that a parameter that is practically unique at a selected significance  $\alpha$  also is practically unique at any significance  $\bar{\alpha} \geq \alpha$ , which simplifies comparison of different data models, data sets, *etc.*

Let us now extend linear independence, since it is necessary and sufficient for uniqueness of a solution of a system of equations in the noise free case and hence it should give us more insight on practical uniqueness. We introduce practical independence and use it to find the regressors that are practically selectable based on observed data. A set of vectors in the matrix  $\Phi = [\phi_1, \dots, \phi_j, \dots, \phi_n]$  is defined as linearly independent, see *e.g.* Friedberg et al. (2003, p. 36-37), if the trivial solution  $\theta = [\theta_1, \dots, \theta_j, \dots, \theta_n]^T = \mathbf{0}$  is the only solution of the homogeneous system of linear equations

$$\sum_{j=1}^n \theta_j \phi_j = \mathbf{0}. \quad (5.5.12)$$

We introduce practical independence as an extension of linear independence by requiring this to hold for every combination of vector realizations that are consistent with the error model.

**Definition 5.5.13. Practical independence.**

The set of uncertain vectors in the matrix  $\tilde{\Phi} = [\phi_1, \dots, \phi_j, \dots, \phi_n]$  is practically (linearly) independent, if, for all consistent  $\tilde{\phi}_j \in \mathcal{U}_{\phi_j}^\alpha \subseteq \mathbb{R}^m$ , the trivial solution  $\theta = [\theta_1, \dots, \theta_j, \dots, \theta_n]^T = \mathbf{0}$  is the only solution of

$$\sum_{j=1}^n \theta_j \tilde{\phi}_j = \mathbf{0}. \quad (5.5.13)$$

We say that a vector is practically independent if it cannot be expressed as a linear combination of the other vectors in the set for any consistent realization. Practical independence can also be expressed in terms of the span of all consistent realizations.

**Corollary 5.5.4. Practical independence.**

The set of uncertain vectors in the matrix  $\tilde{\Phi}$  is practically (linearly) independent, if and only if

$$\dim \text{span } \tilde{\Phi} = n \quad \forall \tilde{\Phi} \in \mathcal{V}_C \quad (5.5.14)$$

or equivalently

$$\begin{aligned} \tilde{\phi}_k \notin \text{pspan } \tilde{\Phi}_{j \neq k} & \quad \forall \text{ consistent } \tilde{\phi}_k \in \mathcal{U}_{\phi_k}^\alpha \subseteq \mathbb{R}^m \\ & \quad \text{and } k \in \{1, 2, \dots, n\}, \end{aligned} \quad (5.5.15)$$

with  $\tilde{\Phi}_{j \neq k} \triangleq [\phi_1, \dots, \phi_{k-1}, \phi_{k+1}, \dots, \phi_n]$ .

Practical independence of a set of regressors implies that no regressor in the set can be expressed as a linear combination of the other regressors for any consistent realization and it is therefore straightforward to show that practical independence is sufficient for practical uniqueness, as formalized in the following theorem.

**Theorem 5.5.5.** All parameters  $\theta_j$  in (5.5.1) are practically unique (Definition 5.5.12) if the regressors are practically independent (Definition 5.5.13).

In short, we can now say that the parameters are unique if the corresponding regressors are practically independent. Note that practical independence of the regressors is strictly speaking not necessary for practical uniqueness, because a solution may not exist for all combinations of regressors and regressand and the existing solutions can be unique even though some of these regressor combinations are not linearly independent. Practical independence of the regressors is however necessary for the set of regressor combinations for which the system of equations has a solution. In principle, we could have defined practical independence such that

only this set of regressor realizations is required to be linearly independent, but then we would first need to determine them, which is as difficult as determining the set of feasible solutions or practical uniqueness directly using Definition 5.5.12. In other words, it would not be as useful nor would it be a simple extension of linear independence. If the set of regressors for any of these realizations is linearly dependent, then at least one of the parameters can take any value and hence lack a unique solution.

Regression instability is often experienced when solving a regression problem with some linearly dependent set of regressors in  $\mathcal{V}_C$ , *i.e.* a small change in the data can lead to a large change in the estimated model, and these inverse problems are then referred to as ill-posed or ill-conditioned, see *e.g.* Aster et al. (2005). Ill-conditioning, strong correlation, multicollinearity, and near collinearity have been extensively studied in order to avoid or at least detect sets of regressors that are collinear, *i.e.* linearly dependent, see *e.g.* Belsley (1991); Stewart (1987); Liao (2010); Alin (2010). However, at best rules of thumb are given for some measure of the degree of near collinearity that may cause instability, or, in our terminology, a solution that is not practically unique, so the conditioning problem remains open. This ambiguity is resolved by practical independence as introduced above and the following introduction of practical collinearity, because near collinearity of the regressors cannot affect the solution at the selected significance level if the regressors are practically independent, while it affects the solution when the regressors are practically collinear.

**Definition 5.5.14. Practical collinearity.**

The set of uncertain vectors in the matrix  $\Phi = [\phi_1, \dots, \phi_k, \dots, \phi_n]$  is practically collinear, or practically (linearly) dependent, if for all consistent  $\phi_k \in \mathcal{U}_{\phi_k}^\alpha$  of some regressor  $\phi_k$  with  $k \in \{1, 2, \dots, n\}$ , one can find consistent  $\tilde{\phi}_j \in \mathcal{U}_{\phi_j}^\alpha \subseteq \mathbb{R}^m$  for all  $j \neq k$  such that a non-trivial solution  $\theta = [\theta_1, \dots, \theta_k, \dots, \theta_n]^T \neq \mathbf{0}$  of

$$\sum_{j=1}^n \theta_j \tilde{\phi}_j = \mathbf{0} \quad (5.5.16)$$

exists.

This definition is equivalent to every realization of one of the vectors being in the practical span of the other vectors.

**Corollary 5.5.6. Practical collinearity.**

*The set of uncertain vectors in the matrix  $\Phi$  is practically collinear if and only if*

$$\tilde{\phi}_k \in \text{pspan } \Phi_{j \neq k} \quad \forall \text{ consistent } \tilde{\phi}_k \in \mathcal{U}_{\phi_j}^\alpha \quad (5.5.17)$$

*for some regressor  $\phi_k$  with  $\Phi_{j \neq k} = [\phi_1, \dots, \phi_{k-1}, \phi_{k+1}, \dots, \phi_n]$ .*

We say that a vector is practically collinear if every consistent realization of it can be expressed as a linear combination of the other vectors for some consistent combination

of realizations. Practical collinearity is the opposite of practical independence in the sense that linearly dependent realizations exist for each consistent realization of one of the vectors in the former, while every realization is linearly independent in the latter. In general, sets of vectors that are neither practically collinear nor practically independent exist, but the set of practically collinear vectors is always a subset of the set of vectors that is not practically independent. Practical collinearity of the set of regressors is sufficient for regression instability of the solution of a system of equations at a selected significance level, while near collinearity merely indicates that instability may occur, as evident in [Belsley \(1991\)](#). Practical independence, on the other hand, is sufficient for stable solution, despite small changes in the data, and ensures robustness in both variable selection and network inference as we show later, as well as in parameter estimation. Many feasible solutions with different sign would exist if a consistent realization such that the set of regressors is linearly dependent exists. In this case, the regressors are clearly not practically assignable nor practically selectable and the corresponding parameters can take any value. Moreover, the absolute value of these parameters is large in many of the feasible solutions due to near collinearity, and the common belief that an interaction is significant if its parameter estimate has a large absolute value is misleading. For a set of regressors that is not practically independent, no relation in general exists between the absolute value of a parameter estimate and existence of the corresponding interaction, and statistical hypothesis tests based on the absolute value of the parameter estimate should not be used, because they in general do not hold due to neglect of practical collinearity.

In the noise free case, the rank of a matrix  $\text{rank}(\Phi)$  is defined as the largest number of columns of  $\Phi \in \mathbb{R}^{m \times n}$  that constitute a linearly independent set ([Horn and Johnson, 1990](#), p. 12). This number is equal to the dimension of the subspace generated or spanned by the columns ([Friedberg et al., 2003](#), p. 67-70,153). Consequently, if  $\text{rank}(\Phi) = m$  or  $\text{rank}([\Phi \ \xi]) = \text{rank}(\Phi)$ , then a solution of  $\Phi\theta = \xi$  exists, and if  $\text{rank}(\Phi) = n$ , then any existing solution is unique, see [Friedberg et al. \(2003](#), p. 153,174) and [Horn and Johnson \(1990](#), p. 13-14). This makes rank a useful tool for calculation that provides insight and we therefore here extend the concept of rank to noisy data by introducing practical rank to complete our extensions of basic concepts in linear algebra, even though it is not necessary for the classification of regressors and interactions. When a set of uncertain vectors is practically independent then each realization of the matrix has rank  $n$  and we define this to be the practical rank, but otherwise the situation is more complicated. The only foreseeable way to always obtain robust results is to implicitly account for the set of feasible solutions by introducing practical rank in the following manner, which also implies that it can be used to assess both practical existence and uniqueness.

**Definition 5.5.15. Practical rank.**

The practical rank of the uncertain matrix  $\Phi = [\phi_1, \dots, \phi_j, \dots, \phi_n]$  is the set

$$\text{prank } \Phi \triangleq \{p \mid \text{rank } \tilde{\Phi} = p \text{ for some } \tilde{\Phi} \in \mathcal{V}_C\}. \quad (5.5.18)$$

The practical rank is, in other words, a set containing the rank of all realizations that are consistent with the error model. The rank of the “true” matrix is therefore in this set if the “true” vectors are consistent with the error model, as formalized in the following theorem.

**Theorem 5.5.7.**

$$\text{rank } \check{\Phi} \in \text{prank } \Phi \quad \text{if } \check{\Phi} \in \mathcal{V}_C. \quad (5.5.19)$$

Practical independence of a set of vectors is equivalent to having a minimal practical rank equal to the number of vectors.

**Corollary 5.5.8.** *The set of uncertain vectors in the matrix  $\Phi$  is practically (linearly) independent, if and only if*

$$\text{prank } \Phi = \{n\}. \quad (5.5.20)$$

We will also use the notion robust rank when the practical rank is equal to the number of columns.

The concept of rank has previously been extended to noisy data by introduction of the so called effective rank (Konstantinides and Yao, 1988; Roy and Vetterli, 2007). The idea behind effective rank is to determine the rank of the latent “true” matrix  $\check{\Phi}$  from the observed matrix  $\Phi$ . The latter is assumed to contain additive errors  $\Phi = \check{\Phi} + \Upsilon$ , with  $\Upsilon$  denoting the errors, as in our data model (5.3.2a). We have previously used effective rank for network inference and found it useful in some cases (Nordling and Jacobsen, 2011). It is, however, in general unsuitable for network inference, because the use of effective rank depends critically on a sufficient separation between the singular values of  $\check{\Phi}$  and  $\Upsilon$ , see Section A.1, which typically does not hold for biological data. The effective rank differs in general from the “true” rank unless the singular values of  $\check{\Phi}$  are all larger than the largest singular value of  $\Upsilon$ . Even if one succeeded in determining the rank of the latent matrix, the result would in general be sensitive to small changes in the data and the robustness that we strive towards would be lacking. Contrary to the effective rank, the practical rank always contains the “true” rank. The practical rank is a set and thus conceptually different from the effective rank, which enables robust results even when the effective rank is influenced by noise.

We have hereby extended the necessary concepts of linear algebra to noisy data and established a set of tools that we next use for classification of the regressors and interactions. The practical span is mainly used to state and prove conditions that are both necessary and sufficient; while practical independence and collinearity are mainly used to establish sufficient conditions and gain insight. How to check practical independence based on observed data is presented later in the section on testing of practical independence (Section 5.10).

## 5.6 Practically selectable regressors

We say that a regressor is practically selectable if and only if it must be selected to explain data, which makes it the key concept in robust variable selection. Only interactions corresponding to practically selectable regressors can be proven to exist, so it is also the key concept in robust network inference. We here use the tools developed in the previous section to first make a general definition of practical selectability and then establish both sufficient and necessary conditions on data for practical selectability. Since the sign of the interactions often also are of interest, we finally establish data conditions for when the sign can be determined with desired confidence, corresponding to practical assignability of the regressor.

We introduced practical selectability based on the set of feasible solutions in Definition 5.5.5, but use of that definition requires knowledge of the feasible set and is restricted to cases in which the feasible set is non-empty. We therefore start by giving a definition of practical selectability based directly on data, using the concept of practical span that we introduced in the previous section. The two definitions are equivalent when a feasible solution exists, as we show below, but the one based on practical span can also be used when no feasible solution exists. It is hence a generalization of the previous one. It also ensures that a regressor proven to be practically selectable at significance level  $\alpha$  is practically selectable for all  $\tilde{\alpha} \geq \alpha$  and we therefore use it henceforth.

### Definition 5.6.1. Practical selectability.

A regressor  $\phi_k$  in (5.5.1) is practically selectable if

$$\tilde{\xi} \notin \text{pspan } \Phi_{j \neq k} \quad \forall \text{ consistent } \tilde{\xi} \in \mathcal{U}_{\tilde{\xi}}^{\alpha} \subseteq \mathbb{R}^m \quad (5.6.1)$$

with  $\Phi_{j \neq k} = [\phi_1, \dots, \phi_{k-1}, \phi_{k+1}, \dots, \phi_n]$ , otherwise it is not practically selectable. The set of regressors in (5.5.1) is practically selectable if each regressor is practically selectable.

**Theorem 5.6.1.** Definition 5.6.1 is equivalent to Definition 5.5.5 when a solution of (5.5.1) practically exists (Definition 5.5.2).

A practically selectable regressor and the corresponding interaction must be present to explain data within a considered model set. Intuitively the number of interactions that can be proven to exist should decrease with increasing uncertainty. We here classify interactions corresponding to practically selectable regressors as existing so let us consider how selectability is affected by increasing uncertainty. If the size of the uncertainty set of the regressand is increased, then (5.6.1) is in general fulfilled in fewer cases. Similarly, if the size of any of the uncertainty sets of the regressors in  $\Phi_{j \neq k}$  is increased, then the set given by  $\text{pspan}(\Phi_{j \neq k})$  is increased and (5.6.1) is in general fulfilled in fewer cases, *i.e.* some  $\tilde{\xi}$  is more likely to be in  $\text{pspan}$ . This implies that our definition of practical selectability in general leads to a decrease in the number of existing interactions with increasing uncertainty and decreasing significance  $\alpha$ , assuming that the size of the uncertainty sets increases

with decreasing  $\alpha$ . This also implies that if a regressor is not practically selectable when no solution practically exists, then it is for sure not practically selectable when one exists, since the error model or significance level must be changed so that the uncertainty sets are increased to enable a feasible solution of (5.5.1). Note that we do not need any explicit condition on the regressor  $\phi_k$ , since a solution of (5.5.1) only exists for realizations such that  $\tilde{\phi}_k \notin \text{span}(\tilde{\Phi}_{j \neq k})$  and  $\tilde{\xi} \in \text{span}(\tilde{\Phi})$ , so (5.6.1) is sufficient. This is also why a regressor is classified as practically selectable based on Definition 5.6.1 even if no feasible solution exists. The condition in (5.6.1) provides a simple definition of practical selectability for all imaginable uncertainty sets and error models, including discrete sets. Evaluation of the practical span is however in general hard, so we provide the following sufficient condition, which in our opinion also provides more insight.

**Theorem 5.6.2. Practical selectability.**

A regressor  $\phi_k$  in (5.5.1) is practically selectable (Definition 5.6.1) if the set of vectors in the matrix  $\Psi_k \triangleq [\phi_1, \dots, \phi_{k-1}, \phi_{k+1}, \dots, \phi_n, \xi]$  is practically independent (Definition 5.5.13).

Practical independence implies that none of the vectors is in the practical span of the other vectors, but in (5.6.1) only the regressand is required not to be, so this condition is in general conservative. However, for data that contain uncertain measurements of continuous variables it is in practice necessary, as shown in the following theorem.

**Theorem 5.6.3.** Assume that each element of the regressors and regressand is uncertain such that

$$\mathcal{B}_\rho(\phi_j) \subseteq \mathcal{U}_{\phi_j}^\alpha \subseteq \mathbb{R}^m, \quad \forall j \in \mathcal{V} = \{1, 2, \dots, n\}, \quad (5.6.2)$$

$$\mathcal{B}_\rho(\xi) \subseteq \mathcal{U}_\xi^\alpha \subseteq \mathbb{R}^m, \quad (5.6.3)$$

i.e. some neighbourhood of each observed variable is contained in its uncertainty set, that all possible combinations within the neighbourhoods are consistent with the error model, and that

$$\tilde{\xi} \in \text{span } \tilde{\Phi} \quad \text{for some } \tilde{\xi} \in \mathcal{B}_\rho(\xi) \text{ and } \tilde{\Phi} \in \mathcal{B}_\rho(\Phi) \subseteq \bigcup_{j \in \mathcal{V}} \mathcal{B}_\rho(\phi_j), \quad (5.6.4)$$

i.e. a solution practically exists within the neighbourhood. Then a regressor  $\phi_k$  in (5.5.1) can only be practically selectable (Definition 5.6.1) if the set of vectors in the matrix  $\Psi_k = [\phi_1, \dots, \phi_{k-1}, \phi_{k+1}, \dots, \phi_n, \xi]$  is practically independent for all realizations within the neighbourhoods. Here the neighbourhood of a vector  $\phi_j$  is defined as

$$\mathcal{B}_\rho(\phi_j) \triangleq \{\tilde{\phi}_j \mid \|\tilde{\phi}_j - \phi_j\| < \rho, \rho > 0, \rho \in \mathbb{R}\}, \quad (5.6.5)$$

and the neighbourhood of a matrix  $\Phi$ , using the corresponding induced matrix norm ([Horn and Johnson, 1990](#), p. 290-295), is defined as

$$\mathcal{B}_\rho(\Phi) \triangleq \{\tilde{\Phi} \mid \|\tilde{\Phi} - \Phi\| < \rho, \rho > 0, \rho \in \mathbb{R}\}. \quad (5.6.6)$$

All biological data sets relevant for network inference that we are aware of contain uncertainty in each element of the regressors and regressand so that at least all combinations of values within a small ball or neighbourhood of the observed values are consistent with the error model. Therefore the set of all vectors in  $\Psi_k$  in practice needs to be practically independent in order for the  $k$ th regressor to be practically selectable. It is trivial to prove that non-rejectable models with and without the interaction corresponding to the  $k$ th regressor, as well as with different signs of the corresponding parameter, can only exist if a realization of  $\Psi_k$  with collinear columns exists.

**Theorem 5.6.4. Collinearity affects selectability.**

A parameter  $\theta_k$  can only have opposite sign in two different non-rejectable models if a consistent realization of  $\Psi_k = [\phi_1, \dots, \phi_{k-1}, \phi_{k+1}, \dots, \phi_n, \xi]$  with collinear columns exists.

The sufficient condition for practical selectability in [Theorem 5.6.2](#) guarantees that we avoid this ambiguity of selection and inference. In other words, in practice we can only be sure that a variable must be selected and that the corresponding interaction exists if no matrix  $\Psi_k$  with collinear columns is consistent with the error model. This is something that can be checked without knowing the set of feasible solutions, as we show later in [Section 5.10](#). Note that practical independence of  $\Psi_k$  implies that any connected uncertainty set of the included variables is bounded in some directions, because the uncertainty sets otherwise would overlap and some linearly dependent realization would exist. We next state the condition in [Theorem 5.6.2](#) in terms of the practical rank, since it highlights that practical selectability is qualitatively a question of rank.

**Corollary 5.6.5.** A regressor  $\phi_k$  in [\(5.5.1\)](#) is practically selectable ([Definition 5.6.1](#)) if

$$\text{prank } \Psi_k = \{n\}, \quad \text{with} \\ \Psi_k = [\phi_1, \dots, \phi_{k-1}, \phi_{k+1}, \dots, \phi_n, \xi]. \quad (5.6.7)$$

The rank of a matrix can never exceed the number of columns or rows, so at least as many data points as variables are in practice required for any regressor to be practically selectable and for the assignment of confidence to any inferred interaction, as shown in [Theorem A.3.2](#). This coincides with the well known requirement of  $m \geq n$  for a unique solution of a system of  $m$  equations and  $n$  unknown parameters to even be possible, see *e.g.* [Ashyraliyev et al. \(2009\)](#). At first sight this appears to be in conflict with recent observations and results suggesting that less than  $n$  data points are required for correct variable selection, estimation of a parameter vector, or

inference of a network, see *e.g.* Candès and Plan (2009); Fan and Lv (2010); Candès and Wakin (2008); Zhao and Yu (2006); Filkov (2005). But no conflict actually exists, since it is one thing to find the correct model and another to show that it is correct, *i.e.* to assign confidence, as discussed in more detail in Section 4.4.

### 5.6.1 Practically assignable regressors

In network inference the sign of the interaction is often also of interest, so here we present data conditions for a parameter to be practically positive or negative so that the corresponding interaction can be classified as activating or repressing, respectively. This classification is only possible if the corresponding regressor is practically assignable—a concept that we introduced in Definition 5.5.8 based on the set of feasible solutions. The idea here is to derive data conditions for practical assignability that can be checked without knowledge of the set of feasible solutions, since the set in general cannot be calculated and calculation of a single solution is sufficient to determine the sign of the interaction if we have established that the regressor is practically assignable. To simplify the presentation of the following results we first introduce the projection matrix, which is defined as (Friedberg *et al.*, 2003, p. 398-400)

$$\mathbf{T}_{j \neq k} \triangleq \Phi_{\mathcal{S}} (\Phi_{\mathcal{S}}^T \Phi_{\mathcal{S}})^{-1} \Phi_{\mathcal{S}}^T = \Phi_{j \neq k} \Phi_{j \neq k}^{\dagger} \in \mathbb{R}^{m \times m}. \quad (5.6.8)$$

This matrix projects any vector in  $\mathbb{R}^m$  onto the subspace spanned by the  $n - 1$  regressors in the matrix  $\Phi_{j \neq k} = [\phi_1, \dots, \phi_{k-1}, \phi_{k+1}, \dots, \phi_n]$  along a vector orthogonal to the subspace. To ensure that the matrix inverse exists, the matrix  $\Phi_{\mathcal{S}}$  only consists of a subset  $\mathcal{S}$  of the regressors in  $\Phi_{j \neq k}$ , which form a basis for this subspace; alternatively the Moore-Penrose generalized inverse  $\Phi_{j \neq k}^{\dagger}$  can be used (Horn and Johnson, 1990, p. 421). The columns of  $\Phi_{\mathcal{S}}$  are thus linearly independent, while the use of all  $n - 1$  regressors in general may lead to a matrix with linearly dependent columns. Further on we use the projection matrix to define hyperplanes that separate the regressand realizations that could lead to a positive parameter, *i.e.* the positive halfspace, from those that could lead to a negative parameter, *i.e.* the negative halfspace. For information on hyperplanes, separation of sets, and halfspaces see *e.g.* Boyd and Vandenberghe (2004, p. 27-29, 46-49). For the moment, we start by using the projection matrix to state necessary and sufficient conditions for practical assignability in a manner that does not require knowledge of the set of feasible solutions.

**Theorem 5.6.6.** *A regressor  $\phi_k$  is practically assignable (Definition 5.5.8) if and only if either*

$$\tilde{\mathbf{h}}_k^T \tilde{\boldsymbol{\xi}} > 0 \quad \forall \text{ consistent } \tilde{\boldsymbol{\xi}} \in \text{pspan } \Phi \quad (5.6.9a)$$

or

$$\tilde{\mathbf{h}}_k^T \tilde{\boldsymbol{\xi}} < 0 \quad \forall \text{ consistent } \tilde{\boldsymbol{\xi}} \in \text{pspan } \Phi \quad (5.6.9b)$$

and a solution of (5.5.1) practically exists (Definition 5.5.2). The normal of the hyperplane of the other regressors, corresponding to the realization  $\tilde{\xi}$ , is defined as

$$\tilde{\mathbf{h}}_k \triangleq (\mathbf{I} - \tilde{\mathbf{T}}_{j \neq k}) \tilde{\phi}_k \quad \text{for any } \tilde{\Phi} \in \mathcal{V}_C \text{ such that } \tilde{\Phi} \boldsymbol{\theta} = \tilde{\xi} \text{ has a solution,} \quad (5.6.10)$$

and  $\tilde{\mathbf{T}}_{j \neq k}$  is the projection matrix defined in (5.6.8).

The conditions in Theorem 5.6.6 ensures that the parameter corresponding to regressor  $\phi_k$  is either practically positive or negative in all feasible solutions. To check these necessary conditions we, however, still have to know the regressand realizations for which a feasible solution exists, hence we next provide sufficient conditions for a regressor to be practically assignable, which only require us to prove existence of one feasible solution and fulfilment of (5.6.9) for all consistent realizations. In other words, we do not need any more to know which combinations of realizations yield a feasible solution.

**Theorem 5.6.7.** *A regressor  $\phi_k$  is practically assignable (Definition 5.5.8) if a solution of (5.5.1) practically exists (Definition 5.5.2) and*

- (a) *for all consistent  $\tilde{\xi} \in \mathcal{U}_{\xi}^{\alpha} \subseteq \mathbb{R}^m$  and  $\tilde{\phi}_j \in \mathcal{U}_{\phi_j}^{\alpha} \subseteq \mathbb{R}^m$ , either  $\tilde{\mathbf{h}}_k^T \tilde{\xi} > 0$  or  $\tilde{\mathbf{h}}_k^T \tilde{\xi} < 0$ , with the normal of the hyperplane of the other regressors defined in (5.6.10), or*
- (b) *the uncertainty set of the regressand  $\mathcal{U}_{\xi}^{\alpha} \subseteq \mathbb{R}^m$  lies in either the intersection of all positive halfspaces*

$$\mathcal{H}_+ \triangleq \bigcap_{\tilde{\mathbf{h}}_k \text{ for consistent } \tilde{\phi}_j \in \mathcal{U}_{\phi_j}^{\alpha}} \{ \tilde{\mathbf{z}} \mid \tilde{\mathbf{z}}^T \tilde{\mathbf{h}}_k > 0, \tilde{\mathbf{z}} \in \mathbb{R}^m \} \quad (5.6.11)$$

*or of all negative halfspaces*

$$\mathcal{H}_- \triangleq \bigcap_{\tilde{\mathbf{h}}_k \text{ for consistent } \tilde{\phi}_j \in \mathcal{U}_{\phi_j}^{\alpha}} \{ \tilde{\mathbf{z}} \mid \tilde{\mathbf{z}}^T \tilde{\mathbf{h}}_k < 0, \tilde{\mathbf{z}} \in \mathbb{R}^m \} \quad (5.6.12)$$

*with  $\tilde{\mathbf{h}}_k$  defined in (5.6.10).*

The sets  $\mathcal{H}_+$  in (5.6.11) and  $\mathcal{H}_-$  in (5.6.12) are convex cones, because every halfspace is a convex set and the intersection of convex sets is a convex set (Boyd and Vandenberghe, 2004, p. 23,25,36). In other words, a parameter is practically positive or negative if the uncertainty set of the regressand is contained in either of the two convex cones. If the uncertainty sets are discrete, then condition (a) is preferable, because it is less conservative than (b), while condition (b) is preferable for continuous variables, because convex optimization techniques can be used. The development of techniques for calculation of these convex sets and checking of practical assignability based on them is however left to future work. For cases when the variables are

continuous and the uncertainty set of each variable is connected then sufficient conditions can be stated in a manner similar to the ones for practical selectability in Theorem 5.6.2. For information on connected sets see *e.g.* Khuri (2003, p. 12).

**Theorem 5.6.8.** *A regressor  $\phi_k$  in (5.5.1) is practically assignable (Definition 5.5.8) if the set of uncertain regressors in the matrix  $\Phi = [\phi_1, \dots, \phi_k, \dots, \phi_n]$  is practically independent, the set of all regressors except  $\phi_k$ , which is replaced by the regressand  $\xi$ , in the matrix  $\Psi_k \triangleq [\phi_1, \dots, \phi_{k-1}, \phi_{k+1}, \dots, \phi_n, \xi]$  is practically independent, and one of the following holds:*

- (a) *The uncertainty sets are connected and  $m = n$ .*
- (b) *The set of regressand realizations  $\tilde{\xi}$  for which a solution of  $\tilde{\Phi}\theta = \tilde{\xi}$  exists for some consistent  $\tilde{\xi} \in \mathcal{U}_{\xi}^{\alpha} \subseteq \mathbb{R}^m$  and  $\tilde{\Phi} \in \mathcal{V}_C$  is a non-empty connected subset of  $\mathbb{R}^m$ .*

A practically assignable regressor must always be practically selectable. We know from Theorem 5.6.2 that practical independence of  $\Psi_k$  is sufficient for practical selectability, so the additional conditions are only needed to prove that  $\theta_k$  has the same sign in all feasible solutions. From (a) above we learn that the additional requirement of practical independence of  $\Phi$  is sufficient when the number of experiments equals the number of variables and the uncertainty sets are connected. When  $m > n$  then this is not in general sufficient because the  $k$ th regressor or regressand can figuratively speaking encircle the subspace spanned by the other regressors. Many different conditions can be constructed that ensure that no such encircling takes place, and the condition given in (b) is just one example that is sufficient for certain cases. Others should be constructed, but we leave it for future work. We can argue based on Theorem 5.6.3 that the practical independence conditions in practice are also necessary for continuous variables, so the conservativeness is essentially determined by the additional conditions.

In conclusion, we now have the necessary results for showing that a variable must be selected to explain data, and for the classification of an interaction as existing and activating or repressing, at least for certain cases.

## 5.7 Practically excludable regressors

So far the focus has been on determining existing interactions and selecting variables that are necessary to explain data, *i.e.* determining when a regressor is practically selectable. Now we will instead look at the determination of non-existing interactions and variables that always can be excluded while explaining data, *i.e.* when a regressor is practically excludable. In other words, the regressors that previously were classified as not being practically selectable are now further analysed and a subset of them are classified as practically excludable. Naturally, we also establish sufficient and necessary data conditions for this classification.

We previously introduced practical selectability of a regressor based on the corresponding parameter being nonzero in every feasible solution of (5.5.1), and it might seem logical to introduce practical excludability based on the parameter being zero in every feasible solution. Such a definition would however be almost useless, because no parameter is in practice zero in every feasible solution due to uncertainty. It is impossible to prove, based on a finite data set of continuous variables, that an interaction does not exist because either the whole data model is rejected or then some model with a weak interaction cannot be rejected. A model with a weak interaction is practically indistinguishable from the model without the interaction. Moreover, when the wrong error model is used or the error model is not properly accounted for in parameter estimation then parameters that should be zero are typically assigned nonzero values in the estimation. This is known as over-fitting, because the parameters are fitted to the specific noise realization of the given data (Cedersund and Roll, 2009; Ljung, 1999, p. 501). On the other hand, the simplest model that can explain data is in general preferable in science and, in particular, in engineering, since it is easier to analyse, comprehend, and use than more complex models, while still being good enough as an approximation to explain data. Typically the accuracy of estimated parameters and predictions is also better, because a simple model in general is less flexible than a more complex model and thereby has fewer unidentifiable parameters and is less prone to over-fitting. A more complex model can sometimes be motivated based on *a priori* knowledge and assumptions, but never based on given data. Preference of the simplest model is also known as the principle of parsimony or Occam's razor, see *e.g.* Cedersund and Roll (2009); Ljung (1999, p. 492). The idea of variable selection is to find variables that are necessary or important for the explanation of data, and the complement can be phrased as finding variables that are unnecessary or unimportant for the explanation of data. The important variables should be included in the model, while the unimportant should be excluded in order to keep it simple. So even though we never can prove that a variable should not be selected, we need to recognize unimportant variables and exclude them. At the same time we should only exclude variables that always can be excluded while explaining available data. The essence of network inference is to separate between variables that are needed to explain data and those that are not needed, since the former correspond to existing interactions and the later to non-existing. We therefore here first introduce practical omissibility and then, based on the former, practical excludability, which captures exactly the variables that can always be excluded while explaining data.

**Definition 5.7.1. Practical omissibility.**

A regressor  $\phi_k$  is practically omissible if

$$\tilde{\xi} \in \text{pspan } \Phi_{j \neq k} \quad \forall \text{ consistent } \tilde{\xi} \in \mathcal{U}_{\xi}^{\alpha} \subseteq \mathbb{R}^m \quad (5.7.1)$$

with  $\Phi_{j \neq k} = [\phi_1, \dots, \phi_{k-1}, \phi_{k+1}, \dots, \phi_n]$ . The set of regressors in (5.5.1) is practically omissible if each regressor is practically omissible.

A practically omissible regressor is not needed to explain the data, since every consistent realization of the regressand can be explained by some realization of the other regressors. The difference between a practically omissible and selectable regressor is that here each consistent realization of the regressand must be in  $\text{pspan}(\Phi_{j \neq k})$ , while selectability requires that no realization is in  $\text{pspan}(\Phi_{j \neq k})$  according to Definition 5.6.1. The set of practically omissible regressors is hence a subset of the set of regressors that are not practically selectable. It is however possible that the regressor can be expressed as a linear combination of the other regressors for some consistent realization and then this regressor might be needed to explain the data if some other regressor is excluded. We therefore introduce practical excludability by requiring practical omissibility and practical independence to ensure that the regressor is not needed to explain the data even if some other regressor is excluded.

**Definition 5.7.2. Practical excludability.**

A regressor  $\phi_k$  is practically excludable if

$$\tilde{\phi}_k \notin \text{pspan } \Phi_{j \neq k} \quad \forall \text{ consistent } \tilde{\phi}_k \in \mathcal{U}_{\phi_k}^\alpha \subseteq \mathbb{R}^m \quad (5.7.2)$$

and

$$\tilde{\xi} \in \text{pspan } \Phi_{j \neq k} \quad \forall \text{ consistent } \tilde{\xi} \in \mathcal{U}_\xi^\alpha \subseteq \mathbb{R}^m \quad (5.7.3)$$

with  $\Phi_{j \neq k} = [\phi_1, \dots, \phi_{k-1}, \phi_{k+1}, \dots, \phi_n]$ . The set of regressors in (5.5.1) is practically excludable if each regressor is practically excludable.

Here condition (5.7.3) is the same as (5.7.1), so every regressor that is practically excludable is also practically omissible. Condition (5.7.2), on the other hand, implies that the regressor can never be written as a linear combination of the other regressors and therefore it can not replace any of the other regressors. This condition also implies that the corresponding parameter  $\theta_k$  is practically unique based on Definition 5.5.12. Inversely, if the regressor is practically omissible and the corresponding parameter is practically unique, then the regressor is practically excludable. If the regressor is practically excludable, then a solution practically exists and we classify the corresponding interaction as non-existing, since this regressor cannot replace any other regressor and it is not needed to explain the data.

**Definition 5.7.3. Non-existing interaction.**

An interaction is non-existing if the corresponding regressor is practically excludable (Definition 5.7.2).

Our classification of interactions as non-existing based on the corresponding regressor being practically excludable is not only intuitive in the sense that these regressors are not needed to explain data, but the number of non-existing interactions also in general decreases with increasing uncertainty. If the size of the uncertainty set of the regressand is increased, then (5.7.3) is in general fulfilled in fewer cases. Similarly, if the size of any of the uncertainty sets of the regressors is increased,

then (5.7.2) is in general fulfilled in fewer cases. This implies that the number of practically excludable regressors and non-existing interactions in general decreases with increasing uncertainty and decreasing significance  $\alpha$ . However, note that non-existing here does not necessarily imply that the interaction does not exist, it merely imply that the interaction is so weak that it is not needed to explain data. The necessary and sufficient data conditions in Definition 5.7.2 are in general difficult to check and we therefore provide the following sufficient conditions for practical excludability, which are sufficient for the classification of the corresponding interaction as non-existing.

**Theorem 5.7.1. Practical excludability.**

*A regressor  $\phi_k$  in (5.5.1) is practically excludable (Definition 5.7.2) if the set of uncertain regressors in the matrix  $\Phi = [\phi_1, \dots, \phi_k, \dots, \phi_n]$  is practically independent (Definition 5.5.13) and the set of vectors in the matrix  $\Psi_k = [\phi_1, \dots, \phi_{k-1}, \phi_{k+1}, \dots, \phi_n, \xi]$  is practically collinear (Definition 5.5.14).*

In general these conditions are conservative for practical excludability, but we can again argue based on Theorem 5.6.3 that they in practice are necessary for data that contain measurements of continuous variables. Compared to the condition for practical selectability in Theorem 5.6.2  $\Psi_k$  is here practically collinear instead of independent. Ideally we would like all regressors to be either practically selectable or excludable so that we can classify all interactions as either existing or non-existing in order to avoid false positives and false negatives. Note however that when more data becomes available then it may be possible to show that interactions that previously were classified as non-existing actually exist, since our definition of practical excludability implies that sufficiently weak interactions are classified as non-existing. This is an unavoidable consequence of classification of interactions as non-existing despite the fact that we can never prove based on finite data with uncertainty that an interaction does not exist. We could for a given data set, in principle at least, calculate an interaction specific detection limit, *i.e.* the strength required for a “true” interaction to not be classified as non-existing. If the detection limit of each interaction classified as non-existing is below a chosen threshold, then we can be sure that all interactions with strength equal or above the selected threshold are correctly inferred and no false negatives exist. This also implies that all interactions corresponding to practically excludable regressors only are true negatives under the assumption that weak interactions should be excluded. This issue arises in all variable selection methods that we know.

## 5.8 Practically negligible and alternative regressors

The determination of variables that must be selected or always can be excluded and the classification of the corresponding interactions as existing or non-existing has been addressed above. Next we address classification of variables and interactions as non-evidential and alternative. This classification is based on lack of information

and we here separate the former practically negligible variables, which cannot be distinguished from the zero vector and hence no information is available about, from the latter practically alternative ones. For this we need to analyse the regressors that previously were found to be not practically selectable nor practically excludable and introduce the concept of practically negligible and alternative regressors.

The idea behind classification of some interactions as non-evidential stems from noise-free data. If a regressor is zero in the noise-free case, then nothing can be said about the existence of the corresponding interaction, since the corresponding parameter can take any value while still being a solution of the system of equations, as shown in Section 4.2. The data contains no information or evidence about the regressor and corresponding interaction and we therefore call it non-evidential. Exclusion of regressors corresponding to non-evidential interactions is based purely on lack of information and it is therefore important in network inference to distinguish these regressors and interactions from practically excludable regressors and non-existing interactions, which are excluded because they are not needed to explain the data despite presence of information. No regressor or regressand is however exactly zero for all consistent realizations when uncertainty is present, so we introduce the notion of a practical zero vector.

**Definition 5.8.1. Practical zero vector.**

A vector  $\phi_k$  is practically zero if a consistent  $\tilde{\phi}_k = \mathbf{0} \in \mathcal{U}_{\phi_k}^\alpha \subseteq \mathbb{R}^m$  exists, *i.e.* if the zero vector is consistent with the error model.

To ensure that the regressor is never needed to explain the data, while it always can be included without affecting the other regressors and their corresponding parameters, *i.e.* that no evidence about it exists in the data, it must be both practically omissible and zero. In such a case, we say that it is practically negligible.

**Definition 5.8.2. Practical negligibility.**

A regressor  $\phi_k$  is practically negligible if a

$$\text{consistent } \tilde{\phi}_k = \mathbf{0} \in \mathcal{U}_{\phi_k}^\alpha \subseteq \mathbb{R}^m \text{ exists} \quad (5.8.1)$$

and

$$\tilde{\xi} \in \text{pspan } \Phi_{j \neq k} \quad \forall \text{ consistent } \tilde{\xi} \in \mathcal{U}_\xi^\alpha \subseteq \mathbb{R}^m \quad (5.8.2)$$

with  $\Phi_{j \neq k} = [\phi_1, \dots, \phi_{k-1}, \phi_{k+1}, \dots, \phi_n]$ . The set of regressors in (5.5.1) is practically negligible if each regressor is practically negligible.

Practical omissibility of a regressor implies that all consistent regressands can be explained by some consistent realizations of the other regressors and a solution of (5.5.1) thus practically exists. A practically negligible regressor is not needed to explain the data and it cannot be distinguished from zero, implying that it typically can replace any of the other regressors. A practically excludable regressor is also not needed to explain the data, but it is distinguishable from zero and cannot replace any of the other regressors. An interesting case occurs when the zero vector is included

in the uncertainty set of the regressand, because the trivial parameter vector  $\theta = \mathbf{0}$  is then a feasible solution and no regressor can consequently be practically selectable. This implies that unsupervised variable selection (see *e.g.* Handl and Knowles, 2006; Jain et al., 2000), *i.e.* selection without a target variable, never can be robust in our sense. Nonetheless, some regressors can still be practically excludable or necessary to explain some of the nonzero realizations of the regressand, implying that the data still may contain useful information from a variable selection perspective. Definition 5.8.2 still correctly captures the case when no information about selection of the regressor exists. From a network inference perspective the lack of information is, however, so severe when the regressand is practically zero that we choose to classify all interactions as non-evidential, since we cannot show that any interaction exists on the selected significance level. We hence classify an interaction as non-evidential in the following cases.

**Definition 5.8.3. Non-evidential interaction.**

An interaction is non-evidential if the corresponding regressor is practically negligible (Definition 5.8.2) or alternatively the regressand is practically zero (Definition 5.8.1) and the regressor is not practically excludable (Definition 5.7.2).

The introduction of the class of non-evidential interactions enables us to distinguish between links that are left out due to lack of information and links left out based on information in the data, *i.e.* interactions classified as non-existing. Inclusion of the zero vector in the uncertainty set implies that we have not received a sufficient signal relative to the uncertainty of the variable and therefore typically lack necessary information.

Even if a regressor is practically zero this does not imply that it is practically negligible, because it might be required to take a nonzero value in order for a solution to practically exist, *i.e.* to explain the data. Therefore, we actually need to check that it also is practically omissible. To foster this check we provide the following two theorems. The first theorem gives sufficient and necessary conditions, while the second theorem provide sufficient but simpler conditions on the data.

**Theorem 5.8.1.** *A regressor  $\phi_k$  in (5.5.1) is practically negligible (Definition 5.8.2) if and only if (5.8.1) holds and a subset  $\mathcal{S}$  of the regressors exists, such that  $k \notin \mathcal{S}$ ,*

$$\tilde{\xi} \in \text{pspan } \Phi_{\mathcal{S}} \quad \forall \text{ consistent } \tilde{\xi} \in \mathcal{U}_{\xi}^{\alpha} \subseteq \mathbb{R}^m, \quad (5.8.3)$$

and

$$\begin{aligned} \tilde{\xi} \notin \text{pspan } \Phi_{\mathcal{S} \setminus l} & \quad \text{for some consistent } \tilde{\xi} \in \mathcal{U}_{\xi}^{\alpha} \subseteq \mathbb{R}^m \\ \text{and } l \in \mathcal{S}, & \quad (5.8.4) \end{aligned}$$

with  $\Phi_{\mathcal{S}}$  consisting of the subset of regressors specified by  $\mathcal{S}$  and  $\Phi_{\mathcal{S} \setminus l}$  of the subset of regressors specified by  $\mathcal{S}$  except  $\phi_l$ . The latter condition ensures that  $\mathcal{S}$  is the minimal set for which the former condition is fulfilled.

**Theorem 5.8.2.** *A regressor  $\phi_k$  in (5.5.1) is practically negligible (Definition 5.8.2) if (5.8.1) holds and a practically independent (Definition 5.5.13) subset  $\mathcal{S}$  of the regressors exists, such that  $k \notin \mathcal{S}$  and the union of this subset of regressors and the regressand is practically collinear (Definition 5.5.14).*

Note that these conditions are not necessary for practical negligibility, since cases when no practically independent subset of the regressors exist such that (5.5.1) has a solution, but they are sufficient. The previous two theorems imply that if we know that a subset  $\mathcal{S}$  of the regressors always can explain the data, then every regressor that is practically zero and does not belong to the subset is practically negligible, which potentially saves us some calculations.

### 5.8.1 Practically alternative

We now introduce practically alternative regressors and alternative interactions in a straight forward manner.

#### Definition 5.8.4. Practical alternative.

A regressor  $\phi_k$  is practically alternative if it is not practically selectable (Definition 5.6.1), excludable (Definition 5.7.2), nor negligible (Definition 5.8.2). The set of regressors in (5.5.1) is practically alternative if each regressor is practically alternative.

An alternative regressor is needed to explain data for some consistent realization, while it is not needed for some other consistent realization, and therefore it cannot be practically selectable nor excludable. The following corollary summarizes the conditions for when a regressor is practically alternative.

**Corollary 5.8.3.** *A regressor  $\phi_k$  is practically alternative (Definition 5.8.4) if and only if either*

$$\tilde{\xi} \in \text{pspan}(\Phi_{j \neq k}) \quad \forall \text{ consistent } \tilde{\xi} \in \mathcal{U}_{\xi}^{\alpha} \subseteq \mathbb{R}^m, \quad (5.8.5a)$$

$$\tilde{\phi}_k \in \text{pspan}(\Phi_{j \neq k}) \quad \text{for some consistent } \tilde{\phi}_k \in \mathcal{U}_{\phi_k}^{\alpha} \subseteq \mathbb{R}^m, \quad \text{and} \quad (5.8.5b)$$

$$\tilde{\phi}_k = \mathbf{0} \notin \mathcal{U}_{\phi_k}^{\alpha}, \quad (5.8.5c)$$

*i.e. the regressor is practically omissible but not excludable nor negligible, or*

$$\tilde{\xi} \in \text{pspan}(\Phi_{j \neq k}) \quad \text{for some consistent} \\ \tilde{\xi} \in \mathcal{U}_{\xi}^{\alpha} \subseteq \mathbb{R}^m, \quad \text{and} \quad (5.8.6a)$$

$$\tilde{\xi} \notin \text{pspan}(\Phi_{j \neq k}) \quad \text{for some consistent} \\ \tilde{\xi} \in \mathcal{U}_{\xi}^{\alpha} \subseteq \mathbb{R}^m, \quad (5.8.6b)$$

*i.e. the regressor is not practically selectable nor omissible, with  $\Phi_{j \neq k} = [\phi_1, \dots, \phi_{k-1}, \phi_{k+1}, \dots, \phi_n]$ .*

The definition of practically alternative strictly speaking implies that a regressor is practically alternative only when the conditions in the previous corollary hold. But we for convenience use the term practically alternative for all regressors that we have not shown to be practically selectable, excludable, or negligible, given that at least some regressor is shown to be either practically selectable, excludable, or negligible. If the regressand is practically zero, then every regressor is either practically negligible, excludable, or alternative, since the zero vector always belongs to  $\text{pspan}(\Phi_{j \neq k})$ . More precisely, each regressor that is practically zero is practically negligible and all other regressors are practically alternative, unless there is no uncertainty in the regressand since then practically independent regressors are practically excludable. Correspondingly, all interactions except those corresponding to practically excludable regressors, which are classified as non-existing, are classified as non-evidential due to lack of information. If the regressor is practically alternative, the regressand is not practically zero, and a solution practically exists, then we classify the corresponding interaction as alternative, since the parameter must be nonzero for some consistent combination of regressors and regressand, *i.e.* for some value of the other parameters.

**Definition 5.8.5. Alternative interaction.**

An interaction is alternative if the corresponding regressor is practically alternative (Definition 5.8.4) and the regressand is not practically zero (Definition 5.8.1).

If a single final network model is desired, then links corresponding to alternative interactions can either be included or left out, but any such selection does not change the fact that we cannot show based on the given data that an alternative interaction exists, nor exclude it. Note that the previously adopted custom to call every regressor practically alternative that is not proven to be anything else implies that the corresponding interaction is called alternative until it is proven to be either existing, non-existing, or non-evidential.

This concludes our introduction of concepts for classification of regressors and interactions, since we now have the basic results that are needed for robust variable selection and network inference when the set of all linear models are considered. The number of definitions and theorems is fairly large and we therefore provide an overview of our classification of interactions, regressors, and corresponding parameters in Table 5.1. We next consider geometrical interpretation of the concepts and demonstrate the use of our results on a two gene example.

## 5.9 Geometrical interpretation

To be present in all models that can explain data, a regressor must be practically selectable, while it must be practically excludable for data to always be explainable without it. A regressor is practically selectable or excludable if the corresponding data matrix is practically independent or collinear, respectively, as established in Theorems 5.6.2 and 5.7.1. These concepts have a simple geometrical interpretation

**Table 5.1: Classification of variables and interactions.** Our classification of regressors, interactions, and corresponding parameters, with data conditions. Here Dx = Definition x, Tx = Theorem x, Cx = Corollary x, & = and, and | = or.

Interaction	Regressor	Parameter	Condition
existing D5.5.6	practically selectable D5.6.1	practically unique D5.5.12 & practically nonzero D5.5.7a	(T5.6.2   C5.6.5) & T5.5.2
activating D5.5.9	practically assignable D5.5.8	practically unique D5.5.12 & practically positive D5.5.7b	T5.6.6   T5.6.7   T5.6.8
repressing D5.5.10	practically assignable D5.5.8	practically unique D5.5.12 & practically negative D5.5.7c	
non-existing D5.7.3	practically excludable D5.7.2	practically unique D5.5.12 & practically zero D5.5.7d	T5.7.1
non-evidential D5.8.3 †	practically negligible D5.8.2	practically zero D5.5.7d	T5.8.1   T5.8.2
alternative D5.8.5 ‡	practically alternative D5.8.4	practically zero D5.5.7d	C5.8.3

† regressor classification or regressand practically zero D5.8.1 and regressor not practically excludable D5.7.2

‡ regressor classification and regressand not practically zero D5.8.1

that we present here in form of a small example to provide insight. Before illustrating the concepts, while performing a systematic analysis according to Table 5.1, we briefly explain the TOEL example and data.

The TOEL example consists of two genes that form a positive feedback loop, which we constructed to illustrate network properties and inference of GRNs. We here use data obtained through an *in silico* perturbation experiment in which the two genes were perturbed one-by-one in two separate steady-state experiments (P1 and P2) and the response recorded. Each regressor  $\phi_j$  is formed by the observed responses in the  $j$ th gene and each regressand  $\xi_i$  by the perturbations applied to the  $i$ th gene. We assume the same deterministic uncertainty that we illustrated in Figure 5.1 (left). According to this error model any vector within the uncertainty set of each of the variables could be its “true” realization, so the selected variables and inferred interactions are only robust if we account for all realizations. All data needed for performing robust inference of the TOEL network, *i.e.* the regressors, regressands, and their uncertainty sets, are presented in Figure 5.2. Details on the example that would not be available in a real inference case, such as the “true” network used to generate the data, are left to Section A.7. We here illustrate everything graphically in the data space, characterized by having one experiment on each axis. This is, however, only one common way of graphically illustrating data and we therefore relate to other common representations in Section A.2.

To simplify the explanations let us first introduce the uncertainty cone that is generated by all realizations in the uncertainty set of a variable. We define the uncertainty cone of a variable as

$$\mathcal{U}_{\phi_j}^c \triangleq \{z \mid z = c\tilde{\phi}_j \text{ for } \tilde{\phi}_j \in \mathcal{U}_{\phi_j}^\alpha, c \in \mathbb{R}\}, \quad (5.9.1)$$

*i.e.* the set of points that are multiples of any vector in the uncertainty set of the variable. Strictly speaking each uncertainty cone is a double infinite cone, since  $c$  can take both positive and negative values, contrary to the convex cones  $\mathcal{H}_+$  in (5.6.11) and  $\mathcal{H}_-$  in (5.6.12). The reason for defining the uncertainty cones in this way is that we want them to be sign invariant and closed under multiplication.

Following the regressor classification in Table 5.1, we first discuss practical selectability and excludability in this paragraph and then practical negligibility and alternative in the next paragraph. Using the concept of uncertainty cones, we can now say that a set of variables is practically independent if each uncertainty cone can be separated from the other by a number of hyperplanes, which intersect in the origin. A realization of variables is linearly dependent if they lay on a hyperplane, because  $\dim \text{span}(\Phi) < n$  on a hyperplane. No linearly dependent realization of variables therefore exists when the uncertainty cones can be separated by a set of hyperplanes and the set of variables is thus practically independent based on Corollary 5.5.4. This corresponds to uncertainty cones that only have the origin in common if the set only contains two regressors, which is illustrated in Figure 5.2 (left) for  $\Psi_1 = [\phi_2, \xi_1]$ . This implies that  $\phi_1$  is practically selectable based on Theorem 5.6.2. On the other hand, the uncertainty cone of  $\phi_1$  and  $\xi_1$  overlap for a wide range of points, so  $\Psi_2 = [\phi_1, \xi_1]$  is not practically independent and consequently  $\phi_2$  cannot be proven practically selectable. If the uncertainty cone of any of the variables lies within the practical span of the other variables, then the set of variables is practically collinear based on Corollary 5.5.6. This corresponds to one uncertainty cone lying within the other uncertainty cone when the set only contains two regressors, which is illustrated in Figure 5.2 (left) for  $\Phi = [\phi_1, \phi_2]$ . This implies that none of the regressors can be practically excludable based on Definition 5.7.2.

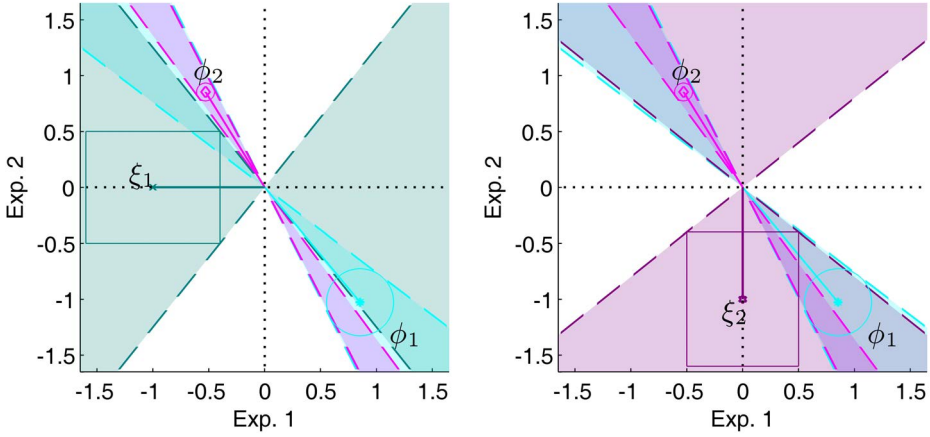
A regressor can only be practically negligible if it is practically zero. A sufficient requirement for a vector to not be practically zero is often conveniently expressed in terms of the SNR, which we previously have mentioned without a proper definition. We define the signal to noise ratio of a variable  $\phi_j$  as

$$\text{SNR}(\phi_j) \triangleq \frac{\|\phi_j\|}{r_{\mathcal{U}_{\phi_j}}}, \quad (5.9.2)$$

with the radius of the uncertainty set

$$r_{\mathcal{U}_{\phi_j}} \triangleq \sup_{\tilde{\phi}_j \in \mathcal{U}_{\phi_j}^\alpha} \|\tilde{\phi}_j - \phi_j\|. \quad (5.9.3)$$

If  $\text{SNR}(\phi_j) < 1$ , then the uncertainty cone of the regressor  $\phi_j$  covers the whole space, implying that the set of regressors is practically collinear. If  $\text{SNR}(\phi_j) > 1$ , then the regressor  $\phi_j$  is not practically zero and it cannot be practically negligible according to Definition 5.8.2. Similarly,  $\text{SNR}(\xi) > 1$  ensures that the regressand is not practically zero. This is fulfilled for both regressors and the regressand in Figure 5.2 (left), because none of the uncertainty cones covers the whole space. In



**Figure 5.2: Illustration of practical independence and collinearity, and demonstration of regressor classification.** The two steady-state perturbation experiments (P1 and P2) that yielded the data are illustrated in Figure A.7 and the deterministic uncertainty sets are explained in Figure 5.1 (left). Data for inference of the first row of the interaction matrix  $\mathbf{A}$  is shown left, while data for the second row is shown right. The two dashed lines with the same colour as the corresponding variable depict the uncertainty cone defined by the uncertainty set of the variable. The cone of  $\phi_1$  and  $\xi_1$  overlap (left), but the cone of  $\phi_2$  and  $\xi_1$  does not, so the set in  $\Psi_1 = [\phi_2, \xi_1]$  is practically independent. The regressor  $\phi_1$  is therefore practically selectable and the corresponding interaction  $a_{11}$  is classified as existing based on the results in Table 5.1. While the set in  $\Phi = [\phi_1, \phi_2]$  is practically collinear, since the cone of  $\phi_2$  is contained within the cone of  $\phi_1$ . Neither regressor can therefore be practically excludable. The SNR of all variables is above one, since the origin is not contained in any of the uncertainty sets and no uncertainty cone covers the whole space. Neither regressor can therefore be practically negligible, hence  $\phi_2$  is practically alternative and the corresponding interaction  $a_{12}$  is classified as alternative (left). Both regressors are practically alternative and  $a_{21}$  and  $a_{22}$  are classified as alternative (right). The effect of a third experiment that makes the data informative for robust variable selection is illustrated in Figure A.8.

other words, the second regressor cannot be practically negligible and we earlier found that it is not practically selectable nor excludable, so it is hence practically alternative based on Definition 5.8.4.

Based on our regressor classification we can now also classify the interactions. The TOEL network consists of two genes and uncertain data from two perturbation experiments is used. The number of variables hence equals the number of measurements and the regressors span the data space, so at least one solution of (5.5.1) exist. We now know that gene  $A$  has a self-loop, since this interaction corresponds to the practically selectable regressor  $\phi_1$  and a solution practically

exists. This interaction is consequently classified as existing based on Definition 5.5.6. The influence from gene  $B$  to gene  $A$  is classified as alternative based on Definition 5.8.5. The data obtained in the two perturbation experiments is not informative enough to robustly determine if it exists or not. The corresponding illustration for the second row of the interaction matrix  $\mathbf{A}$  in Figure 5.2 (right) reveals that both interactions are alternative, since the uncertainty cone of the regressand overlaps with both regressors and the regressors based on the reasoning above therefore are practically alternative. Informative data for robust inference of all four interactions, which contains an additional perturbation experiment, is illustrated in Figure A.8. Graphical classification of regressors and interactions is limited to small examples of great pedagogical, but typically little practical, value, so we next study computational testing for practical independence.

## 5.10 Testing for practical independence

To perform robust variable selection and network inference, in practice, one needs to determine if an uncertain matrix is practically independent (Definition 5.5.13), *i.e.* show that no realization of lower rank than the observed one exists within the uncertainty sets of the variables. Practical independence is, in particular for continuous variables, useful for establishing that a regressor is practically selectable (Definition 5.6.1), *i.e.* must be present to explain data, and that the corresponding interaction exists. Here we therefore prove that two classical tools for computations in matrix algebra and robust control—the singular value decomposition (SVD) and structured singular value (SSV) (see *e.g.* Horn and Johnson 1990, p. 414-415; Skogestad and Postlethwaite 1996, p. 331)—can be used to test if an uncertain matrix is practically independent. In other words, we provide computationally feasible tools for implicitly verifying that a regressor and interaction is present in all models that can explain data.

We start by establishing the connection between a matrix with uncertain variables and the structured singular value problem, followed by a proof of practical independence using the SSV and comments on calculation. All realizations that are consistent with the error model can be expressed as the observed matrix  $\Phi$  minus an error realization  $\Upsilon$ ,  $\tilde{\Phi} = \Phi - \Upsilon$ . The idea behind the SSV is to find the closest rank deficient  $\tilde{\Phi}$  under structured uncertainty in terms of bounds on the norm of  $\Upsilon$ . Three basic uncertainty structures exist: bounded elements  $|v_{ij}| \leq 1$ , bounded columns  $\|\mathbf{v}_j\|_2 \leq 1$ , and matrix bounds  $\|\Upsilon\|_2 \leq 1$ . We focus on the second one because it corresponds to variable specific error bounds and the third one because it provides an exact test for normally distributed errors. Even though an one-to-one relation between the uncertainty sets and norm bounds in general does not exist, an upper bound can always be found for finite uncertainty sets such that the norm bound realizations constitute a superset of the uncertainty sets. If  $\text{rank}(\Phi - \Upsilon) = n$  for all  $\Upsilon$  within the norm bounds, then the set of uncertain variables is practically independent based on Corollary 5.5.8. Formally the structured singular value  $\mu(\mathbf{M})$

is defined as (Doyle, 1982; Safonov, 1982; Zhou and Doyle, 1998, p. 189)

$$\frac{1}{\mu(\mathbf{M})} \triangleq \min_{\Delta} \{\bar{\sigma}(\Delta) \mid \det(\mathbf{I} - \mathbf{M}\Delta) = 0 \text{ for structured } \Delta\}. \quad (5.10.1)$$

In our applications,  $\mathbf{M}$  will be a function of the observed data matrix  $\Phi$  and  $\Delta$  represents the structure and norm bounds on  $\Upsilon$ . We report  $\gamma \triangleq \mu^{-1}$  instead of  $\mu$  and call it the confidence score, because the set of vectors is practically independent if the uncertainty sets are scaled by any factor smaller than  $\gamma$  and each  $\gamma$  value corresponds to a significance level. The following theorem proves sufficient conditions for practical independence, by utilizing the SSV with norm bounds on individual columns.

**Theorem 5.10.1. Practical independence for column uncertainty.**

The set of uncertain vectors in the matrix  $\Phi = [\phi_1, \dots, \phi_j, \dots, \phi_n] \in \mathbb{R}^{m \times n}$ , with each realization  $\tilde{\phi}_j \in \mathcal{U}_{\phi_j}^\alpha \subseteq \mathcal{U}_{\phi_j}^{\text{ell}}$  and  $m \geq n$ , is practically independent at significance level  $\alpha$  (Definition 5.5.13) if the confidence score  $\gamma(\Phi) > 1$ . Here the confidence score is defined as

$$\gamma(\Phi) \triangleq \frac{1}{\mu(\Phi)} \triangleq \min_{\Delta} \left\{ \bar{\sigma}(\Delta) \mid \det(\mathbf{I} - \Phi^\dagger \mathbf{W}_1 \Delta) = 0 \right. \\ \left. \text{for } \Delta = \text{diag}(\delta_1, \dots, \delta_j, \dots, \delta_n), \text{ with } \delta_j \in \mathbb{R}^m \right\}, \quad (5.10.2)$$

and the scaling matrix, i.e. the right transformation matrix of  $\Phi^\dagger$ , as

$$\mathbf{W}_1 \triangleq [\mathbf{Q}_{\phi_1}^{-\frac{1}{2}}, \dots, \mathbf{Q}_{\phi_j}^{-\frac{1}{2}}, \dots, \mathbf{Q}_{\phi_n}^{-\frac{1}{2}}], \quad (5.10.3)$$

with the ellipsoidal weights  $\mathbf{Q}_{\phi_j}^{\frac{1}{2}T} \mathbf{Q}_{\phi_j}^{\frac{1}{2}} = \mathbf{Q}_{\phi_j}$  of the ellipsoidal set  $\mathcal{U}_{\phi_j}^{\text{ell}}$ , which is defined as

$$\mathcal{U}_{\phi_j}^{\text{ell}} \triangleq \{\phi_j + \mathbf{v}_j \mid \mathbf{v}_j^T \mathbf{Q}_{\phi_j} \mathbf{v}_j \leq 1, \mathbf{v}_j \in \mathbb{R}^m\}. \quad (5.10.4)$$

Here  $\bar{\sigma}(\Delta)$  denotes the largest singular value of  $\Delta$  and  $^\dagger$  the Moore-Penrose generalized inverse (Horn and Johnson, 1990, p. 414-415, 421), while  $\text{diag}$  denotes a matrix with the argument on the diagonal and all other elements zero.

This test of practical independence applies to any data matrix with variables that can be bounded by ellipsoids and  $\gamma(\Phi) > 1$  is always sufficient when the SSV is calculated as (5.10.2), but it is only necessary if the uncertainty sets are ellipsoids and all variable combinations within are consistent with the error model. This test is therefore in general conservative. The confidence score reveals the factor by which the ellipsoidal uncertainty sets should be scaled to have a realization with linearly dependent variables on the closure of the sets. In other words, it also tells us how much larger the uncertainty could be while practical independence still can be

guaranteed, or how much smaller the uncertainty need to be in order to guarantee practical independence. Each confidence score  $\gamma$  corresponds to a specific selection of the significance level  $\alpha$  for construction of the uncertainty sets, so the significance level at which practical independence can be guaranteed can be calculated from  $\gamma$ . But  $\alpha$  is in general a nonlinear function of  $\gamma$  and the insight provided by use of  $\gamma$  would hence be lost, so we prefer to use  $\gamma$ .

The structured singular value is by now a classical tool in robust control and efficient methods for calculating upper and lower bounds on it exist, see *e.g.* Skogestad and Postlethwaite (1996); Dullerud and Paganini (2000); Zhou and Doyle (1998); Zhou et al. (1996). These bounds are used in practice because exact calculation of the structured singular value is in general NP-hard. *E.g.* the `mussv` implementation in Matlab ([www.mathworks.com](http://www.mathworks.com)) can be used to calculate the SSV bounds (Gu et al., 2005). It is worth noting that many algorithms for the calculation of the SSV assume that the  $\Delta$  matrix in (5.10.2) is allowed to take complex values. The set of complex numbers is however a superset of the real numbers, so  $\gamma > 1$  will still be sufficient for practical independence, but in certain cases it makes this test conservative, since we only allow real errors. We recommend the use of the lower bound on  $\gamma$ , corresponding to the upper bound on  $\mu$ , since no rank deficient realization is included in the uncertainty sets that results from scaling by any value smaller than this one, implying that the data matrix is guaranteed to be practically independent up to this value. Three examples of how to calculate the SSV, using `mussv` in Matlab, and confidence score are given in Section A.8. Practical independence of  $\Psi_{ki} = [\phi_1, \dots, \phi_{k-1}, \phi_{k+1}, \dots, \phi_n, \xi_i]$  guarantees that the regressor  $\phi_k$  is practically selectable (Theorem 5.6.2), *i.e.* selected in all feasible models, for the variable selection problem faced on row  $i$  of the interaction matrix. So when applied to  $\Psi_{ki}$  the SSV based test, proposed here, provides a reliable way to calculate a lower bound on the confidence score of every possible interaction and ensures that no model with an alternative structure lacking the  $a_{ik}$  interaction exists within the set of all practically indistinguishable models up to the significance level that  $\gamma(\Psi_{ki})$  corresponds to.

The test of practical independence above can almost always be used, but in general it provides a conservative confidence score. Other confidence scores that are exact and conditions that are both sufficient and necessary for practical independence, however, exist for certain error models. One such case is errors drawn from a multivariate normal distribution with zero mean and covariance  $\lambda \mathbf{I}$ , *i.e.* white Gaussian noise, as we prove in the following theorem. White Gaussian noise is the most common noise assumption in system identification, signal processing, and perhaps modelling in general, so we believe that this case deserves special attention. It also turns out that the singular value decomposition, which is a standard tool in matrix algebra, in this case can be used to calculate the exact confidence score, so it provides further insight on testing for practical independence to everyone familiar with it. The singular values of a matrix  $\Phi \in \mathbb{R}^{m \times n}$  arranged in decreasing order  $\bar{\sigma} \geq \sigma_k \geq \underline{\sigma}$ , left singular vectors  $\mathbf{U}_\Phi \triangleq [\mathbf{u}_1, \dots, \mathbf{u}_k, \dots, \mathbf{u}_m]$ , and right singular vectors  $\mathbf{V}_\Phi \triangleq [\mathbf{v}_1, \dots, \mathbf{v}_k, \dots, \mathbf{v}_n]$  are given by the singular value decomposition,

defined as (Horn and Johnson, 1990, p. 414-415)

$$\begin{aligned}\Phi &= \mathbf{U}_\Phi \Sigma_\Phi \mathbf{V}_\Phi^T, \text{ with } \Sigma_\Phi = \text{diag}(\bar{\sigma}, \dots, \sigma_k, \dots, \underline{\sigma}), \\ \mathbf{U}_\Phi &\in \{\mathbb{R}^{m \times m} | \mathbf{U}_\Phi^T \mathbf{U}_\Phi = \mathbf{I}\}, \\ \mathbf{V}_\Phi &\in \{\mathbb{R}^{n \times n} | \mathbf{V}_\Phi^T \mathbf{V}_\Phi = \mathbf{I}\}.\end{aligned}\tag{5.10.5}$$

The  $n$ th largest singular value  $\sigma_n$  is the distance to rank deficiency, measured by the induced 2-norm, of a matrix with rank  $n \leq m$ . It is a special case of the structured singular value above obtained when no structure is assumed on  $\Delta$  and the 2-norm of  $\Delta$  is bounded. We use it to calculate an exact confidence score for normally distributed errors based on the following theorem.

**Theorem 5.10.2. Practical independence for normally distributed errors.** *The set of uncertain vectors in the matrix  $\Phi = [\phi_1, \dots, \phi_j, \dots, \phi_n] \in \mathbb{R}^{m \times n}$ , with the corresponding identical and independent errors  $\Upsilon = [\mathbf{v}_1, \dots, \mathbf{v}_j, \dots, \mathbf{v}_n]$  drawn from a multivariate normal distribution with zero mean and covariance matrix  $\lambda \mathbf{I}$ ,  $\tilde{\Upsilon} \sim \mathbf{N}(\mathbf{0}, \lambda \mathbf{I})$ , and  $m \geq n$ , is practically independent at significance level  $\alpha$  (Definition 5.5.13) if and only if the confidence score  $\gamma(\Phi) > 1$ . In this case the confidence score is defined as*

$$\gamma(\Phi) \triangleq w_1^{-1} \sigma_n(\Phi),\tag{5.10.6}$$

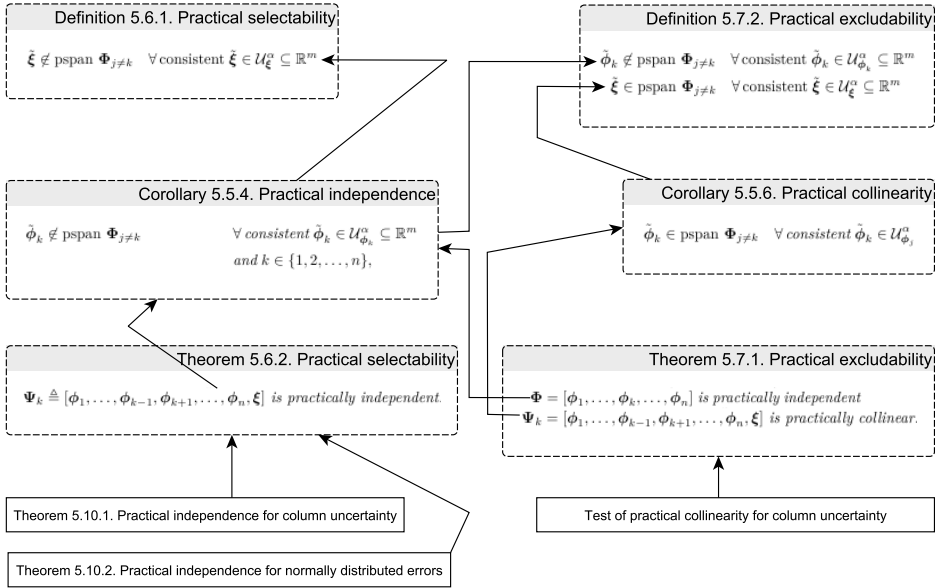
and the scaling constant as

$$w_1 \triangleq \sqrt{\chi^{-2}(\alpha, nm)} \lambda.\tag{5.10.7}$$

Here  $\vec{\cdot}$  denotes vectorization column by column, i.e.  $\vec{\Upsilon} = [\mathbf{v}_1^T, \dots, \mathbf{v}_n^T]^T$ , and  $\sigma_n$  denotes the  $n$ th largest singular value (Horn and Johnson, 1990, p. 414-415). In addition,  $\chi^{-2}(\alpha, nm)$  is the inverse of the chi-square cumulative distribution with  $nm$  degrees of freedom, such that  $\mathbb{P}[\chi^2(nm) > \chi^{-2}(\alpha, nm)] = \alpha$  (Chew, 1966).

If the covariance matrix  $\Lambda_\Upsilon \neq \lambda \mathbf{I}$ , then the variables should be transformed such that the covariance matrix becomes an identity matrix while preserving the relation among the variables. One such example is given in Section A.8.

To summarize, the structured singular value and singular value decomposition provide computationally efficient tools to test if an uncertain matrix is practically independent (Definition 5.5.13), i.e. show that no realization of lower rank than the observed one exists within the uncertainty sets of the variables. Together with the tools developed in Sections 5.5 and 5.6, we now have everything needed to robustly infer the interactions that exist and prove that they exist (Definition 5.5.6), i.e. must be present to explain data within the class of all linear network models at the selected significance level when the error model is assumed to be correct. In other words, we now have computational tools for implicitly verifying that a regressor and interaction is present in all solutions to the typically infinite number of systems of linear equations in (5.5.1). The formal path from these two tests of practical



**Figure 5.3: Graph over the path to proving practical selectability and excludability using tests for practical independence and collinearity.** The tests are at the bottom and the path is marked by arrows leading to the definitions at the top.

independence (Theorems 5.10.1 and 5.10.2) to proving practical selectability of the regressor (Definition 5.6.1) corresponding to the existing interaction is shown in Figure 5.3. Note that to show that an interaction is non-existing we need to prove that the corresponding regressor is practically excludable (Definition 5.7.2) and this requires a test of practical collinearity, as indicated by the ‘Test of practical collinearity for column uncertainty’ in the figure. Development of such a test of practical collinearity is, however, left for future work.

### 5.11 Conclusions

Observed data contain measurement errors and can therefore typically be explained by infinitely many models. These models are indistinguishable in the sense that none of them can be rejected at a desired significance/confidence level based on the error model of the measurement process. Selection of the variables that optimize some quantity, e.g. maximize the likelihood of the observed data or minimize the prediction or classification error, is typically the objective of existing variable

selection methods (see *e.g.* Dash and Liu, 1997; Peng et al., 2005; Molina et al., 2002; Guyon and Elisseeff, 2003; Fan and Lv, 2010; Hara and Sillanp, 2009; George, 2000). On the contrary, our objective is to determine the variables that must be included in or that always can be excluded from models explaining the data at the desired confidence level. The subset of variables that we select are hence true positives, assuming that the “true” model belongs to the set of indistinguishable models, while the subset that we exclude are true negatives, assuming that variables with weak relation should be excluded, and we say that the selection is robust. This ensures that the validity of the selection only depends on the validity of the investigated class of functions and assumed error model, which for each case should be discussed, while in existing methods it largely depends on the method. Checking of every model in the infinite set of indistinguishable models is, of course, computationally infeasible, but we have found a way to indirectly do this without even fitting any model to the data, by reformulating the variable selection problem as a rank problem of the data matrices as we demonstrated in Chapter 4.

In general, variables and their corresponding interactions, in models viewed as network graphs, can in layman terms be classified as present/existing, absent/non-existing, non-evidential, or alternative. Due to the close connection between variable selection and network inference, we have here developed the beginning of a theory for both robust variable selection and network inference. It is to the best of our knowledge the first theory that both can be used to determine the variables and interactions that must be included in and always can be excluded from the model with desired confidence and prove this to be the case. We say that the former subset of variables/regressors are practically selectable (Definition 5.6.1) and the later practically excludable (Definition 5.7.2), while the corresponding interactions are existing (Definition 5.5.6) and non-existing (Definition 5.7.3), respectively. Since a model can never be validated, we do the next best thing and prove that the observed data cannot be explained by a chosen class of functions or model set without the practically selectable variables. It is philosophically impossible to prove for finite uncertain data that a variable does not influence the target variable, so we do the next best thing and prove that the observed data can always be explained by a chosen class of functions, or model set, without the practically excludable variables. The remaining variables are classed as either practically negligible (Definition 5.8.2) or practically alternative (Definition 5.8.4) and the corresponding interactions as non-evidential (Definition 5.8.3) or alternative (Definition 5.8.5), depending on the degree of lack of information. The practically negligible variables cannot even be distinguished from the zero vector, so we know that no information is available about them.

In particular biological data sets are typically only partly informative and suffer from low signal to noise ratios, errors-in-variables, near collinearity, and few data points (Holter et al., 2000; Tegnér and Björkegren, 2007; Cosgrove et al., 2010; Nordling and Jacobsen, 2009a). We have not found any current theory or method that can cope with the combination of these in a robust manner like the one presented here. Our theory hence appears to be the only one to provide the much

sought tools for inferring direct causal interactions, *e.g.* between genes, in biology and medicine. Inference of causal interactions from observed data is necessary to gain mechanistic insight. Philosophically, non-robust methods can only generate hypotheses, while robust methods perform testing of implicitly generated hypotheses. The method presented here therefore offers automated discovery of the type sought in systems biology, see *e.g.* [Ideker et al. \(2001\)](#). Our two tests of practical independence (Theorems [5.10.1](#) and [5.10.2](#)) are based on the structured singular value and singular value decomposition (see *e.g.* [Skogestad and Postlethwaite 1996](#), p. 331; [Horn and Johnson 1990](#), p. 414-415), so existing efficient methods and implementations for their calculation can now be used to prove practical selectability of variables and existence of interactions.

The robust variable selection problem (Definition [5.2.1](#)) and network inference problem (Definition [5.3.1](#)) can both be solved without fitting any model to data—mere knowledge that a solution exists is sufficient—so we have decoupled variable selection from parameter estimation. This is of both theoretical and practical value, because variable selection is a part of the model structure selection that should precede parameter estimation and it ensures that our selection only depends on the observed data and specified class of models. Existing methods for model selection, such as the F-test or likelihood ratio test ([Cedersund and Roll, 2009](#); [Casella and Berger, 2001](#)), information criteria, such as Akaike information and Bayesian information ([Stoica and Selen, 2004](#); [Akaike, 1973](#); [Schwarz, 1978](#)), or regularization, such as the penalized likelihood framework ([Fan and Lv, 2010](#); [Fan and Li, 2001](#)), however, require estimation of model parameters and the selection therefore depends on the parameter estimation. This is problematic because any property of a model, *e.g.* the variance of its parameters, only holds for the chosen model structure ([Burnham and Anderson, 2002](#), p. 14). Moreover, traditionally a single model is constructed by fitting the parameters to a training data set and it is then accepted if it cannot be invalidated based on a validation data set. We instead use all data to invalidate as many models as possible, here limited to the class of linear models, and accept all models that cannot be invalidated. In addition to robustness, this provides a straightforward way to add data later on. Additional data can only invalidate more models and variables that previously were practically selectable will continue to be so, so we only need to test if any additional variables become practically selectable. For these reasons, we believe that the creation of our robust theory could spark a new research direction in statistics, estimation theory, system identification, and econometrics. As a side result, by introducing practical independence (Definition [5.5.13](#)) and practical collinearity (Definition [5.5.14](#)) of a set of vectors in an uncertain matrix, based on our current understanding, we also give the first exact solution of the conditioning problem. The conditioning problem is known to cause instability of estimators, large uncertainty in the estimated parameters, sign errors even in statistically significant parameters, and good fit to data despite statistically insignificant parameters, see *e.g.* [Belsley \(1991\)](#); [Liao \(2010\)](#); [Rao and Toutenburg \(1999\)](#); [Larose \(2005\)](#). The key advantage of this work can, however, be summarized in one word—confidence. If our conditions for robust

classification are not fulfilled, then no confidence can be assigned even though almost all network models have the same interactions and we can say that it is the most likely network, implying that no method can do better.

## 5.12 Proofs

We here prove all results presented as theorems or corollaries in this chapter.

**Proof of Corollary 5.5.1.**

$$\begin{aligned}
 \text{pspan } \Phi &= \left\{ \sum_{j=1}^n \theta_j \tilde{\phi}_j \mid \theta_j \in \mathbb{R}, \tilde{\phi}_j \in \tilde{\Phi}, \tilde{\Phi} \in \mathcal{V}_C \right\} \\
 &= \bigcup_{\tilde{\Phi} \in \mathcal{V}_C} \left\{ \sum_{j=1}^n \theta_j \tilde{\phi}_j \mid \theta_j \in \mathbb{R}, \tilde{\phi}_j \in \tilde{\Phi} \right\} \\
 &= \bigcup_{\tilde{\Phi} \in \mathcal{V}_C} \text{span } \tilde{\Phi}. \tag{5.12.1}
 \end{aligned}$$

□

**Proof of Theorem 5.5.2.** Equation (5.5.8) is based on Corollary 5.5.1 equivalent to

$$\tilde{\xi} \in \bigcup_{\tilde{\Phi} \in \mathcal{V}_C} \text{span } \tilde{\Phi} \text{ for some consistent } \tilde{\xi} \in \mathcal{U}_{\xi}^{\alpha}. \tag{5.12.2}$$

This is equivalent to  $\tilde{\xi} \in \text{span}(\tilde{\Phi})$  for some realization  $\tilde{\Phi}$ , which is both necessary and sufficient for existence of a solution of  $\tilde{\Phi}\theta = \tilde{\xi}$  (Friedberg et al., 2003, p. 174), given that this regressand and regressor combination is consistent, i.e.  $\tilde{\xi} \in \mathcal{U}_{\xi}^{\alpha}$  and  $\tilde{\Phi} \in \mathcal{V}_C$ . This implies that a feasible solution exists, which is the requirement in Definition 5.5.2, and that it only exists when (5.5.8) holds. Note that if  $\text{pspan}(\Phi)$  contained linear combinations of regressors from different realizations  $\tilde{\Phi}$ , which characterizes subspaces, then this would not have been sufficient, but it does not. □

**Proof of Theorem 5.5.3.** When the opposite of (5.5.9) holds, i.e.  $\tilde{\xi} \notin \text{pspan } \Phi_{j \neq k} \forall$  consistent  $\tilde{\xi} \in \mathcal{U}_{\xi}^{\alpha}$ , then a solution to  $\tilde{\Phi}\theta = \tilde{\xi}$  can only exist for realizations such that  $\tilde{\phi}_k \notin \text{span}(\tilde{\Phi}_{j \neq k})$  and  $\theta_k$  has a unique value in each feasible solution, since the corresponding regressor  $\tilde{\phi}_k$  cannot be compensated for by a linear combination of the other regressors. A solution can never exist for realizations such that  $\tilde{\phi}_k \in \text{span}(\tilde{\Phi}_{j \neq k})$  when  $\tilde{\xi} \notin \text{pspan } \Phi_{j \neq k} \forall$  consistent  $\tilde{\xi} \in \mathcal{U}_{\xi}^{\alpha}$ , so it does not matter if  $\tilde{\phi}_k$  is a linear combination of the other regressors. When (5.5.9) holds, then (5.5.10) ensures that  $\theta_k$  has a unique value in each feasible solution, since the corresponding regressor  $\tilde{\phi}_k$  cannot be compensated for by a linear combination of the

other regressors. For each consistent regressor matrix  $\tilde{\Phi}$  and regressand  $\tilde{\xi}$ ,  $\tilde{\xi} \in \mathcal{U}_{\tilde{\xi}}^{\alpha}$ , such that a feasible solution exists:  $\tilde{\Phi}\tilde{\theta} = \tilde{\xi}$  and  $\tilde{\Phi}\tilde{\theta}' = \tilde{\xi}$ , then these can be inserted into (5.5.11)

$$\left\| \tilde{\Phi}\tilde{\theta} - \tilde{\Phi}\tilde{\theta}' \right\| \leq r_{\tilde{\xi}}. \quad (5.12.3)$$

The other parameters need not be unique even though  $\theta_k$  is, but if we always select the solutions of  $\tilde{\Phi}\tilde{\theta} = \tilde{\xi}$  and  $\tilde{\Phi}\tilde{\theta}' = \tilde{\xi}$  such that  $\tilde{\theta} - \tilde{\theta}' = \mathbf{0}$  if  $\tilde{\Phi}(\tilde{\theta} - \tilde{\theta}') = \mathbf{0}$ , e.g. by minimizing  $\sum_j |\theta_j|$ , then a constant  $0 < \tilde{c} < \infty$  exists such that

$$\tilde{c} \left\| \tilde{\theta} - \tilde{\theta}' \right\| \leq \left\| \tilde{\Phi}(\tilde{\theta} - \tilde{\theta}') \right\| \leq r_{\tilde{\xi}} \quad (5.12.4)$$

follows from (5.12.3). Here

$$\tilde{c} \|\boldsymbol{\theta}\| \leq \|\tilde{\Phi}\boldsymbol{\theta}\| \quad (5.12.5)$$

due to our selection of  $\boldsymbol{\theta} = \mathbf{0}$  if  $\tilde{\Phi}\boldsymbol{\theta} = \mathbf{0}$ , which is always possible since a suitable linear combination of solutions to the corresponding homogeneous systems can be subtracted such that as many parameters as the nullity of  $\tilde{\Phi}$  becomes zero (Friedberg et al., 2003, p. 69,172), and the norm property:  $\|\boldsymbol{\theta}\| = 0$  if and only if  $\boldsymbol{\theta} = \mathbf{0}$  (Friedberg et al., 2003, p. 333). Equation (5.12.4) implies that a constant  $0 < \frac{r_{\tilde{\xi}}}{\tilde{c}} < \infty$  exists, such that

$$\left\| \tilde{\theta} - \tilde{\theta}' \right\| \leq \frac{r_{\tilde{\xi}}}{\tilde{c}} \quad \forall \text{ consistent } \tilde{\xi}, \tilde{\xi} \in \mathcal{U}_{\tilde{\xi}}^{\alpha}. \quad (5.12.6)$$

A constant  $\frac{r_{\tilde{\xi}}}{\tilde{c}} < c < \infty$  therefore exists, such that  $\{\boldsymbol{\theta} \mid \theta_k = c\} \notin \mathcal{F}$ , which implies that Definition 5.5.3 is fulfilled. On the other hand, if  $\tilde{\phi}_k \in \text{pspan}(\tilde{\Phi}_{j \neq k})$  for some consistent  $\tilde{\phi}_k \in \mathcal{U}_{\tilde{\phi}_k}^{\alpha}$ , then some consistent realization  $\tilde{\Phi}_{j \neq k}$  such that  $\tilde{\phi}_k \in \text{span}(\tilde{\Phi}_{j \neq k})$  exists based on Corollary 5.5.1. This implies that  $\tilde{\phi}_k$  is a linear combination of the other regressors and if a feasible solution exists, then  $\theta_k$  can take any value in  $\tilde{\Phi}\boldsymbol{\theta} = \tilde{\xi}$  and Definition 5.5.3 is not fulfilled. A feasible solution only exists when a solution of (5.5.1) practically exists based on Definition 5.5.2 and Definition 5.5.12 is then equivalent to Definition 5.5.3 for all cases with bounded uncertainty in the regressand, i.e. (5.5.11) fulfilled.  $\square$

**Proof of Corollary 5.5.4.** If and only if  $\dim \text{span } \tilde{\Phi} = n$  or  $\tilde{\phi}_k \notin \text{span}(\tilde{\Phi}_{j \neq k})$  for all  $k$ , then the  $n$  vectors in  $\tilde{\Phi}$  are linearly independent and form a basis (Friedberg et al., 2003, p. 46-48). This is sufficient and necessary for the trivial solution to be the only solution of  $\tilde{\Phi}\boldsymbol{\theta} = \mathbf{0}$  for all consistent  $\tilde{\phi}_j \in \mathcal{U}_{\tilde{\phi}_j}^{\alpha}$ , since (5.5.14) holds for all  $\tilde{\Phi} \in \mathcal{V}_C$  and (5.5.15) is equivalent to  $\tilde{\phi}_k \notin \text{span}(\tilde{\Phi}_{j \neq k})$  for all  $\tilde{\Phi} \in \mathcal{V}_C$  based on Corollary 5.5.1.  $\square$

**Proof of Theorem 5.5.5.** Definition 5.5.13 implies that

$$\begin{aligned} \tilde{\phi}_k \notin \text{pspan } \tilde{\Phi}_{j \neq k} & \quad \forall \text{ consistent } \tilde{\phi}_k \in \mathcal{U}_{\phi_k}^\alpha \subseteq \mathbb{R}^m \\ & \quad \text{and } k \in \{1, 2, \dots, n\}, \end{aligned} \quad (5.12.7)$$

based on Corollary 5.5.4, so (5.5.10) is fulfilled for each regressor, independent of (5.5.9), and so is Definition 5.5.12.  $\square$

**Proof of Corollary 5.5.6.**  $\tilde{\phi}_k \in \text{pspan } \tilde{\Phi}_{j \neq k}$  for all consistent  $\tilde{\phi}_k \in \mathcal{U}_{\phi_k}^\alpha$  if and only if a consistent realization  $\tilde{\Phi} \in \mathcal{V}_C$  exists such that  $\tilde{\phi}_k \in \text{span } \tilde{\Phi}_{j \neq k}$  based on Corollary 5.5.1. The set of vectors in this realization is linearly dependent (Friedberg et al., 2003, p. 36, 39). This is thus sufficient and necessary for the existence of a non-trivial solution of  $\tilde{\Phi}\theta = \mathbf{0}$  and fulfilment of Definition 5.5.14.  $\square$

**Proof of Theorem 5.5.7.** If  $\tilde{\Phi} \in \mathcal{V}_C$ , then this follows directly from Definition 5.5.15.  $\square$

**Proof of Corollary 5.5.8.** Practical independence of the set of vectors in the matrix  $\tilde{\Phi}$  is equivalent to  $\dim \text{span}(\tilde{\Phi}) = n$  for all  $\tilde{\Phi} \in \mathcal{V}_C$  based on Corollary 5.5.4. The later is equivalent to  $\text{rank}(\tilde{\Phi}) = n$  for all  $\tilde{\Phi} \in \mathcal{V}_C$ , since  $\text{rank}(\tilde{\Phi}) = \dim \text{span}(\tilde{\Phi})$  (Friedberg et al., 2003, p. 153). This is equivalent to (5.5.20) based on Definition 5.5.15.  $\square$

**Proof of Theorem 5.6.1.** If (5.6.1) holds, then a solution to  $\tilde{\Phi}\theta = \tilde{\xi}$  can only exist for realizations such that  $\tilde{\phi}_k \notin \text{span}(\tilde{\Phi}_{j \neq k})$ . Therefore  $\theta_k \neq 0$  in each feasible solution and the corresponding regressor  $\phi_k$  is practically selectable according to Definition 5.5.5, so this condition is sufficient. If  $\tilde{\xi} \in \text{pspan}(\tilde{\Phi}_{j \neq k})$  for some consistent  $\tilde{\xi} \in \mathcal{U}_\xi^\alpha$ , then some consistent realization  $\tilde{\Phi}_{j \neq k}$  such that  $\tilde{\xi} \in \text{span}(\tilde{\Phi}_{j \neq k}) \subseteq \text{span}(\tilde{\Phi})$  exists based on Corollary 5.5.1. This implies that a feasible solution to  $\tilde{\Phi}\theta = \tilde{\xi}$  exists according to Theorem 5.5.2, which has  $\theta_k = 0$  and the corresponding regressor  $\phi_k$  is not practically selectable according to Definition 5.5.5, so (5.6.1) is necessary. A feasible solution only exists when a solution of (5.5.1) practically exists based on Definition 5.5.2, and the sufficiency and necessity then implies that Definition 5.6.1 is equivalent to Definition 5.5.5.  $\square$

**Proof of Theorem 5.6.2.** When  $\Psi_k$  is practically independent then the set of vectors in  $\tilde{\Psi}_k$  is linearly independent for all consistent  $\tilde{\phi}_j \in \mathcal{U}_{\phi_j}^\alpha \subseteq \mathbb{R}^m$  and  $\tilde{\xi} \in \mathcal{U}_\xi^\alpha \subseteq \mathbb{R}^m$ , based on Definition 5.5.13, and  $\tilde{\xi}$  cannot be expressed as a linear combination of the regressors in any of them. Therefore  $\tilde{\xi} \notin \text{pspan}(\tilde{\Phi}_{j \neq k})$  and  $\phi_k$  is practically selectable according to Definition 5.6.1.  $\square$

**Proof of Theorem 5.6.3.** The assumption that all possible combinations within the neighbourhoods are consistent with the error model implies that within each neighbourhood of any linearly dependent realization  $\tilde{\Phi}$  there exists a linearly independent

realization  $\tilde{\Phi}$  based on Lemma A.3.3. Together with the assumption that  $\tilde{\xi} \in \text{span}(\tilde{\Phi})$  for some  $\tilde{\xi} \in \mathcal{B}_\rho(\phi_j)$  and  $\tilde{\Phi} \in \mathcal{B}_\rho(\Phi)$ , on the other hand, it implies that within each neighbourhood we can find a realization  $\tilde{\Phi}$  and  $\tilde{\xi}$  such that  $\tilde{\Phi}\tilde{\theta} = \tilde{\xi}$  has a solution based on Lemma A.3.4. From Definition 5.6.1 we have that  $\tilde{\xi} \notin \text{pspan}(\Phi_{j \neq k})$  for all consistent realizations with  $\Phi_{j \neq k} = [\phi_1, \dots, \phi_{k-1}, \phi_{k+1}, \dots, \phi_n]$  is both necessary and sufficient for practical selectability of  $\phi_k$ , which within the neighbourhoods is equivalent to

$$\tilde{\xi} \notin \text{span} \tilde{\Phi}_{j \neq k} \quad \forall \tilde{\Phi} \in \mathcal{B}_\rho(\Phi), \tilde{\xi} \in \mathcal{B}_\rho(\xi), \quad (5.12.8)$$

based on Corollary 5.5.1. We next prove that this is equivalent to requiring that  $\Psi_k$  is practically independent, i.e.  $\tilde{\Psi}_k$  is linearly independent for all realizations within the neighbourhoods (Definition 5.5.13). If a linearly dependent realization  $\tilde{\Psi}_k$  exists, then either (i)  $\tilde{\xi} \in \text{span}(\tilde{\Phi}_{j \neq k})$  or (ii)  $\tilde{\Phi}_{j \neq k}$  is linearly dependent. In case (i) then we directly have a contradiction of (5.12.8). In case (ii), if  $\tilde{\Phi}_{j \neq k}$  is linearly dependent, then a linearly independent realization  $\tilde{\Phi}_{j \neq k} \in \mathcal{B}_{\tilde{\rho}}(\tilde{\Phi}_{j \neq k}) \subseteq \mathcal{B}_\rho(\Phi_{j \neq k})$  exists such that  $\tilde{\xi} \in \text{span}(\tilde{\Phi}_{j \neq k})$  for some  $\tilde{\xi} \in \mathcal{B}_\rho(\xi)$  based on Lemma A.3.3, i.e. again a contradiction of (5.12.8). In conclusion,  $\Psi_k$  needs to be practically independent for all realizations within the neighbourhoods for Definition 5.6.1 to be fulfilled.  $\square$

**Proof of Theorem 5.6.4.** This follows from the necessity of Definition 5.6.1 for practical selectability of the corresponding regressor, since the parameter is zero only when  $\tilde{\xi} \in \text{span}(\tilde{\Phi}_{j \neq k})$ , which is equivalent to  $\tilde{\Psi}_k$  having collinear columns.  $\square$

**Proof of Corollary 5.6.5.** This follows from Theorem 5.6.2, since practical independence of the set of vectors in the matrix  $\Psi_k$  is equivalent to  $\dim \text{span}(\tilde{\Psi}_k) = n$  for all consistent  $\tilde{\phi}_j \in \mathcal{U}_{\phi_j}^\alpha$  and  $\tilde{\xi} \in \mathcal{U}_\xi^\alpha$  based on Corollary 5.5.4. The latter is equivalent to  $\text{rank}(\tilde{\Psi}_k) = n$  for all consistent  $\tilde{\phi}_j \in \mathcal{U}_{\phi_j}^\alpha$  and  $\tilde{\xi} \in \mathcal{U}_\xi^\alpha$ , since  $\text{rank}(\tilde{\Psi}_k) = \dim \text{span}(\tilde{\Psi}_k)$  (Friedberg et al., 2003, p. 153). The use of Definition 5.5.15 finally enables us to recognize (5.6.7).  $\square$

**Proof of Theorem 5.6.6.** For each consistent realization such that  $\tilde{h}_k^T \tilde{\xi} > 0$ , we have that  $\theta_k > 0$  if  $\tilde{\Phi}\tilde{\theta} = \tilde{\xi}$  has a solution, based on Theorem A.3.6. Similarly, for each consistent realization such that  $\tilde{h}_k^T \tilde{\xi} < 0 \Leftrightarrow -\tilde{h}_k^T \tilde{\xi} > 0$ , then  $-\theta_k > 0 \Leftrightarrow \theta_k < 0$  if  $\tilde{\Phi}\tilde{\theta} = \tilde{\xi}$  has a solution, based on Theorem A.3.6.  $\tilde{\Phi}\tilde{\theta} = \tilde{\xi}$  has a feasible solution if and only if  $\tilde{\xi} \in \text{pspan}(\Phi)$ , since only then a consistent regressor realization such that  $\tilde{\xi} \in \text{span}(\tilde{\Phi})$  exists based on Corollary 5.5.1, which is necessary and sufficient for existence of a solution (Friedberg et al., 2003, p. 174). A solution of (5.5.1) practically exists if and only if  $\mathcal{F} \neq \emptyset$  based on Definition 5.5.2. Hence (5.6.9) is equivalent to  $\theta_k > 0 \forall \theta \in \mathcal{F}$  or  $\theta_k < 0 \forall \theta \in \mathcal{F}$  and Definition 5.6.6 is equivalent to Definition 5.5.8.  $\square$

**Proof of Theorem 5.6.7.** We prove (a) first and then (b) based on it.

- (a) The set of consistent realizations for which a feasible solution exists is a subset of all consistent realizations, so together with the requirement of practical existence the condition in (a) implies that Theorem 5.6.6 is fulfilled.
- (b) The intersection of all positive halfspaces (5.6.11) or negative halfspaces (5.6.12) is a set such that  $\tilde{\mathbf{z}}^T \tilde{\mathbf{h}}_k > 0$  in the former case for each vector  $\tilde{\mathbf{z}} \in \mathcal{H}_+$  for all consistent regressor combinations, while  $\tilde{\mathbf{z}}^T \tilde{\mathbf{h}}_k < 0$  in the later case for each vector  $\tilde{\mathbf{z}} \in \mathcal{H}_-$  for all consistent regressor combinations. The condition in (a) is therefore fulfilled.

□

**Proof of Theorem 5.6.8.** When  $\Psi_k$  is practically independent then the regressor  $\phi_k$  is practically selectable based on Theorem 5.6.2, the set of vectors in  $\tilde{\Psi}_k$  is linearly independent for all consistent  $\tilde{\phi}_j \in \mathcal{U}_{\phi_j}^\alpha \subseteq \mathbb{R}^m$  and  $\tilde{\xi} \in \mathcal{U}_\xi \subseteq \mathbb{R}^m$  based on Definition 5.5.13, and  $\tilde{\xi}$  cannot be expressed as a linear combination of the regressors in any of them. Similarly, when  $\Phi$  is practically independent then the set of regressors in  $\tilde{\Phi}$  is linearly independent for all consistent  $\tilde{\phi}_j \in \mathcal{U}_{\phi_j}^\alpha \subseteq \mathbb{R}^m$ , based on Definition 5.5.13, and  $\tilde{\phi}_k$  cannot be expressed as a linear combination of the other regressors in any of them. The regressor  $\phi_k$  is therefore always outside of the subspace defined by the other regressors and  $\tilde{\mathbf{h}}_k \neq \mathbf{0}$  based on its definition (5.6.10) and Lemma A.3.5. Therefore all consistent realizations, such that  $\tilde{\Phi}\theta = \tilde{\xi}$  has a solution, contain a  $\tilde{\xi}$  such that either  $\tilde{\mathbf{h}}_k^T \tilde{\xi} > 0$  or  $\tilde{\mathbf{h}}_k^T \tilde{\xi} < 0$ . In all other consistent realizations  $\tilde{\Phi}\theta = \tilde{\xi}$  has no solution. To prove that Theorem 5.6.6 is fulfilled all consistent realizations  $\tilde{\phi}_k$ , however, need to be on the same side or in the separating hyperplane defined by  $\tilde{\mathbf{h}}_k$  in (5.6.10), and a feasible solution of (5.5.1) needs to exist. Additional conditions are needed and we next prove sufficiency of each of the ones given in the theorem, one-by-one.

- (a) When  $m = n$ , practical independence of the regressors implies  $\text{span}(\tilde{\Phi}) = \mathbb{R}^m$  for all consistent  $\tilde{\Phi} \in \mathcal{V}_C$ , so each consistent realization  $\tilde{\Phi}_{i \neq k}$  defines a separating hyperplane  $\tilde{\mathbf{h}}_k$  based on (5.6.10) and a solution of  $\tilde{\Phi}\theta = \tilde{\xi}$  exists for every consistent realization. The uncertainty sets are connected sets, so if a consistent realization  $[\tilde{\Phi}, \tilde{\xi}]$  with  $\tilde{\mathbf{h}}_k^T \tilde{\xi} > 0$  and another consistent realization  $[\tilde{\Phi}, \tilde{\xi}]$  with  $\tilde{\mathbf{h}}_k^T \tilde{\xi} < 0$  exist, then also a consistent realization  $[\tilde{\Phi}, \tilde{\xi}]$  with  $\tilde{\mathbf{h}}_k^T \tilde{\xi} = 0$  exists, since a change from positive to negative requires zero at an intermediary point. Practical independence of  $\Phi$  implies that  $\tilde{\mathbf{h}}_k \neq \mathbf{0}$ , which together with  $\tilde{\mathbf{h}}_k^T \tilde{\xi} = 0$  and  $\tilde{\Phi}\theta = \tilde{\xi}$  having a solution implies that  $\tilde{\xi}$  lies in the separating hyperplane and  $\tilde{\Psi}_k$  needs to be linearly dependent. This constitutes a contradiction, since  $\Psi_k$  is practically independent, so  $\tilde{\mathbf{h}}_k^T \tilde{\xi}$  must therefore be either positive or negative for all consistent realizations with a solution and Theorem 5.6.6 is fulfilled.

(b) If the set of regressand realizations  $\tilde{\xi}$  for which a solution of  $\tilde{\Phi}\tilde{\theta} = \tilde{\xi}$  exists for some consistent  $\tilde{\Phi} \in \mathcal{V}_C$  is a non-empty connected subset of  $\mathbb{R}^m$ , then either only one regressand realization with a solution exists or there exists at least one other realization with a solution belonging to the set in a neighbourhood of each  $\tilde{\xi}$  as shown in Lemma A.3.4, i.e. within a ball of radius  $\rho > 0$ , see e.g. (Khuri, 2003, p. 11-12). This implies that from each  $\tilde{\xi}$  with a solution there is a path to each other  $\tilde{\xi}$  with a solution that only contains regressand realizations  $\tilde{\xi}$  that have a solution. Therefore if a realization  $[\tilde{\Phi}, \tilde{\xi}]$  with  $\tilde{h}_k^T \tilde{\xi} > 0$  and a solution to  $\tilde{\Phi}\tilde{\theta} = \tilde{\xi}$ , and another realization  $[\tilde{\Phi}, \tilde{\xi}']$  with  $\tilde{h}_k^T \tilde{\xi}' < 0$  and a solution to  $\tilde{\Phi}\tilde{\theta}' = \tilde{\xi}'$  exist, then also a realization  $[\tilde{\Phi}, \tilde{\xi}^*]$  with  $\tilde{h}_k^T \tilde{\xi}^* = 0$  and a solution to  $\tilde{\Phi}\tilde{\theta}^* = \tilde{\xi}^*$  exists, since a change from positive to negative requires zero at an intermediary point. Practical independence of  $\Phi$  implies that  $\tilde{h}_k \neq 0$ , which together with  $\tilde{h}_k^T \tilde{\xi} = 0$  and  $\tilde{\Phi}\tilde{\theta} = \tilde{\xi}$  having a solution implies that  $\tilde{\Psi}_k$  need to be linearly dependent. This constitutes a contradiction, since  $\Psi_k$  is practically independent, so  $\tilde{h}_k^T \tilde{\xi}$  must therefore be either positive or negative for all consistent realizations with a solution and Theorem 5.6.6 is fulfilled.  $\square$

**Proof of Theorem 5.7.1.** Practical independence of  $\Phi$  implies that  $\tilde{\phi}_k \notin \text{pspan}(\Phi_{j \neq k})$  for all consistent  $\tilde{\phi}_k \in \mathcal{U}_{\phi_k}^\alpha$  and any  $k \in \{1, 2, \dots, n\}$ , with  $\Phi_{j \neq k} = [\phi_1, \dots, \phi_{k-1}, \phi_{k+1}, \dots, \phi_n]$ , based on Corollary 5.5.4. Therefore practical collinearity of  $\Psi_k$  implies that  $\tilde{\xi} \in \text{pspan}(\Phi_{j \neq k})$  for all consistent  $\tilde{\xi} \in \mathcal{U}_\xi^\alpha \subseteq \mathbb{R}^m$  based on Corollary 5.5.6 and Definition 5.7.2 is fulfilled. Note that practical collinearity of  $\Psi_k$  is not even sufficient for practical omissibility, since some of the regressors might instead of the regressand be in the practical span of the other vectors of  $\Psi_k$ .  $\square$

**Proof of Theorem 5.8.1.** Here  $\mathcal{S}$  is a subset of the regressors in  $\Phi_{j \neq k}$  and (5.8.2) therefore holds if (5.8.3) holds. The subset  $\mathcal{S}$  may contain all regressors except  $\phi_k$ , so (5.8.2) only holds if (5.8.3) holds. Condition (5.8.4) only implies that the subset  $\mathcal{S}$  contains a minimal set of regressors required for existence of a solution of (5.5.1). The conditions in this theorem are therefore equivalent to the conditions in Definition 5.8.2.  $\square$

**Proof of Theorem 5.8.2.** Practical independence of the regressors in the subset  $\mathcal{S}$  implies that  $\tilde{\phi}_l \notin \text{pspan}(\Phi_{\mathcal{S} \setminus l})$  for all consistent  $\tilde{\phi}_l \in \mathcal{U}_{\phi_l}^\alpha$  and any  $l \in \mathcal{S}$ , with  $\Phi_{\mathcal{S} \setminus l}$  defined as in Theorem 5.8.1, based on Corollary 5.5.4. Therefore practical collinearity of  $\{\phi_j | j \in \mathcal{S}\} \cup \xi$  implies that (5.8.3) holds based on Corollary 5.5.6 and together with the other conditions that Theorem 5.8.1 is fulfilled.  $\square$

**Proof of Corollary 5.8.3.** These are the conditions for when a regressor is not practically selectable, excludable, nor negligible, taken from Definition 5.6.1, Definition 5.7.1, Definition 5.7.2, and Definition 5.8.2.  $\square$

**Proof of Theorem 5.10.1.** Express each realization of the variables as  $\tilde{\phi}_j = \phi_j - \mathbf{v}_j$  with  $\mathbf{v}_j \in \mathcal{U}_{\phi_j}^\alpha$ , i.e. a sum of the observed variable and an error realization. Based on Definition 5.5.13, the set of uncertain vectors in the matrix  $\Phi = [\phi_1, \dots, \phi_j, \dots, \phi_n]$  is practically independent if  $\theta = \mathbf{0}$  is the only solution of

$$(\Phi - \Upsilon)\theta = \mathbf{0}, \quad \forall \mathbf{v}_j \in \mathcal{U}_{\phi_j}^{\text{ell}}, \quad (5.12.9)$$

with  $\Upsilon = [\mathbf{v}_1, \dots, \mathbf{v}_j, \dots, \mathbf{v}_n]$ , since  $\mathcal{U}_{\phi_j}^\alpha$  is a subset of  $\mathcal{U}_{\phi_j}^{\text{ell}}$ . Solution of the structured singular value problem (5.10.2) yields the smallest

$$\gamma(\Phi) = \bar{\sigma}(\Delta) = \|\Delta\|_2 = \|\text{diag}(\delta_1, \dots, \delta_j, \dots, \delta_n)\|_2 = \max_j \|\delta_j\|_2, \quad (5.12.10)$$

such that

$$\det(\mathbf{I} - \Phi^\dagger \mathbf{W}_1 \Delta) = 0, \quad (5.12.11)$$

see e.g. Skogestad and Postlethwaite (1996, p. 331, 540), so we only need to show that  $\theta = \mathbf{0}$  is the only solution of (5.12.9) when  $\gamma(\Phi) > 1$ . To do this, we first show that (5.12.9) has a non-trivial solution  $\theta \neq \mathbf{0}$  when  $\Upsilon = \mathbf{W}_1 \Delta$  is inserted with the  $\Delta$  obtained by solving (5.10.2) and then that this  $\Delta$  has smallest induced 2-norm of all  $\Delta$  for which (5.12.9) has a non-trivial solution. Note that if  $\text{rank}(\Phi) < n$ , then a non-trivial solution of (5.12.9) exists for  $\Upsilon = \mathbf{0}$  and  $\gamma(\Phi) = 0$ . Otherwise  $\text{rank}(\Phi) = n$  and  $\Phi^T \Phi$  is invertible and the Moore-Penrose generalized inverse  $\Phi^\dagger = (\Phi^T \Phi)^{-1} \Phi^T$  (Meyer, 2000, p. 423, 428). To obtain  $\mathbf{I} - \Phi^\dagger \mathbf{W}_1 \Delta$  from (5.12.9) we first pre-multiply by  $\Phi^T$ , which gives

$$\Phi^T (\Phi - \Upsilon)\theta = \mathbf{0}, \quad (5.12.12)$$

we then pre-multiply by  $(\Phi^T \Phi)^{-1}$  and insert  $\Upsilon = \mathbf{W}_1 \Delta$ , which after simplification gives

$$(\mathbf{I} - \Phi^\dagger \mathbf{W}_1 \Delta)\theta = \mathbf{0}. \quad (5.12.13)$$

Here  $(\Phi^T \Phi)^{-1}$  is invertible with rank  $n$  and  $\mathbf{I} - \Phi^\dagger \mathbf{W}_1 \Delta$  is a square matrix, so a non-trivial solution  $\theta \neq \mathbf{0}$  of (5.12.12) exists if and only if (5.12.11) holds (Horn and Johnson, 1990, p. 13-14). Clearly, (5.12.12) has a non-trivial solution when (5.12.9) has one, but it also has one for any  $\mathbf{z} \neq \mathbf{0}$  such that

$$(\Phi - \Upsilon)\theta = \mathbf{z} \quad \text{and} \quad \Phi^T \mathbf{z} = \mathbf{0}, \quad (5.12.14)$$

i.e. when  $\mathbf{z}$  belongs to the null space of  $\Phi^T$ . We need to prove that the non-trivial solution of (5.12.12) with smallest induced 2-norm is the non-trivial solution of (5.12.9), which we do next. Any such  $\mathbf{z} \neq \mathbf{0}$ , however, belongs to the range of  $\Phi - \Upsilon$ , which is orthogonal to the null space of  $\Phi^T - \Upsilon^T$  (Horn and Johnson, 1990, p. 5,

17), and  $\mathbf{Y}$  must therefore remove one dimension from the range of  $\Phi$  and add the dimension of  $\mathbf{z}$  to the range of  $\Phi - \mathbf{Y}$ . Any such  $\mathbf{Y}$  for which (5.12.14) has a solution is, in other words, composed of three orthogonal parts  $\mathbf{Y} = \mathbf{Y}_1 + \mathbf{Y}_2 + \mathbf{Y}_3$ , where  $\mathbf{Y}_1$  removes one dimension of the range of  $\Phi$  such that  $(\Phi - \mathbf{Y}_1)\theta = 0$  for a  $\theta \neq \mathbf{0}$ , while  $\mathbf{Y}_2$  adds a dimension of the null space of  $\Phi^T$  such that  $(\Phi - \mathbf{Y}_1 - \mathbf{Y}_2)\theta = \mathbf{z}$  for a  $\theta \neq \mathbf{0}$  and  $\Phi^T \mathbf{z} = \mathbf{0}$ . The third part is orthogonal to the former two, in the range of  $\Phi$ , and has no influence on the existence of a non-trivial solution of (5.12.12). The ranges of these three matrices are orthogonal and for any induced norm

$$\|\mathbf{Y}_1\| < \|\mathbf{Y}_1 + \mathbf{Y}_2 + \mathbf{Y}_3\| \quad \text{when } \mathbf{Y}_2 \neq \mathbf{0} \text{ or } \mathbf{Y}_3 \neq \mathbf{0}, \quad (5.12.15)$$

since  $\|\mathbf{v}_1 + \mathbf{v}_2\| = \|\mathbf{v}_1\| + \|\mathbf{v}_2\|$  for any  $\mathbf{v}_1^T \mathbf{v}_2 = 0$  and  $\|\mathbf{v}_k\| > 0$  for any  $\mathbf{v}_k \neq \mathbf{0}$  (Horn and Johnson, 1990, p. 256, 290, 294). In other words, addition of any part but the one that removes a dimension from the range of  $\Phi$  increases the induced norm of  $\mathbf{Y}$ . Here  $\mathbf{Y} = \mathbf{W}_1 \Delta$  with  $\mathbf{W}_1$  being a constant matrix of rank  $n$ , so the smallest induced 2-norm of  $\Delta$  such that a non-trivial solution of (5.12.9) exists is given by solution of (5.10.2). Finally, if  $\gamma(\Phi) > 1$ , then  $\|\delta_j\|_2 > 1$  based on (5.12.10) and

$$\mathbf{v}_j^T \mathbf{Q}_{\phi_j} \mathbf{v}_j = \delta_j^T \mathbf{Q}_{\phi_j}^{-\frac{1}{2}T} \mathbf{Q}_{\phi_j} \mathbf{Q}_{\phi_j}^{-\frac{1}{2}} \delta_j = \delta_j^T \delta_j = \|\delta_j\|_2^2 > 1 \quad (5.12.16)$$

for some  $j \in \{1, \dots, n\}$ . This implies that the closest realization  $\mathbf{Y} = \mathbf{W}_1 \Delta$  such that a non-trivial solution  $\theta \neq \mathbf{0}$  of (5.12.9) exists is outside of the ellipsoidal uncertainty set in (5.10.4) and  $\Phi$  is therefore practically independent.  $\square$

**Proof of Theorem 5.10.2.** Each realization of the variables is expressed as  $\tilde{\phi}_j = \phi_j - \mathbf{v}_j$ , i.e. a sum of the observed variable and an error realization. To prove this theorem we first show that  $\theta = \mathbf{0}$  is the only solution of

$$(\Phi - \mathbf{Y})\theta = \mathbf{0}, \quad (5.12.17)$$

for error realizations  $\mathbf{Y}$  with  $\|\mathbf{Y}\|_2 < \sigma_n(\Phi)$  and then that all realizations with larger induced 2-norm are rejected. The induced 2-norm of any matrix  $\mathbf{Y} \in \mathbb{R}^{m \times n}$  such that  $\text{rank}(\Phi - \mathbf{Y}) < n$  is

$$\|\mathbf{Y}\|_2 \geq \sigma_n(\Phi) \quad (5.12.18)$$

for any matrix  $\Phi \in \mathbb{R}^{m \times n}$  with  $m \geq n$  (Meyer, 2000, p. 421). In other words, there exist no matrix with smaller induced 2-norm than  $\mathbf{Y} = \sigma_n \mathbf{u}_n \mathbf{v}_n^T$ , which is constructed using the  $n$ th largest singular value of  $\Phi$  and corresponding singular vectors  $\mathbf{u}_n$  and  $\mathbf{v}_n$ , since it has  $\|\mathbf{Y}\|_2 = \bar{\sigma}(\mathbf{Y}) = \sigma_n(\Phi)$  (Meyer, 2000, p. 412). Hence for all error realizations  $\mathbf{Y}$  with smaller induced 2-norm  $\text{rank}(\Phi - \mathbf{Y}) = n$  and  $\theta = \mathbf{0}$  is the only solution of (5.12.17), since the columns of  $\Phi - \mathbf{Y}$  are linearly independent (Friedberg et al., 2003, p. 36-37, 153). Next, any realization of normally distributed errors  $\mathbf{Y}$  such that

$$\tilde{\mathbf{Y}}^T \Lambda_{\tilde{\mathbf{Y}}}^{-1} \tilde{\mathbf{Y}} > \chi^2(\alpha, nm) \quad (5.12.19)$$

is rejected based on the  $\chi^2$  hypothesis test at significance level  $\alpha$  (see e.g. [Chew, 1966](#)). For the matrix with smallest induced 2-norm,  $\mathbf{Y} = \sigma_n \mathbf{u}_n \mathbf{v}_n^T$ , we get

$$\tilde{\mathbf{Y}}^T \Lambda_{\mathbf{Y}}^{-1} \tilde{\mathbf{Y}} = \frac{1}{\lambda} \|\mathbf{Y}\|_{\text{Fro}}^2 \geq \frac{1}{\lambda} \|\mathbf{Y}\|_2^2 = \chi^{-2}(\alpha, nm) w_1^{-2} \sigma_n^2(\Phi) = \chi^{-2}(\alpha, nm) \gamma^2(\Phi). \quad (5.12.20)$$

by using  $\Lambda_{\mathbf{Y}} = \lambda \mathbf{I}$ , [\(5.10.7\)](#), [\(5.10.6\)](#), and standard properties of norms ([Horn and Johnson, 1990](#), p. 290,294). In general the Frobenius norm of a matrix is greater or equal to the induced 2-norm ([Horn and Johnson, 1990](#), p. 314), but for a rank one matrix such as  $\mathbf{Y} = \sigma_n \mathbf{u}_n \mathbf{v}_n^T$  it is always equal, since  $\|\mathbf{Y}\|_{\text{Fro}} = \sqrt{\sum \sigma_j^2(\mathbf{Y})}$  ([Skogestad and Postlethwaite, 1996](#), p. 556), implying that any error realization  $\mathbf{Y}$  with  $\|\mathbf{Y}\|_2 > \sigma_n(\Phi)$  is rejected based on [\(5.12.19\)](#) if and only if the confidence score  $\gamma(\Phi) > 1$  and  $\Phi$  is therefore practically independent for all non-rejectable realizations at significance level  $\alpha$ .  $\square$



---

# Inference of subnetworks—impact of latent states

---

*“The organism is not a closed, but an open system.”*

Karl Ludwig von Bertalanffy, General system theory:  
Foundations, development, applications, 1969.

In inference of gene regulatory networks it is common practice to only observe the change in mRNA abundance, neglecting other components that affects the state of the cell such as proteins and metabolites. These constitute unobserved and unperturbed latent state variables and we here study their impact in network inference. We use linear time invariant models and well known results from systems theory to show that it is necessary to know a linearly independent subset of as many relations between the states of interest, which constitute nodes in the inferred subnetwork, and responses or perturbations as there are states of interest. Otherwise the states of interest are not uniquely defined and consequently neither is the structure of the subnetwork. No information about the latent states is required, but their existence implies that a subnetwork of pseudo-direct causal influences, accounting for all environmental effects, in general is inferred. In principle, the number of latent states and different paths between the states of interest can be estimated, but their identity cannot be determined without direct perturbation or observation of them. For illustration, we construct an extended model of the published synthetic IRMA network by including latent states based on published mechanisms and succeed to explain the steady-state data observed by culturing this strain of *S. cerevisiae* with addition of only one unknown component. Our IRMA model illustrates that interactions inferred from perturbation experiments capture influences of regulatory importance. But these links do not in general correspond to physical binding between observed molecules, which also has been noted elsewhere.

## 6.1 Introduction

Biological functions are created through interactions among a large number of components within complex networks. One of the key problems in systems biology is that of inferring these networks based on experimental response data. In particular, the inference of gene regulatory networks (GRNs) based on perturbation experiments has been minutely investigated and reviewed during the past decade (Lecca et al., 2011; Baralla et al., 2009; Hecker et al., 2009; Karlebach and Shamir, 2008; Bonneau, 2008; Tegnér and Björkegren, 2007; Cho et al., 2007; Bansal et al., 2007; Doyle and Lauffenburger, 2005; Filkov, 2005; de Jong, 2002; D’haeseleer et al., 2000). The potential of inference of GRNs from perturbation experiments has been demonstrated in different contexts for a number of organisms and using different types of *in vivo* data, most notably in Gardner et al. (2003); Bonneau et al. (2006); Lorenz et al. (2009); Cantone et al. (2009); Jörnsten et al. (2011). Ideally, the inferred interactions should correspond to direct causal influences of regulatory importance that exist between the genes, thereby providing mechanistic insight, and at the same time be useful for prediction of the response to novel perturbations. However, it is not always clear how inferred interactions should be interpreted.

So far, the main focus within this area has been on algorithms for obtaining models from given response data. However, prior to feeding experimental response data to an inference algorithm, one should ask what can in fact be deduced about the underlying network based on the given data. The ability to infer interactions and networks is limited partly by the fundamental information content in the available data set, and partly by the uncertainty present in the data. The latter problem, limitations imposed by uncertainty in measurements and perturbations was treated in Chapters 4-5 of this thesis. The former problem has received significant attention in past years, but then almost exclusively limited to the problem resulting from having a limited number of experiments and samples. In particular, the underdetermined inference problems that result when the number of samples falls short of the network size, *i.e.* the number of genes in case of GRNs, has been addressed in numerous studies (Hecker et al., 2009; Tegnér and Björkegren, 2007). Its popularity reflects the current cost structure of biological experiments; it is considerably cheaper to measure the expression change of a thousand genes in one sample than the expression change of one gene in a thousand experiments. The fact that no reliable measure of confidence can be assigned to inferred interactions when the number of nodes exceed the number of samples, as was shown in Chapter 4, has, however, been overlooked. The current cost structure makes it important to focus on a subset of the most important genes when significance at the link level is desired.

In this work we consider limitations imposed by the fact that not all components in the underlying network can be directly observed or perturbed. We will refer to components that take part in the network interactions, but which are not measured and not perturbed, as latent states. Such states are present in essentially all networks of biological relevance. For instance, it is common practice when inferring GRNs to measure only the mRNA abundance for each considered gene, while the true

regulatory interactions also involve proteins, metabolites, and other biochemical components. Furthermore, while a given cell typically contains thousands of active genes, one usually only attempts to infer a network for a selected subset of these. The problem of inferring the interactions among a subset of the system states is here referred to as subnetwork inference. In this work we derive necessary conditions for inference of subnetworks based on data from perturbation experiments.

The motivation behind the work presented here partly stems from a previous study of the five gene IRMA network engineered by [Cantone et al. \(2009\)](#) in *S. cerevisiae*. When analysing their *in vivo* steady-state response data we found that their data could not be explained without a transcriptional activation of the *SWI5* gene by the *CBF1* gene, see Section 4.6. This interaction should not exist based on the known engineered structure of the network, nor have we found any evidence for it in the literature; see Section B.1.2 for details. We also found that several of the interactions present in the engineered structure were not needed to explain the response data. To explain this apparent discrepancy, we consider the influence of mRNAs and proteins neglected by Cantone *et al.* These constitute unperturbed and unobserved latent state variables of the system, and as we show they have a significant impact on the network behaviour. Furthermore, we show that the latent states also have a strong impact on the meaning of an inferred interaction. As an example, we construct an extended IRMA network that can explain the data and account for previously published mechanisms in the simplest possible way.

Limitations in network inference based on perturbation data in which not all states are measured or perturbed have previously been discussed in [Margolin and Califano \(2007\)](#); [Goncalves and Warnick \(2008\)](#); [He et al. \(2009\)](#); [Hendrickx et al. \(2011\)](#). Implications of latent states on network inference based on mutual information is discussed in [Margolin and Califano \(2007\)](#). They use a simple co-regulation model in which two genes are regulated by a single transcription factor (TF) to show that an edge erroneously will be inferred between the co-regulated genes instead of the edges between the TF and the genes, unless the latent protein state of the TF correlates strongly enough with the TF mRNA. They conclude that it is important to characterize the necessary correlation strength for correct inference, since evidence for relatively weak correlation between the mRNA and protein of TFs is increasing. This is a requirement for inference algorithms based on mutual information or correlations, but it is not clear based on their work if it has any implications on the fundamental information content required for inference. [Goncalves and Warnick \(2008\)](#) establish necessary and sufficient conditions for dynamical structure reconstruction of linear time invariant (LTI) systems from knowledge of the transfer function from input (perturbations) to output (responses). The dynamical structure function is a representation of LTI systems introduced in [Goncalves et al. \(2007\)](#), which contains the transfer functions between the observed states and the transfer functions from the inputs to the observed states. It is a special case in the sense that the true full state-space description of the underlying system in general only can be recovered from it if all state variables are observed. For the case of inferring subnetworks, [Goncalves and Warnick \(2008\)](#) derive a necessary

condition on the experimental data for inference of the dynamical structure function. Essentially, the condition they derive states that as many experiments as there are nodes in the subnetwork are necessary for network reconstruction. However, this condition is contradicted by the earlier results in [Schmidt et al. \(2005\)](#) where a complete GRN is reconstructed based on a single experiment with time-series data. We here provide an explanation for the contradiction and derive an alternative and strictly necessary condition for inference of subnetworks.

As stated above, one usually seeks direct interactions between components when inferring intracellular networks. Inference of physical interactions versus regulatory influences are discussed in [Gardner and Faith \(2005\)](#). The physical approach seeks the regulatory protein factors that physically bind to the promoter of the regulated transcript, while the influence approach seeks transcripts whose concentration changes can explain the changes of a considered transcript. Interactions discovered by the former approach are clearly direct due to the physical binding. In the latter case, however, components can act on each other directly, *e.g.* a protein acting as a transcription factor for a gene, or indirectly, *e.g.* an mRNA producing a protein that binds to another protein and the resulting complex acts as a transcription factor. Usually, it is not clear what type of interactions are reported in the literature on GRN inference, which inevitably causes confusion. Considering the number of works demonstrating the potential of reverse engineering and the limited actual contribution towards understanding how gene regulation works, we think it is time to examine and explain what an inferred interaction represents. We show that the number of latent intermediary states and the length of the paths that constitute an inferred interaction can be determined from the order and relative degree of its corresponding transfer function.

In order to make the presentation readily accessible to system biologists with diverse backgrounds we restrict our consideration to deterministic and linear models of GRNs, and use a GRN synthesized in *S. cerevisiae* for illustration. We start by presenting and further motivating the mentioned model formalism and assumptions we employ. We also present the IRMA example that will be used for illustration throughout the paper. In [Section 6.3](#) we employ well known results from system theory to show that in order to infer the full network one needs to know for each state either how it is related to the observed quantities or how it is affected by the perturbed quantities, as well as ensure that both observability and controllability is fulfilled. For the case with latent states, we show in [Section 6.4](#) that knowledge of a linearly independent subset of as many relations as states of interest, between the states of interest and responses or perturbations, is a prerequisite. More precisely, observability and controllability of the states of interest together with the above condition are necessary and sufficient for inference of a subnetwork. We then in [Section 6.5](#) re-examine the results derived by [Goncalves and Warnick \(2008\)](#). As discussed above, it is usually of interest to know if interactions are direct or indirect and we therefore in [Section 6.6](#) consider conditions under which one can classify inferred interactions as direct or indirect and, in the latter case, the minimum number of latent states that exist between two nodes in the network. At the end

we discuss some implications of the results for inference of GRNs, reflecting over previous work and giving recommendations for future works.

## 6.2 Problem definition and the IRMA gene regulatory network

We consider inference of interactions between biochemical components based on perturbation experiments. The change in concentration, or activity, of component  $i$  is assumed to be described by a linear time-invariant differential equation

$$\frac{dx_i}{dt}(t) = \sum_{j=1}^n a_{ij}x_j(t) + \sum_{k=1}^l b_{ik}p_k(t) \quad (6.2.1)$$

where the  $n$  concentrations  $x_j$  are the states of the system. Here the parameter  $a_{ij}$  describes the direct effect of a change in the concentration of component  $x_j$  on  $x_i$  and  $b_{ik}$  is the direct effect of the applied perturbation  $p_k$  on  $x_i$ . The linearity assumption can be motivated either by having linear reaction kinetics and transport mechanisms, or by having relatively small changes in the concentrations  $x_j$ . We also assume that we have measurements  $y_i$  of  $o$  quantities that can be described as a linear combination of the states  $x_j$

$$y_i(t) = \sum_{j=1}^n c_{ij}x_j(t). \quad (6.2.2)$$

In matrix form we then obtain the standard linear state-space model

$$\frac{d\mathbf{x}}{dt}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{p}(t) \quad (6.2.3a)$$

$$\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t). \quad (6.2.3b)$$

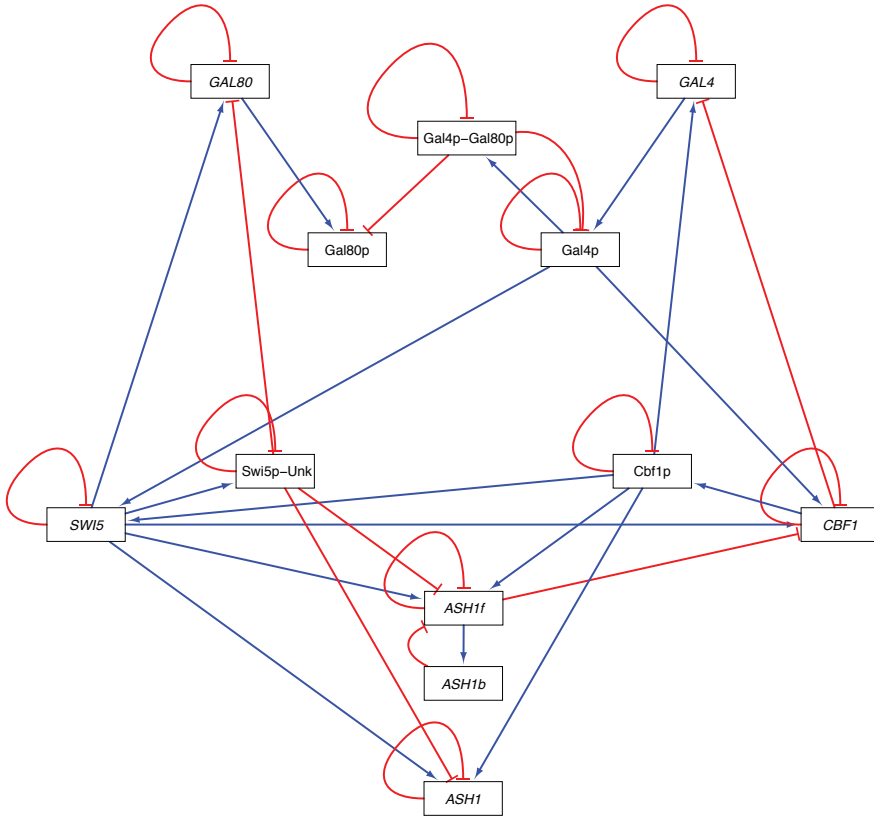
The interaction matrix  $\mathbf{A}$  describes the network structure; a non-zero element  $a_{ij}$  implies a direct interaction, or edge, from component  $j$  to component  $i$  which is upregulating (downregulating) if  $a_{ij}$  is positive (negative). The problem considered in network inference is that of determining the interactions that exist between the components, *i.e.* primarily to determine the structure of  $\mathbf{A}$ . The sign of the interactions is considered secondary and the strength of the interactions, *i.e.* the magnitude of the non-zero elements  $a_{ij}$ , is considered subsidiary. We further restrict ourselves to the case when  $\mathbf{A}$  is completely unknown, *i.e.* no *a priori* information on which links that exist, their signs, or strength is available, but  $\mathbf{C}$  and  $\mathbf{B}$  are assumed known. This situation is common in reverse engineering of GRNs. It is important to note that the problem of inferring the network structure in general is different from that of identifying a predictive model. In particular, a good predictive model does not need to capture the structure of the underlying system as discussed in Section 3.6.

As stated in the Introduction, one motivation for choosing a deterministic linear time invariant description of the network here is to keep the presentation at a level that is accessible to biologists in general and systems biologists in particular. However, we also note that deterministic LTI models are a subset of the class of nonlinear and stochastic models, and hence if one cannot solve the network inference problem for this subset then one cannot expect to solve it for more complex models either. We assume here that the response in all observed variables in the vector  $\mathbf{y}(t)$  to every possible perturbation of the perturbed variables in the vector  $\mathbf{p}(t)$  is known precisely without any noise or measurement errors, *i.e.* that we know the input-output behaviour of the system. In this way we can distinguish between limitations imposed by the fundamental information content in the data and limitations imposed by uncertainty which were discussed in Chapters 4-5. All the necessary conditions that we establish for network inference based on the deterministic case are also necessary when real noisy data are used.

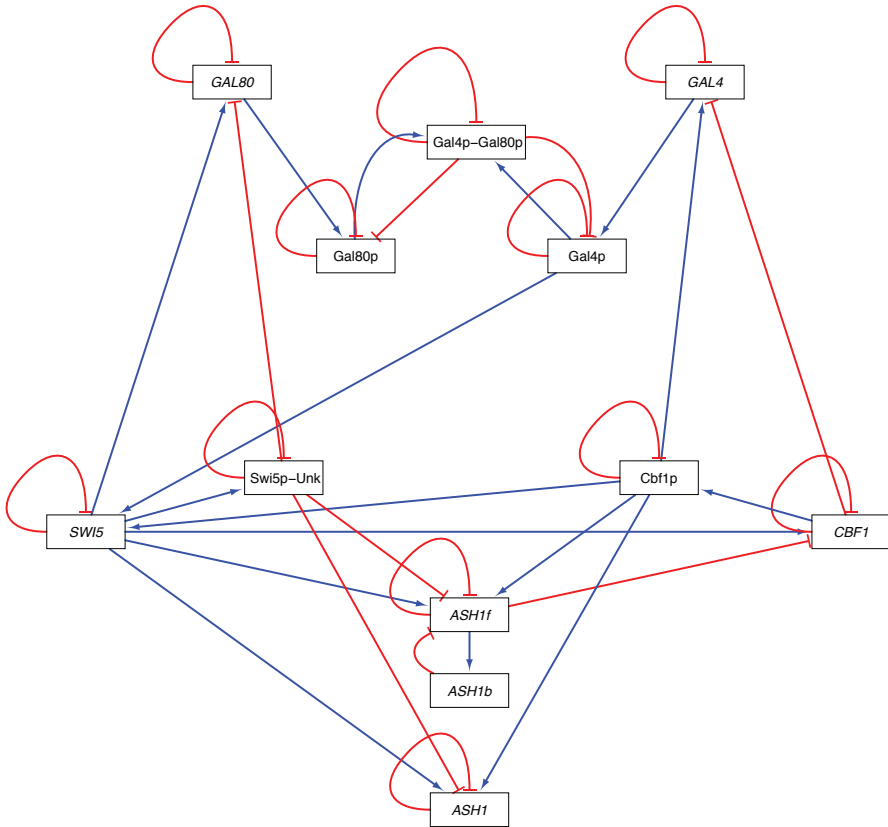
As an example we consider an extended version of the IRMA network that we constructed to explain the steady-state data recorded by Cantone *et al.* (2009). The extension was performed by as far as possible incorporating published mechanisms and in the simplest of ways. Note that we did not attempt to include all known mechanisms nor explain the time-series data recorded by Cantone *et al.* (2009), since the former is beyond our purpose and the latter requires a nonlinear or hybrid model that can describe the shift from the Glucose to Galactose mode. The Glucose and Galactose models that we inferred in Section 4.6 have different structure and strengths of the interactions, which tells us that the system has two distinct dynamical modes; see Nordling *et al.* (2007b) for a discussion on dynamical modes in GRNs.

The extended IRMA model has 12 states and is schematically illustrated in Figures 6.1 and 6.2 for the Glucose and Galactose modes, respectively. The states are the log-fold change in total mRNA of each of the five genes of the network  $x_1 = CBF1$ ,  $x_2 = GAL4$ ,  $x_3 = SWI5$ , and  $x_4 = ASH1$ ,  $x_5 = GAL80$ , the log-fold change in the proteins-protein complexes  $x_6 = Gal4p$ ,  $x_7 = Gal80p$ ,  $x_8 = Gal4p-Gal80p$ ,  $x_9 = Cbf1p$ , and  $x_{10} = Swi5p-Unk$ , and the log-fold change of free and bound *ASH1* mRNA  $x_{11} = ASH1f$  and  $x_{12} = ASH1b$ . Cantone *et al.* perturbed and observed each of the five genes independently, so  $b_{ii} = c_{ii} = 1$  for  $i = 1, 2, \dots, 5$ , whereas the other coefficients are zero.

The interaction matrix  $\mathbf{A}$  is given in Table B.1 and values for the coefficients in Table B.2 for the two modes, together with a detailed description of how the example was constructed. Note that many of the interaction coefficients take different values in the Glucose and Galactose models and the structure of the two networks differ, as seen in Figures 6.1 and 6.2. The interaction matrix contains information about how the state variables regulate each other, *i.e.* the links between the states, and one would therefore in network inference like to determine it. We next investigate conditions for determining the interaction matrix  $\mathbf{A}$  from perturbation data, *i.e.* perturbations  $\mathbf{p}(t)$  and measured responses  $\mathbf{y}(t)$ .



**Figure 6.1: The extended IRMA network in Glucose.** The structure of the extended IRMA model for *S. cerevisiae* containing the gene regulatory network engineered by Cantone *et al.* (2009) grown in a medium that contains Glucose but does not contain Galactose. This network contains an activation of *CBF1* by *Gal4p* that does not exist in the Galactose network in Figure 6.2 and lacks the activation of *Gal4p-Gal80p* by *Gal80p* that is present in the Galactose network. An arrow shaped head on the link indicates up-regulation or activation, while a bar shaped head indicates down-regulation or repression. One possible set of strengths of the interactions that gives a model that can explain the Glucose steady-state data recorded by Cantone *et al.* is shown in Table B.2. This figure was generated in Cytoscape (Shannon *et al.*, 2003).



**Figure 6.2: The extended IRMA network in Galactose.** The structure of the extended IRMA model for *S. cerevisiae* containing the gene regulatory network engineered by Cantone *et al.* (2009) grown in medium that contains Galactose. This network contains an activation of Gal4p-Gal80p by Gal80p that does not exist in the Glucose network in Figure 6.1 and lacks the activation of CBF1 by Gal4p that is present in the Glucose network. An arrow shaped head on the link indicates up-regulation or activation, while a bar shaped head indicates down-regulation or repression. One possible set of strengths of the interactions that gives a model that can explain the Galactose steady-state data recorded by Cantone *et al.* is shown in Table B.2. This figure was generated in Cytoscape (Shannon *et al.*, 2003).

### 6.3 Conditions for inference of the full network

We first consider the problem of inferring the full interaction matrix  $\mathbf{A}$  in the case when only a subset of the states are perturbed and observed. We consider conditions on the set-up of the perturbation experiments that are necessary for reconstruction based on knowledge of the input-output behaviour of the system, which is a prerequisite for inference both from time-series and steady-state data. If two state-space representations (6.2.3) with structurally different  $\mathbf{A}$  matrices have identical responses to all perturbations, then they are indistinguishable based on data and the structure of  $\mathbf{A}$  cannot be inferred. We therefore start by in this section studying uniqueness of the state-space representation, which is a prerequisite for inference of the full network.

Since we consider noise-free measurements and allow for time-series data we can determine the exact relationship between the inputs (perturbations) and the outputs (responses). However, it is well known from systems theory that the input-output relationship in general does not uniquely define the state-space representation in (6.2.3), see *e.g.* Kailath (1980, p. 53-54, 80-90) or Trentelman et al. (2001, p. 39-43). This is partly due to the existence of similarity transformations of the system states

$$\mathbf{x}(t) = \mathbf{T}\tilde{\mathbf{x}}(t), \quad \text{with } \det \mathbf{T} \neq 0, \quad (6.3.1)$$

and partly due to existence of realizations, *i.e.* system representations  $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ , that are not observable or/and not controllable, loosely speaking, realizations containing states that do not have any influence on the output or are not influenced by the input. Since  $\mathbf{A}$  describes the interactions among the states, uniquely defined states are clearly a prerequisite for a unique state-space representation and inference of  $\mathbf{A}$ . We therefore first focus on uniqueness of the states and discuss the similarity transformation  $\mathbf{T}$ , and later comment on the controllability and observability issue.

The condition in (6.3.1) on the determinant of  $\mathbf{T}$ , which is an  $n \times n$  matrix, ensures that the transformation is invertible. If not fulfilled, then the same dynamical relations cannot in general be represented by both sets of states, so all alternative sets of states are obtained via the similarity transformation. With the transformation we get a new state basis for the state-space representation and we can rewrite (6.2.3) as

$$\frac{d\tilde{\mathbf{x}}}{dt}(t) = \underbrace{\mathbf{T}^{-1}\mathbf{A}\mathbf{T}}_{\triangleq \tilde{\mathbf{A}}}\tilde{\mathbf{x}}(t) + \underbrace{\mathbf{T}^{-1}\mathbf{B}}_{\triangleq \tilde{\mathbf{B}}}\mathbf{p}(t) \quad (6.3.2a)$$

$$\mathbf{y}(t) = \underbrace{\mathbf{C}\mathbf{T}}_{\triangleq \tilde{\mathbf{C}}}\tilde{\mathbf{x}}(t), \quad (6.3.2b)$$

with  $\tilde{\cdot}$  indicating the transformed state-space matrices. The structure of  $\tilde{\mathbf{A}}$  will in general differ from that of the original  $\mathbf{A}$  when  $\mathbf{T} \neq \mathbf{I}$ , *i.e.* when the new states differ from the original ones. Moreover, for any system that has as many distinct eigenvalues as states, the eigenvectors form a state transformation matrix that

diagonalizes the interaction matrix (Kailath, 1980, p. 54), *i.e.* completely decouples the network so that no interactions exist between the states whatsoever. As an example, consider the IRMA model that we inferred from the Glucose steady-state data of Cantone et al. (2009) in Section 4.6 with interaction matrix

$$\hat{\mathbf{A}}_{\text{Glu}} = \begin{bmatrix} -0.143 & 0.043 & 0.073 & 0 & 0 \\ 0 & -0.183 & 0 & 0 & 0 \\ 0.103 & 0.091 & -0.219 & 0 & 0 \\ 0.036 & 0 & 0 & -0.177 & 0 \\ 0 & 0 & 0.009 & 0 & -0.152 \end{bmatrix}, \quad (6.3.3)$$

and assume for the moment that the system only has five states, then the state-space representation obtained by the transformation matrix

$$\mathbf{T} = \begin{bmatrix} 0 & 0 & 0.474 & -0.751 & -0.162 \\ 0 & 0 & 0 & 0 & 0.177 \\ 0 & 0 & -0.861 & -0.583 & -0.016 \\ 0 & 1 & -0.173 & -0.298 & 0.971 \\ 1 & 0 & 0.063 & -0.08 & 0.005 \end{bmatrix} \quad (6.3.4)$$

has the diagonal interaction matrix

$$\mathbf{T}^{-1} \hat{\mathbf{A}}_{\text{Glu}} \mathbf{T} = \begin{bmatrix} -0.152 & 0 & 0 & 0 & 0 \\ 0 & -0.177 & 0 & 0 & 0 \\ 0 & 0 & -0.276 & 0 & 0 \\ 0 & 0 & 0 & -0.086 & 0 \\ 0 & 0 & 0 & 0 & -0.183 \end{bmatrix}. \quad (6.3.5)$$

Contrary to the eigenvalues or characteristic/natural frequencies of the system, which are conserved and visible in the diagonal matrix above, the structure of the interaction matrix is not invariant under similarity transformations. It is trivial to realize that in general transformations such that the interaction matrix becomes full also exist, *i.e.* that every state variable interact with all other states including itself. This exemplifies that uniquely defined states is a prerequisite for inference of the full interaction matrix  $\mathbf{A}$ . We next examine criteria to make the states unique.

The solution of the ODEs in the state-space model (6.2.3) from any initial state  $\mathbf{x}_0$  (see *e.g.* Trentelman et al., 2001, p. 38)

$$\mathbf{y}(t, \mathbf{x}_0) = \mathbf{C}e^{\mathbf{A}t} \mathbf{x}_0 + \int_0^t \underbrace{\mathbf{C}e^{\mathbf{A}(t-\tau)} \mathbf{B}}_{\triangleq \mathbf{K}(t)} \mathbf{p}(\tau) d\tau \quad (6.3.6)$$

contains the input-output relationship in its second term. By observing the output relative to the initial output, we can without loss of generality henceforth assume

$\mathbf{x}_0 = \mathbf{0}$ . The impulse response matrix  $\mathbf{K}(t)$  contains the relation between the  $\mathbf{A}$  matrix and inputs, as well as  $\mathbf{A}$  matrix and outputs. Conditions such that only the state transformation  $\mathbf{T} = \mathbf{I}$  fulfils

$$\mathbf{K}(t) = \underbrace{\mathbf{C}\mathbf{T}}_{=\tilde{\mathbf{C}}} e^{\mathbf{T}^{-1}\mathbf{A}\mathbf{T}t} \underbrace{\mathbf{T}^{-1}\mathbf{B}}_{=\tilde{\mathbf{B}}} \quad (6.3.7)$$

are needed for the states to be uniquely defined. We have restricted ourselves to the case when  $\mathbf{A} \in \mathbb{R}^{n \times n}$  is completely unknown, but  $\mathbf{C} \in \mathbb{R}^{o \times n}$  and  $\mathbf{B} \in \mathbb{R}^{n \times l}$  are known, so we next examine which properties they need to fulfil to make the states unique.

Because  $\mathbf{C}$  and  $\mathbf{B}$  are known, the corresponding transformed matrices  $\tilde{\mathbf{C}}$  and  $\tilde{\mathbf{B}}$  have to be equal to them in (6.3.7), and we obtain the system of equations

$$\begin{aligned} \mathbf{C}\mathbf{T} &= \mathbf{C} \\ \mathbf{T}^{-1}\mathbf{B} &= \mathbf{B} \end{aligned} \quad \Leftrightarrow \quad \begin{bmatrix} \mathbf{C} \\ \mathbf{B}^T \end{bmatrix} \mathbf{T} = \begin{bmatrix} \mathbf{C} \\ \mathbf{B}^T \end{bmatrix}. \quad (6.3.8)$$

Obviously more than one  $\mathbf{T}$  satisfies these equations unless the number of known rows in  $\mathbf{C}$  plus the number of known columns in  $\mathbf{B}$  at least equals the number of states  $n$ . This also implies that the sum of the number of different quantities that are perturbed and observed at least must equal the number of states. However, the relation between the rows of  $\mathbf{C}$  and columns of  $\mathbf{B}$  also matter. More precisely, this linear system of equations has a unique solution if and only if  $[\mathbf{C}^T \ \mathbf{B}]^T$  has full column rank (see *e.g.* Anton and Rorres, 2000; Friedberg et al., 2003; Horn and Johnson, 1990). Phrased differently, we need to know a linearly independent subset of  $n$  relations between the states and responses or perturbations. This condition is necessary for inference of the full network both in the case of time-series and steady-state data, and constitutes the main result in this section. Considering that each observed and perturbed quantity typically correspond to a single state, in practice for each state we need to know either how it is related to an observed response or how it is affected by an applied perturbation. Based on this discussion it trivially follows that the states are uniquely defined when  $\mathbf{C}$  has full column rank or  $\mathbf{B}$  has full row rank, but this is rarely met and we therefore later study inference of subnetworks.

The above condition for uniqueness of the states is clearly necessary for inference of  $\mathbf{A}$ , but in general not sufficient, since it does not ensure that the realization is unique, *i.e.* observable and controllable. Any unobservable or uncontrollable states cannot be seen in the input-output behaviour. A realization of (6.2.3) is observable

if and only if the observability matrix

$$\mathbf{O}_{(C,A)} \triangleq \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{n-1} \end{bmatrix} \quad (6.3.9)$$

has rank  $n$  and controllable if and only if the controllability matrix

$$\mathbf{S}_{(A,B)} \triangleq \begin{bmatrix} B & AB & A^2B & \dots & A^{n-1}B \end{bmatrix} \quad (6.3.10)$$

has rank  $n$ , see *e.g.* [Kailath \(1980, p. 80-90\)](#) or [Trentelman et al. \(2001, p. 40-42\)](#). Both observability and controllability depend on the interaction matrix  $\mathbf{A}$  of the “true” system, and it is required to determine precise conditions for uniqueness of the realization. Observability and controllability can thus in general only be guaranteed in some exceptional cases, such as full column rank of  $\mathbf{C}$  and full row rank of  $\mathbf{B}$ .

An example of inference of the full network from time-series data, when all states are observed but only one state is perturbed, is given in [Schmidt et al. \(2005\)](#). They prove that the full network can be inferred, under stated conditions, when the system is controllable. Observability and uniqueness of the states is in their case guaranteed by observing all states, so the example illustrates the conditions that we have pointed out here. Interestingly, the necessary conditions for dynamical structure reconstruction of LTI systems established by [Goncalves and Warnick \(2008, Theorem 2\)](#), which we describe later, are not fulfilled for the example in Schmidt *et al.*. It thus proves by contradiction that the conditions of Goncalves and Warnick are not necessary, which we will return to later.

In conclusion, observability and controllability together with knowledge of a linearly independent subset of  $n$  relations between the states and responses or perturbations are necessary for inference of the full network. In fact, they are also sufficient because they ensure that the realization, state definition, and interaction matrix are unique. We next consider the problem of inferring the interactions among a subset of the states, *i.e.* subnetwork inference.

## 6.4 Conditions for inference of a subnetwork

When inferring GRNs it is common practice to only observe the change in mRNA abundance, neglecting all protein, metabolite, and other state variables of the system, and perturb selected genes or/and quantities with unknown effect on the observed genes ([Hecker et al., 2009](#); [Tegnér and Björkegren, 2007](#); [Brazhnik et al., 2002](#)). In this case we say that the system contains latent states, characterised by all elements in some column of  $\mathbf{C}$  and the corresponding row of  $\mathbf{B}$  being zero, because a latent state per definition is not directly observed nor perturbed. The aim is now

to determine necessary conditions for inferring the interactions among a subset of the states, which we call the states of interest, possibly including the indirect interactions mediated by the latent states. Formally we call any network that does not contain all state variables of the system as nodes a subnetwork. We here use the phrase latent states and not hidden states, as the latter commonly refers to a basis of the unobservable or uncontrollable subspace, *i.e.* the set of states that do not impact the input-output behaviour of the system (see *e.g.* Skogestad and Postlethwaite, 1996, p. 137).

Assume for now that only a subset of the state variables are either directly observed or perturbed, divide the vector of states into two parts,  $\mathbf{x} = [\mathbf{x}_1^T, \mathbf{x}_2^T]^T$ , such that  $\mathbf{x}_1$  contains the  $n_1$  states of interest that either are directly observed or perturbed and  $\mathbf{x}_2$  the remaining latent states. To see the impact of the latent states we rewrite (6.2.3) so that this division becomes explicit

$$\frac{d\mathbf{x}_1}{dt}(t) = \mathbf{A}_{11}\mathbf{x}_1(t) + \mathbf{A}_{12}\mathbf{x}_2(t) + \mathbf{B}_1\mathbf{p}(t) \quad (6.4.1a)$$

$$\frac{d\mathbf{x}_2}{dt}(t) = \mathbf{A}_{21}\mathbf{x}_1(t) + \mathbf{A}_{22}\mathbf{x}_2(t) \quad (6.4.1b)$$

$$\mathbf{y}(t) = \mathbf{C}_1\mathbf{x}_1(t). \quad (6.4.1c)$$

As previously, we assume that  $\mathbf{C}_1$  and  $\mathbf{B}_1$  are known with dimension  $o \times n_1$  and  $n_1 \times l$ , respectively. The interaction matrix of the subnetwork that is sought from the input-output relationship of the system consists of both the direct interactions in  $\mathbf{A}_{11}$  and the indirect interactions through the latent states. It can be expressed concisely using transfer functions by taking the Laplace transform of (6.4.1),

$$\mathbf{s}\mathbf{x}_1(s) = \underbrace{\left( \mathbf{A}_{11} + \overbrace{\mathbf{A}_{12}(s\mathbf{I} - \mathbf{A}_{22})^{-1}\mathbf{A}_{21}}^{\triangleq \mathbf{H}(s)} \right)}_{\triangleq \mathbf{A}(s)} \mathbf{x}_1(s) + \mathbf{B}_1\mathbf{p}(s) \quad (6.4.2a)$$

$$\mathbf{y}(s) = \mathbf{C}_1\mathbf{x}_1(s), \quad (6.4.2b)$$

assuming that all initial values are zero. Here  $\mathbf{A}(s)$  is the sought interaction matrix of the subnetwork and  $\mathbf{H}(s)$  contains the indirect interactions. Instead of being described in the time domain, the interactions are now described in the Laplace or frequency domain—see Section B.2 for an introduction to Laplace transforms.

Based on the results in the previous section, we know that the latent states  $\mathbf{x}_2$  cannot be uniquely defined given only the input-output behaviour. We therefore start by establishing necessary conditions on the set-up of the perturbation experiments for unique definition of the states of interest  $\mathbf{x}_1$ . Instead of (6.3.8), we obtain, based on knowledge of  $\mathbf{C}_1$  and  $\mathbf{B}_1$  and the same reasoning, a system of equations that constrains the possible similarity transformations, partitioned like  $\mathbf{A}$  above,

$$\begin{bmatrix} \mathbf{C}_1 & \mathbf{0} \\ \mathbf{B}_1^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{T}_{11} & \mathbf{T}_{12} \\ \mathbf{T}_{21} & \mathbf{T}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{C}_1 & \mathbf{0} \\ \mathbf{B}_1^T & \mathbf{0} \end{bmatrix}. \quad (6.4.3)$$

This gives the following constraints on the two blocks of the states of interest,  $\mathbf{T}_{11}$  and  $\mathbf{T}_{12}$ , which are required to be  $\mathbf{I}$  and  $\mathbf{0}$ , respectively,

$$\begin{bmatrix} \mathbf{C}_1 \\ \mathbf{B}_1^T \end{bmatrix} \mathbf{T}_{11} = \begin{bmatrix} \mathbf{C}_1 \\ \mathbf{B}_1^T \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \mathbf{C}_1 \\ \mathbf{B}_1^T \end{bmatrix} \mathbf{T}_{12} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}. \quad (6.4.4)$$

The states of interest are thus uniquely defined if and only if  $[\mathbf{C}_1^T \ \mathbf{B}_1]^T$  has full column rank. Consequently, a necessary condition for inference of a subnetwork is knowledge of a linearly independent subset of  $n_1$  relations between the states of interest and responses or perturbations. This is essentially the same condition as for unique definition of all states in Section 6.3, but limited to the states of interest  $\mathbf{x}_1$ . The condition ensures that only state transformations of the form

$$\begin{bmatrix} \mathbf{x}_1(t) \\ \mathbf{x}_2(t) \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{T}_{21} & \mathbf{T}_{22} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{x}}_1(t) \\ \tilde{\mathbf{x}}_2(t) \end{bmatrix}, \quad (6.4.5)$$

with  $\mathbf{x}_2(t) = \mathbf{T}_{21}\tilde{\mathbf{x}}_1(t) + \mathbf{T}_{22}\tilde{\mathbf{x}}_2(t)$  exist. This can be used to rewrite (6.4.1) in a manner similar to (6.3.2), proving that the latent states are not unique. Based on the results of the previous section it is therefore not feasible to infer the interactions among the latent states nor can they be uniquely associated to any physical quantity.

From the previous section we know that observability and controllability also are needed to obtain a unique realization from input-output behaviour alone. The exact conditions on  $\mathbf{C}_1$  and  $\mathbf{B}_1$  for observability and controllability depends on the interaction matrix of the “true” system, including the latent states. However, we are only interested in the subnetwork formed by the states of interest, so we only require uniqueness of the related part of any realization that can represent the input-output behaviour. The input-output relation is unaffected by hidden states, consequently existence of a minimal realization with the states of interest as a subset of its states is sufficient for uniqueness of the subnetwork. Actually, existence of hidden latent states in the “true” system is an advantage because it reduces the number of latent states that affects the input-output relation and the length of the paths between the states of interest. In other words, in agreement with the principle of parsimony we seek the minimal realization of the system, *i.e.* the realization in which  $\mathbf{A}$  has the smallest possible dimension and number of states (see *e.g.* Skogestad and Postlethwaite, 1996, p. 137). Other realizations in agreement with the input-output behaviour always exist, but they are not minimal and contain hidden states that do not affect the input-output behaviour. Note that a minimal realization is observable and controllable per definition. Existence of this minimal realization together with the above condition for a unique definition of the states of interest are then necessary and sufficient for inference of a subnetwork.

In conclusion, observability and controllability of the states of interest together with knowledge of a linearly independent subset of  $n_1$  relations between the states of interest and responses or perturbations are necessary for inference of a subnetwork. These conditions are also sufficient for inference of a unique minimal realization of

the system. The required knowledge on the set-up of the perturbation experiments, in other words, scales linearly with the number of states of interest, which implies that inference of a subnetwork typically is feasible. We next relate these conditions to previously published conditions by [Goncalves and Warnick \(2008\)](#).

## 6.5 Relation to previously published conditions

Our work is closely related to the work in [Goncalves and Warnick \(2008\)](#) on necessary and sufficient conditions for dynamical structure reconstruction of LTI systems when unobserved states exist, although we do not use dynamical structure functions. We therefore repeat and re-examine some of their results in order to explain the apparent discrepancy noted at the end of Section 6.3. The aim is still to infer the interactions among a subset of the states.

We now study the transfer matrix in a manner similar to [Goncalves and Warnick \(2008\)](#). The input-output relationship of the system in (6.2.3) can be described by the transfer matrix

$$\mathbf{G}(s) = \mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} \quad (6.5.1)$$

when the initial state  $\mathbf{x}_0 = \mathbf{0}$  (see *e.g.* [Trentelman et al., 2001](#), p. 38). It is the Laplace transform of the impulse response—see Section B.2 for an introduction to Laplace transforms. This transfer matrix thus contains the same information about the input-output relationship as (6.3.6), but the convolution has now been simplified to an algebraic equation. Similar to [Goncalves and Warnick \(2008\)](#) we now distinguish between observed states  $\hat{\mathbf{x}}_1$  and unobserved states  $\hat{\mathbf{x}}_2$ , and assume that  $\mathbf{C}$  is known. Note that the observed states are a subset of the states of interest  $\mathbf{x}_1$  and the latent states  $\mathbf{x}_2$  are a subset of the unobserved states. The Laplace transformed analogue of (6.2.3) is in this case

$$s\hat{\mathbf{x}}_1(s) = \underbrace{\left( \hat{\mathbf{A}}_{11} + \overbrace{\hat{\mathbf{A}}_{12}(s\mathbf{I} - \hat{\mathbf{A}}_{22})^{-1}\hat{\mathbf{A}}_{21}}^{\triangleq \hat{\mathbf{H}}(s)} \right)}_{\triangleq \hat{\mathbf{A}}(s)} \hat{\mathbf{x}}_1(s) + \underbrace{\left( \hat{\mathbf{B}}_1 + \hat{\mathbf{A}}_{12}(s\mathbf{I} - \hat{\mathbf{A}}_{22})^{-1}\hat{\mathbf{B}}_2 \right)}_{\triangleq \hat{\mathbf{B}}(s)} \mathbf{p}(s) \quad (6.5.2a)$$

$$\mathbf{y}(s) = \hat{\mathbf{C}}_1 \hat{\mathbf{x}}_1(s), \quad (6.5.2b)$$

assuming that all initial values are zero. Here  $\hat{\mathbf{B}}_1$  and  $\hat{\mathbf{B}}_2$  contain the direct effect of the perturbations on the observed and unobserved states, respectively. If and only if  $\hat{\mathbf{B}}_2 = \mathbf{0}$ , then the set of latent states equals the set of unobserved states, *i.e.*  $\hat{\mathbf{x}}_1 = \mathbf{T}_1\mathbf{x}_1$  and  $\hat{\mathbf{x}}_2 = \mathbf{T}_2\mathbf{x}_2$  for some similarity transforms  $\mathbf{T}_1$  and  $\mathbf{T}_2$ . The interactions among the observed states are described by the transfer functions in  $\hat{\mathbf{A}}(s)$ , which are the sum of the direct interactions in  $\hat{\mathbf{A}}_{11}$  and the indirect interactions

mediated through the unobserved states in  $\dot{\mathbf{H}}(s)$ . The input matrix  $\dot{\mathbf{B}}(s)$  is a sum of the direct effect of the perturbations  $\dot{\mathbf{B}}_1$ , and the indirect effect that acts through the unobserved states  $\dot{\mathbf{A}}_{12}(s\mathbf{I} - \dot{\mathbf{A}}_{22})^{-1}\dot{\mathbf{B}}_2$ . The transfer matrix is in this case

$$\mathbf{G}(s) = \dot{\mathbf{C}}_1(s\mathbf{I} - \dot{\mathbf{A}}(s))^{-1}\dot{\mathbf{B}}(s). \quad (6.5.3)$$

Similar to the case in Section 6.4, we seek necessary conditions for uniqueness of  $\dot{\mathbf{A}}(s)$  based on knowledge of the transfer matrix  $\mathbf{G}(s)$ .

We start by an example demonstrating that independent observations of all states in the subnetwork and knowledge of  $\mathbf{G}(s)$  are not sufficient to determine the interaction matrix  $\dot{\mathbf{A}}(s)$ . Consider the extended IRMA model for the Glucose mode introduced above, and assume that only the first five states corresponding to the five gene mRNAs are perturbed and observed independently so that

$$s\dot{\mathbf{x}}_1(s) = \begin{bmatrix} a_{11}(s) & a_{12}(s) & a_{13}(s) & 0 & 0 \\ a_{21}(s) & a_{22} & 0 & 0 & 0 \\ a_{31}(s) & a_{32}(s) & a_{33} & 0 & 0 \\ a_{41}(s) & 0 & a_{43}(s) & a_{44} & 0 \\ 0 & 0 & a_{53}(s) & 0 & a_{55} \end{bmatrix} \dot{\mathbf{x}}_1(s) + \mathbf{I}\mathbf{p}(s) \quad (6.5.4a)$$

$$\mathbf{y}(s) = \mathbf{I}\dot{\mathbf{x}}_1(s) \quad (6.5.4b)$$

corresponds to (6.5.2), with the transfer functions  $a_{ij}(s)$  given in Section B.1.3. Assume for a moment that we do not know the input matrix  $\mathbf{B}$ , *i.e.* we do not know the direct effect of the perturbations, and add one additional non-zero element to it, namely  $b_{75} = 1$ . This corresponds to the perturbation applied to the *GAL80* gene ( $x_5$ ) also having a direct effect on the protein Gal80p ( $x_7$ ). Then, the system given by

$$s\dot{\mathbf{x}}_1(s) = \begin{bmatrix} a_{11}(s) & a_{12}(s) & a_{13}(s) & 0 & 0 \\ a_{21}(s) & a_{22} & a_{23}(s) & 0 & 1 \\ a_{31}(s) & a_{32}(s) & a_{33} & 0 & 0 \\ a_{41}(s) & 0 & a_{43}(s) & a_{44} & 0 \\ 0 & 0 & a_{53}(s) & 0 & a_{55} \end{bmatrix} \dot{\mathbf{x}}_1(s) \quad (6.5.5a)$$

$$+ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & b_{25}(s) \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \mathbf{p}(s) \quad (6.5.5a)$$

$$\mathbf{y}(s) = \mathbf{I}\dot{\mathbf{x}}_1(s), \quad (6.5.5b)$$

with

$$a_{23}(s) = \frac{a_{27}a_{73}(s - a_{10,10}) + a_{27}a_{7,10}a_{10,3}}{(s - a_{77})(s - a_{10,10})}, \quad (6.5.6)$$

$$b_{25}(s) = -\frac{1}{s - a_{55}}. \quad (6.5.7)$$

has exactly the same input-output behaviour for the five genes as our extended IRMA model for Glucose. In other words, the transfer matrices  $\mathbf{G}(s)$  are identical. This network is illustrated in Figure 6.3 and was obtained by making the following modifications of the original full interaction matrix;  $a_{25} = 1$ ,  $a_{27} = -1$ ,  $a_{73} = a_{53}$ ,  $a_{75} = 0$ ,  $a_{77} = a_{55}$ ,  $a_{78} = 0$ ,  $a_{7,10} = a_{5,10}$ , and  $a_{87} = 0$ . Note that, in this example, the resulting  $\dot{\mathbf{A}}(s)$  and  $\dot{\mathbf{B}}(s)$  only have a total of three elements that differ from their counterparts in (6.5.4a), because we aimed at keeping it simple. In general, there exists an infinite number of  $\dot{\mathbf{A}}(s)$  and  $\dot{\mathbf{B}}(s)$  that are consistent with the same  $\mathbf{G}(s)$ , because both  $\dot{\mathbf{A}}(s)$  and  $\dot{\mathbf{B}}(s)$  in (6.5.2) contain  $\dot{\mathbf{A}}_{12}(s\mathbf{I} - \dot{\mathbf{A}}_{22})^{-1}$  and a trade-off can be made between them. To avoid this trade-off it is necessary to plan the experiments such that  $\dot{\mathbf{B}}_2$  is zero, *i.e.* that the perturbations do not affect any unobserved state, or such that  $\dot{\mathbf{B}}_2$  is known and the states of interest  $\mathbf{x}_1$  are uniquely defined. Conditions for the latter case were established in Section 6.4 and we next follow the approach of Goncalves and Warnick (2008) to establish conditions for the former case.

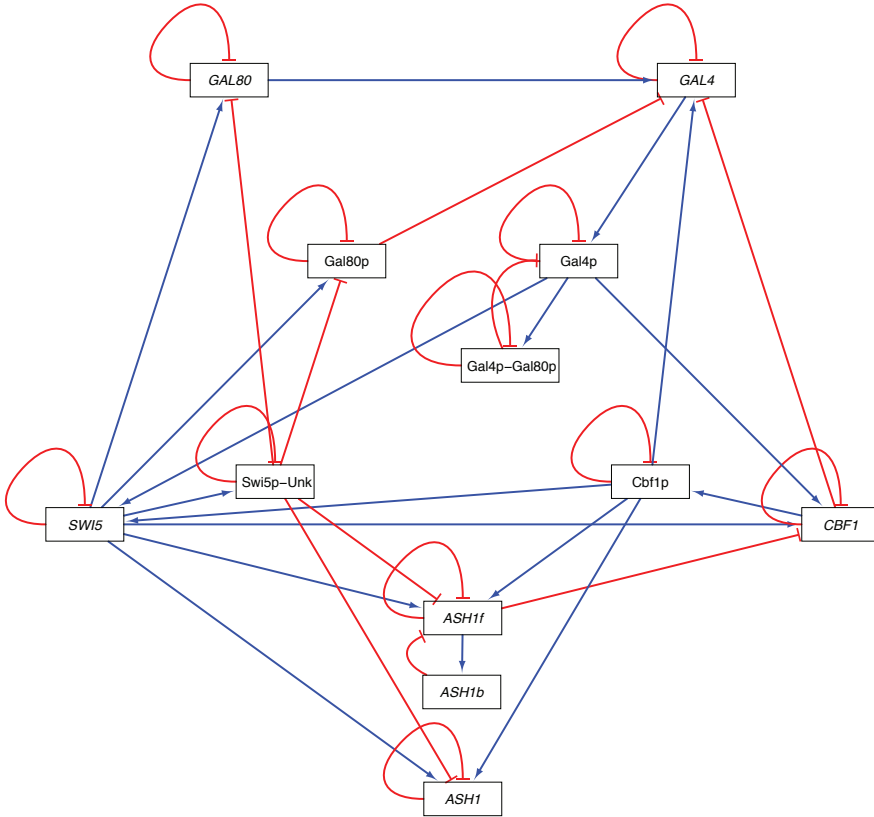
We now study conditions for existence of a unique pair  $\dot{\mathbf{A}}(s)$  and  $\dot{\mathbf{B}}(s)$  fulfilling (6.5.3) based on knowledge of  $\mathbf{G}(s)$  under the assumption that  $\dot{\mathbf{C}}_1$  is known and invertible. From (6.5.2) and  $\dot{\mathbf{x}}_1 = \dot{\mathbf{C}}_1^{-1}\mathbf{G}(s)\mathbf{p}$  with introduction of  $\tilde{\mathbf{G}} \triangleq \dot{\mathbf{C}}_1^{-1}\mathbf{G}(s)$ , we derive

$$s\tilde{\mathbf{G}}(s) = \dot{\mathbf{A}}(s)\tilde{\mathbf{G}}(s) + \dot{\mathbf{B}}(s), \quad (6.5.8)$$

which can be put on vector form

$$s\tilde{\mathbf{G}}^H(s) = [\tilde{\mathbf{G}}^H(s) \mathbf{I}] \begin{bmatrix} \dot{\mathbf{A}}^H(s) \\ \dot{\mathbf{B}}^H(s) \end{bmatrix}. \quad (6.5.9)$$

Here  $^H$  denotes the conjugate transpose. Any combination of  $\dot{\mathbf{A}}(s)$  and  $\dot{\mathbf{B}}(s)$  that fulfils this equation fully describes the input-output behaviour given by  $\mathbf{G}(s)$ . In general, infinitely many such combinations exist and we need enough knowledge about  $\dot{\mathbf{B}}(s)$  such that only one unique solution  $\dot{\mathbf{A}}(s)$  exists to the inference problem, corresponding to the “true” interaction matrix of the subnetwork of observed states. This is the analogue of the relationship used by Goncalves and Warnick (2008, equation 7) to establish necessary and sufficient conditions for reconstruction of the dynamical structure functions. The dynamical structure functions are an alternative representation of the system in which all diagonal elements of the matrix corresponding to  $\dot{\mathbf{A}}(s)$  is zero and thus has one less parameter per row. The



**Figure 6.3: The modified extended IRMA network in Glucose.** The structure of the extended IRMA model for *S. cerevisiae* modified such that the input-output behaviour of the system is identical to the extended IRMA model for Glucose in Figure 6.1 despite differences in the interaction matrix and input matrix. An arrow shaped head on the link indicates up-regulation or activation, while a bar shaped head indicates down-regulation or repression. This figure was generated in Cytoscape (Shannon et al., 2003).

equivalence of their necessary and sufficient condition for reconstruction (Theorem 2) is therefore as many known elements per column in  $[\dot{\mathbf{A}}(s) \dot{\mathbf{B}}(s)]^H$  as nodes in the subnetwork. In other words,  $\dot{\mathbf{B}}^H(s)$  must at least be of dimension  $n_1 \times n_1$  and completely known unless *a priori* information on  $\dot{\mathbf{A}}(s)$  is available. This corresponds to requiring at least as many unique experiments as there are nodes in the subnetwork, and furthermore with known effects of the perturbations on the nodes. This condition is necessary and sufficient to ensure that the null space of the operator  $[\tilde{\mathbf{G}}^H(s) \mathbf{I}]$  is empty.

The condition above is not fulfilled in the case considered by [Schmidt et al. \(2005\)](#), because they only perform a single experiment to infer a network with four nodes and  $\mathbf{B}$  is unknown and of dimension  $n_1 \times 1$ . In addition, Schmidt *et al.* showed that it is not necessary to know  $\mathbf{B}$ , because it could be inferred from the response data. Indeed, the example considered in Schmidt *et al.* fulfils the conditions derived in Section 6.3 and the full network can be inferred from a single time-series experiment since all states are observed and the system is controllable. The example considered by Schmidt *et al.* does not involve any unobserved states and we therefore below consider a small example with one latent state before explaining why the conditions in [Goncalves and Warnick \(2008\)](#) are contradictory to the results in Schmidt *et al.* and the conditions derived in Section 6.3.

Consider the following 3 gene system

$$\frac{d\mathbf{x}}{dt}(t) = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 1 \\ 1 & 0 & -1 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \mathbf{p}(t) \quad (6.5.10a)$$

$$\mathbf{y}(t) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \mathbf{x}(t), \quad (6.5.10b)$$

with two observed states of interest  $x_1$  and  $x_2$ . The “true” interaction matrix of the subnetwork that they form and the corresponding input matrix are

$$\mathbf{A}(s) = \begin{bmatrix} -1 & 0 \\ \frac{1}{s+1} & -1 \end{bmatrix} \quad \mathbf{B}(s) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \quad (6.5.11)$$

Assuming that only  $\mathbf{C}$  and the transfer matrix

$$\mathbf{G}(s) = \begin{bmatrix} \frac{1}{s+1} \\ \frac{1}{(s+1)^3} \end{bmatrix} \quad (6.5.12)$$

are known, our aim is to infer the interaction matrix of the subnetwork. Knowledge of  $\mathbf{C}$  alone implies that the two states of interest are uniquely defined. The realization in (6.5.10) is also observable and controllable, so the necessary and sufficient conditions for inference of a subnetwork established in Section 6.4 are fulfilled. Nonetheless, the operator  $[\mathbf{G}^H(s) \mathbf{I}]$  has a null space of dimension two and (6.5.9) is satisfied by every interaction matrix  $\mathbf{A}(s)$  and input matrix  $\mathbf{B}(s)$  of the form

$$\mathbf{A}(s) = (1 + k_3(s)) \begin{bmatrix} -1 & 0 \\ 0 & s \end{bmatrix} + \begin{bmatrix} k_1(s) \\ k_2(s) \end{bmatrix} \begin{bmatrix} -1 & (s+1)^2 \end{bmatrix} \quad (6.5.13a)$$

$$\mathbf{B}(s) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} + k_3(s) \begin{bmatrix} \frac{1}{s+1} \\ -\frac{s}{(s+1)^3} \end{bmatrix}, \quad (6.5.13b)$$

with  $k_i(s)$  being arbitrary transfer functions. As noted by [Goncalves and Warnick \(2008\)](#), the existence of a non-empty nullspace implies that there exists an infinite number of pairs  $\mathbf{A}(s), \mathbf{B}(s)$  that give the same input-output behaviour  $\mathbf{G}(s)$ . However, as we note here, the feasible pairs  $\mathbf{A}(s), \mathbf{B}(s)$  have to fulfil constraints imposed by properness and maximum order of the “true” system. In other words, all elements in  $\mathbf{A}(s)$  and  $\mathbf{B}(s)$  must be proper, *i.e.* have numerator degree less than or equal to denominator degree, and the overall order should not exceed the order of the known  $\mathbf{G}(s)$ . To see how this constrains the use of the nullspace for the example considered here, we consider allowable choices of the polynomials  $k_i(s)$ ,  $i = 1, \dots, 3$  in [Section 6.5.13](#).

The order of the system, corresponding to the McMillan degree of  $\mathbf{G}(s)$ , is three in this case (see *e.g.* [Skogestad and Postlethwaite, 1996](#), p. 137). This restricts the order of each element of  $\mathbf{A}(s)$  and  $\mathbf{B}(s)$  to one since two states are observed. Considering first  $\mathbf{B}(s)$ , we see that  $k_3(s)$  must be 0 if we require  $b_{21}(s)$  to have order one or less and  $b_{11}(s)$  to be proper. We then get  $a_{11}(s) = -1 - k_1(s)$  and  $a_{12}(s) = k_1(s)(s+1)^2$  from which we deduce that  $k_1(s)$  can be of at most order one for  $a_{11}(s)$  to have order one or less, while  $k_1(s)$  must be zero or of at least order two to make  $a_{12}(s)$  proper. Thus,  $k_1(s) = 0$  is the only feasible choice. Finally, we have  $a_{21}(s) = -k_2(s)$  and  $a_{22}(s) = s + k_2(s)(s+1)^2$ . Thus,  $k_2(s)$  can be of at most order one and must furthermore be chosen such that  $a_{22}(s)$  is proper. This yields a unique  $k_2(s) = -1/(s+1)$  and we have arrived at a unique set of interaction and input matrices that are identical to the “true” ones in [\(6.5.11\)](#). Thus, in this case a unique minimal realization exists since the null space of the operator  $[\mathbf{G}^H(s) \mathbf{I}]$  cannot be utilized due to constraints on properness and order of the elements of  $\mathbf{A}(s)$  and  $\mathbf{B}(s)$ .

The simple case considered here serves to explain why [Schmidt et al. \(2005\)](#) was able to infer a network with four nodes in a single experiment, despite the claim by [Goncalves and Warnick \(2008\)](#) that at least as many unique experiments as there are nodes in the network are required for unique recovery of  $\mathbf{A}(s)$  and  $\mathbf{B}(s)$ . [Goncalves et al.](#) failed to note that the nullspace of  $[\mathbf{G}^H(s) \mathbf{I}]$  in general can not be fully utilized, and their condition is therefore not necessary but only sufficient.

## 6.6 Direct versus indirect interactions

As seen above, the existence of latent states affects the requirements on the experimental setup and the data recorded for network inference. The presence of latent states also affects the meaning of an inferred interaction. Biologists usually desire direct causal interactions in accordance with the tradition in molecular biology to map out causal interactions between molecules of the type: “Gal80p binds directly to the activation domain of Gal4p” ([Sellick et al., 2008](#)). Although the interactions inferred from perturbation experiments are causal—thanks to the distinction between perturbations and responses—they are in general not direct. Latent states namely mediate influences from one observed state to another. In [\(6.4.2\)](#) the matrix  $\mathbf{A}_{11}$

contains the direct interactions between the states of interest and  $\mathbf{H}(s)$  the indirect interactions mediated by the latent states. From a systems perspective  $\mathbf{H}(s)$  captures the influence of the environment, which in general is equally important for regulation and function as the direct interactions.

It is often hard to say if an inferred interaction is direct or indirect in a subnetwork model. From steady-state experiments alone this is impossible to establish, as seen if we take the limit of (6.4.2) as the frequency approaches zero and replace the states of interest by observations from  $m$  experiments  $\mathbf{Y}(0) \triangleq [\mathbf{y}_1(0) \dots \mathbf{y}_m(0)]$

$$\mathbf{Y}(0) = \mathbf{C}_1 \underbrace{(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})}_{=\mathbf{A}(0)}^{-1} \mathbf{B}_1 \mathbf{P}(0), \quad (6.6.1)$$

with the step perturbations  $\mathbf{P}(0) \triangleq [\mathbf{p}_1(0) \dots \mathbf{p}_m(0)]$ . It is then clearly impossible to distinguish contributions from the indirect interactions  $\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}$  and the direct interactions  $\mathbf{A}_{11}$ . Thus, elements of the static interaction matrix  $\mathbf{A}(0)$  correspond to direct or indirect interactions, or combinations thereof, and there is no possibility to separate them based on steady-state observations. Moreover, both  $\mathbf{C}_1$  and  $\mathbf{B}_1$  must have rank  $n_1$  and be known for existence of a unique solution of  $\mathbf{A}(0)$ . This is consistent with the results in Chapter 4—in the sense that we need at least as many experiments as there are nodes in the network for inference. The good news is that we can focus on a subnetwork and thereby manage with less experiments. The bad news is that we in the steady-state case cannot distinguish between direct and indirect influences, no matter how many experiments we perform, unless we observe all states.

From the input-output relationship that can be obtained from time-series experiments, it is possible under the conditions established in Section 6.4 to estimate the transfer function matrix  $\mathbf{A}(s)$  in (6.4.2). In this case it is possible to distinguish direct from indirect interactions, and also determine the maximum and minimum path lengths, in terms of the number of latent states, in the indirect interactions. To see this, consider element  $a_{ij}(s)$  of  $\mathbf{A}(s)$  which describes the effect of changes in gene  $j$  on gene  $i$

$$a_{ij}(s) = \frac{N(s)}{D(s)}. \quad (6.6.2)$$

Assume there are no cancellations between the numerator  $N(s)$  and denominator  $D(s)$  polynomials. Then the order of the denominator polynomial  $D(s)$  gives an upper bound on the number of latent states involved in the longest path(s) from gene  $j$  to gene  $i$ . Similarly, the relative degree, corresponding to the difference between the order of the denominator  $D(s)$  and the order of the numerator  $N(s)$ , gives an upper bound on the number of latent states in the shortest path from gene  $j$  to gene  $i$ . Thus, a direct influence from gene  $j$  to gene  $i$  exists if the relative degree of  $a_{ij}(s)$  is zero. If the order of the denominator  $D(s)$  is non-zero and the relative degree is zero, then it implies that the interaction is a combination of direct and indirect influences.

As an example, consider the transfer function from *SWI5* to *GAL80* in our extended IRMA model for the Glucose mode

$$a_{53}(s) = a_{53} + \frac{a_{5,10}a_{10,3}}{s - a_{10,10}}. \quad (6.6.3)$$

As can be seen in Figure 6.1, the influence from gene 3 to gene 5 consists of a direct effect  $a_{53}$  in combination with an indirect effect through the feedforward loop that involves the latent state *Swi5p-Unk* ( $x_{10}$ ). Correspondingly, the transfer-function  $a_{53}(s)$  is first order with relative degree zero.

The ability to estimate the order and relative order of each individual transfer function in  $\mathbf{A}(s)$  depends on how informative the time-series data is and how well the experiments excite the latent states. However, as shown above, this information is required to distinguish between direct and indirect interactions. Considering that most interactions inferred in the literature are not based on time-series data and that the presence of latent states must be assumed in almost every biological system of interest, it is likely that inferred interactions in general do not correspond to physical binding of the states, *e.g.* binding of transcription factors. This is also pointed out in [Gardner and Faith \(2005\)](#), who therefore propose the phrase regulatory influence for inferred network interactions. To stress that these influences may be mediated by latent states but never by any observed state we say that they are pseudo or quasi direct.

## 6.7 Discussion and conclusions

The purpose of network inference is primarily to find the interactions that exist among a set of variables of interest, *i.e.* the  $n_1$  states of interest. To be able to do this based on data from perturbation experiments it is necessary to know a linearly independent subset of  $n_1$  relations between the states of interest and responses or perturbations. Otherwise the states of interest are not uniquely defined. In addition, a minimal realization with the states of interest need to exist, which is observable and controllable per definition. These requirements are also necessary and sufficient for inference of the full network when the set of states of interest contains all states of the system, implying that the fundamental information requirements scale linearly with the number of states of interest. Given that the state variables of interest typically represent molecules of some sort, *e.g.* mRNA, proteins, metabolites, or ions, which often can be directly observed and/or perturbed. Fulfilment of these necessary conditions can therefore typically be ensured during the planing of the experiments.

Direct observation of changes in biological experiments together with knowledge of the applied perturbations make it possible to assign causality to the interactions, *i.e.* determine the direction of the links. Each perturbation is per definition the cause and the resulting change of the states is the effect mediated by the directed interactions. *E.g.* over-expression of *GAL4* leads to increased expression of *SWI5* in the IRMA example depicted in Figure 6.1. The direction is in general hard

to determine without doing perturbation experiments, *e.g.* if we did not know the perturbations, then we could at best infer correlations between the states of interest. To highlight the value of knowing the perturbations and response in the states of interest consider the common case faced in econometrics. In analysis of econometric data it is typically not known which variables are perturbations and which are responses of the observed ones and the notion of Granger causality has been introduced (Granger, 1969). In short, a variable is said to Granger cause another variable if the prediction of the observed values of the later is improved by use of the observed values of the former. A variables influence on the prediction of another variable however depends on how it relates to the set of other variables used in the prediction (Lukacs et al., 2009; Guyon and Elisseeff, 2003; Freedman, 1983). Cases when an unrelated variable gives better prediction than a related variable have been documented. The possible existence of feedback loops makes it in general impossible to say that one variable causes the response seen in another. Granger causality should therefore not be mixed with causality despite the name. The directions of the interactions are essential to understand the mechanism behind biological functions and we therefore recommend inference based on data from perturbation experiments. Even though the perturbations are needed to assign causality to inferred interactions, they have in many cases been neglected, as evident by *e.g.* counting the number of works in which the network is inferred from expression changes alone in recent reviews such as De Smet and Marchal (2010); He et al. (2009); Hecker et al. (2009); Karlebach and Shamir (2008); Li et al. (2008b); Bonneau (2008); Tegnér and Björkegren (2007); Markowitz and Spang (2007); Cho et al. (2007).

Gardner and Faith (2005) classify inferred networks into two groups: physical and influential models. Due to indirect influences mediated by latent states only an influential model is in general obtained from perturbation experiments, meaning that the interactions describe regulatory influences between the states. Only direct interactions can correspond to physical binding between molecules. Since the interactions in general are not physical, it is strictly speaking wrong to evaluate or validate inferred networks based on physical interactions obtained by other techniques. This could in part explain the discrepancy seen between inferred networks and interactions found in chromatin immunoprecipitation experiments *e.g.* in Lorenz et al. (2009). However, to correct, modify, and create a specific behaviour we need to know the mechanisms that generate the behaviour, *i.e.* the regulatory influences. We therefore argue that it is regulatory interactions that are of primary interest in both medicine and biotechnology. Actually network inference based on perturbation experiments is the only method that captures the interactions of regulatory importance, as far as we know. Perturbation of the network creates the type of signals that environmental changes give rise to and the system responds in the natural way. The interactions essential for the behaviour are activated and it is these regulatory influences that are inferred. Indirect interactions through latent states also capture the effect of the environment on the system consisting of the states of interest. The current multitude of experimental techniques, inference algorithms, and networks makes it

important to investigate what an interaction actually represents in each network and distinguish between different types of interactions.

The implication of latent states has, with the exception of previous discussions *e.g.* in [Goncalves and Warnick \(2008\)](#) and [Margolin and Califano \(2007\)](#), largely been neglected in inference of networks based on perturbation experiments. Considering that today in inference of GRNs it is common practice to either only observe the change in mRNA, protein, or metabolite abundance, neglecting all other state variables of the system ([Hecker et al., 2009](#); [Tegnér and Björkegren, 2007](#); [Brazhnik et al., 2002](#)), it is clear that unobserved and unperturbed latent states almost always are present and their existence need to be accounted for. Existence of latent states, among other things, implies that it is necessary to know a linearly independent subset of  $n_1$  relations between the states of interest and responses or perturbations. The sum of the observed and perturbed variables must therefore at least equal the number of genes of interest to infer a subnetwork. In general this is not sufficient but it is always necessary, contrary to the necessary conditions for inference of the dynamical structure function established by [Goncalves and Warnick \(2008\)](#), which we showed not to be necessary. However, when only steady-state data is available then it is necessary to independently both perturb and observe each variable of interest to infer the subnetwork existing among them. Independent perturbation and observation of each gene of interest is always sufficient for inference of subnetworks. These conditions are already in many cases fulfilled in inference of GRNs, because it is common place to measure mRNA abundances of individual genes and perturb them *e.g.* by insertion of a plasmid containing an extra copy ([Gardner and Faith, 2005](#); [Cantone et al., 2009](#)). RNA interference and endogenous perturbations resulting in copy number aberrations also affect single genes ([Wheeler et al., 2005](#); [Jörnsten et al., 2011](#)). In other words, the techniques and methods to even fulfil this sufficient condition already exist and have been used in network inference. It is worth stressing that the ability to infer subnetworks offers a mean to avoid underdetermined problems and achieve confidence when the number of variables exceeds the number of samples. This is in particular important in steady-state experiments, because even though all genes are observed it is then still only possible to infer the subnetwork of genes that were perturbed. These are the genes that should be included in the network model. If one tries to infer the full network, then the resulting model is determined by the sparsity assumption or *a priori* knowledge and no confidence can be assigned to any interaction, as discussed in Chapter 4. While the existence or lack-off of intermediary states never can be determined from steady-state data, it can in principle be determined from time-series data. In practice it depends on how informative the time-series data is and how well the perturbations excite the latent states. One way to represent this knowledge is by estimating transfer functions from the data, which concisely represent the interactions between states of interest. Transfer functions are widely used in engineering and systems theory, see *e.g.* [Lindner \(1999\)](#); [Kailath \(1980\)](#).

To increase the impact of the theoretical limitations of inference based on perturbation experiments that we have explored here, we now give some practical

---

advice to practitioners dealing with inferred networks or inferring networks. Latent states have a profound impact both on what can be inferred and the meaning of the inferred interactions. So unless it is known that all states of the system are observed and perturbed, one should always assume that latent states exist and recognize that a subnetwork in fact is inferred. One should therefore always ask what the links represent and check the assumptions made during inference, in particular regarding existence of latent states. When using network inference to determine the genetic mechanism behind a function of interest we recommend the following procedure: Perturb something that affects the function and measure the change in all genes when the system has reached a new steady-state. Next perturb a gene that changed. If this perturbation affected the function, then this gene is essential and should be included in the subnetwork, otherwise it can be excluded. Repeat this step until a subset of genes that are essential emerges and hence forms the set of states of interest. After this, experiment design should be employed to ensure that the data is informative for inference of this subnetwork, as described in Chapter 7. Finally, time-series experiments can be performed if knowledge about intermediary states is desired.



---

## Design of perturbation experiments for network inference

---

*“It is a capital mistake to theorize before one has data. Insensibly one begins to twist facts to suit theories, instead of theories to suit facts.”*

Sir Arthur Ignatius Conan Doyle, *The adventures of Sherlock Holmes: A scandal in Bohemia*, 1892.

The intrinsic ability of gene regulatory networks to amplify and attenuate different signals makes inference of existing causal interactions from expression data challenging. We demonstrate that recently published data on the *Snf1* signalling pathway in *S. cerevisiae* lack necessary information for inference, even though the authors perturbed all 10 genes one-by-one in separate experiments in an attempt to obtain enough information. As we show, the only foreseeable remedy is iterative design of correlated multi-gene perturbation experiments that explores the multivariable gains of the system and thereby counteracts the signal attenuation. We establish sufficient conditions on data for robust variable selection and network inference, and propose the following design principle—select the next perturbation such that the expected response practically spans an additional dimension in the state space. This principle is here numerically demonstrated by designing additional steady-state experiments for inference of the *Snf1* signalling pathway. On average 15 and in the worst case 35 additional experiments were needed to increase the strength of the weakest signal 33 times and obtain informative data in 1000 Monte Carlo simulations. In comparison, designs using only random or single gene perturbations yielded poor data in all simulations even after 100 additional experiments. Large improvements in data quality are hence obtainable by replacing the random and single gene designs currently used. The ability to combine and scale previous perturbations is essential for the iterative experiment design and hence calls for development of improved experimental protocols.

## 7.1 Introduction

Interacting genes, proteins and metabolites form a dynamical system controlling cellular processes (Wolkenhauer et al., 2005b; Sontag, 2005; Csete and Doyle, 2002). Several studies have demonstrated that network inference based on gene expression data obtained from *in vivo* experiments, in which the system is perturbed by known disturbances and the following response measured, has the potential to reveal all direct causal influences within a set of observed genes (Tegnér and Björkegren, 2007; Cho et al., 2007; Crampin, 2006; Gardner and Faith, 2005; Goncalves and Warnick, 2008). Inferred gene regulatory networks (GRNs) are approximations/abstractions of the intracellular system of interacting molecules providing insight on the central genetic control structure, *e.g.* feedback and feedforward loops, while hiding details such as mediator proteins and metabolites. To successfully infer the GRN that underlies a biological function of interest, based on expression changes, two basic problems must be solved. First, informative data that allow discrimination between models with different network structure must be recorded by performing *in vivo* experiments on the system in the physiological mode where the function is active. Second, the “true” network model with a structure including only the active gene interactions must be selected based on the recorded data set. A multitude of traditional estimation methods and novel inference algorithms have been adopted within the last decade to recover the network from given gene expression data; see the review articles (Hecker et al., 2009; Li et al., 2008b; Tegnér and Björkegren, 2007; Cho et al., 2007; Bansal et al., 2007; Gardner and Faith, 2005; Styczynski and Stephanopoulos, 2005; Doyle and Lauffenburger, 2005; van Someren et al., 2002; D’haeseleer et al., 2000). In particular, incorporation of network sparsity in the selection algorithm has been of major interest. Currently one can, in our opinion, speak about a race towards the best solution of the second problem, boosted by the DREAM challenges in which the solutions are benchmarked against each other (Marbach et al., 2010). Relative to the popularity of the second problem, the first one is more or less being neglected even though a solution of it is a prerequisite for solving the second one, since obviously no selection algorithm can create information. Indeed, current data sets are characterized by high dimensionality but few samples, low signal to noise ratios, errors-in-variables, and ill-conditioned response matrices in which most of the variation can be explained by a few linear combinations of the variables (see *e.g.* Holter et al., 2000; Alter et al., 2000; Tegnér and Björkegren, 2007; Faith et al., 2007; Cosgrove et al., 2010; Nordling and Jacobsen, 2009a; Wu and Wu, 2010). Inference algorithms applied to poor data lacking necessary information do in general not reveal that the data is poor but rather select a false network model. Poor data implies existence of at least one false network model that cannot be falsified by the available data, since it explains the data well.

We here address the first problem, that of generating sufficiently informative data for network inference through design of perturbation experiments. To stress the necessity of better design, we first demonstrate that recently published data from a reverse engineering study of the Snf1 signalling pathway in *S. cerevisiae* (Lorenz

*et al.*, 2009) lacks necessary information, even though all genes were perturbed. Motivated by the lack of information, we analyse the cause of it, and based on our theory for robust variable selection in Chapter 5 we establish sufficient criteria on the data for robust variable selection and network inference in Section 7.3. Based on observations of signal amplification and attenuation by the system, we propose iterative design of correlated perturbations as the only foreseeable solution in Section 7.4. We then identify the fundamental principle for design of the next perturbation in Section 7.5, and demonstrate the use of it for design of steady-state experiments in Section 7.6. Before presenting our results, we give an introduction to design of experiments, and point out the distinction between previous work and this work. The results are followed by a discussion on the demand for experimental multi-gene perturbation protocols and importance of data quality and analysis.

### 7.1.1 An overview of design of experiments for model selection

Design of experiments (DoE) is since long used in statistics to obtain more information in fewer experiments by tailoring the experiments to the purpose of the study (Pronzato, 2008; Chaloner and Verdinelli, 1995; Steinberg and Hunter, 1984; Ljung, 1999; Fisher, 1935). The idea is to plan the experiments such that when a mathematical model is fitted to or trained on the recorded data then it provides a sufficiently good approximation of the properties of interest, while minimising the experimental effort. Model construction can in general be divided into two steps: selection of the model structure, and estimation of parameters (Ljung, 1999). From a modelling perspective network inference is model selection, since the primal concern is which interactions to include in the model, while determination of the strength of the interactions is parameter estimation. It is tradition to differentiate between DoE for parameter estimation, which is used to decrease parameter uncertainty for a chosen model structure, and DoE for model/variable selection, which is used to distinguish among a set of alternative model structures, since the two tasks typically require different data (Schwaab *et al.*, 2008). In particular, experiments designed to minimise parameter covariance should not be used for model selection, since measures of parameter uncertainty are only meaningful for the selected model structure. Three recent reviews of DoE related to systems biology, Kreutz and Timmer (2009); Banga and Balsa-Canto (2008); Franceschini and Macchietto (2008), provide a good overview of DoE for parameter estimation and, in particular, a subset of methods known as optimal experimental design (OED). However, they only give a brief introduction to DoE for model selection, which is the main interest here. The combined introductions of Schwaab *et al.* (2008) and Michalik *et al.* (2010) provide an excellent overview of OED for model selection, which we next complement by other DoE works related to systems biology and published within the last decade. Previous works can roughly be grouped into three categories: OED for discrimination among dynamical models, discrimination among non-dynamical models, and heuristic methods or rules for choosing experiments.

The first group concerns OED for discrimination among a specified set of

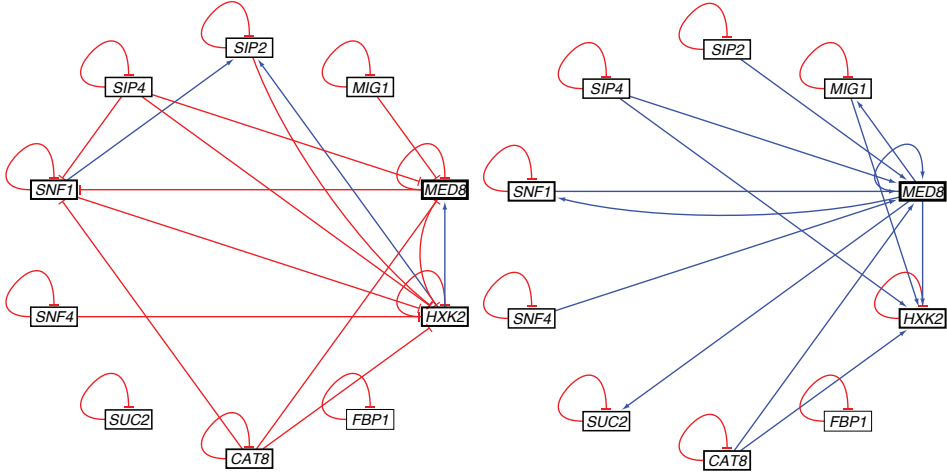
alternative nonlinear dynamical models, *i.e.* hypothesis testing, and require *a priori* knowledge or data both for constructing the alternative models and estimation of their parameters (Michalik et al., 2010; Skanda and Lebiecz, 2010; Kremling et al., 2004). This implies that the proposed methods are unsuitable for generation of interaction hypotheses, which network inference so far mainly has been used for, and that model and parameter errors may have a large impact on the designed experiments and discriminatory power. The current research theme is to account for model and parameter uncertainty through modified objective functions (see *e.g.* Donckels et al., 2009; Schwaab et al., 2008), an approach which further increases the *a priori* information and data demand. In the second group, discrimination among Boolean and Bayesian models has been investigated (Ideker et al., 2000; Steinke et al., 2007; Yoo and Cooper, 2003). These are computationally expensive when the sets of alternative models and possible perturbations are large. They fail to properly account for noise in the data, but they are suitable for generating and testing interaction hypotheses in Boolean or Bayesian models. For instance, Steinke et al. (2007) assume noise-free measurements of the steady-state responses. The third group consists of works where some heuristic method or rule for choosing experiments is proposed. An example of this is Tegnér et al. (2003), who suggest two heuristic rules for selecting which gene to perturb next. Experiments generated by their rules are suitable for both hypothesis generation and testing, but it may be computationally expensive to determine all models consistent with data as needed and informative data is not guaranteed. Apgar et al. (2008), on the other hand, use model based controllers to design dynamical stimuli for driving candidate models through a target trajectory so that the models can be distinguished. They select the target trajectory based on suitability for the available measurement methods and do not explicitly optimize the discriminatory power. Their method is made for discrimination among a few candidate models and it therefore requires *a priori* knowledge or data, implying that it is not suitable for generation of interaction hypotheses.

Contrary to previous works, we present a simple yet powerful principle for how to design the next perturbation experiment given recorded data, shown to be fundamental for robust variable selection and network inference. Application of the principle does not require any *a priori* information, nor does it at any stage require an explicit model of the system, so it can be applied at any stage of the data collection. At the same time prior information, data and system models can easily be incorporated to decrease the required number of experiments. It thereby avoids the cons of previous DoE approaches, while maintaining their pros. We show that proper application of the principle yields informative data. We exemplify the principle by implementing an algorithm for design of steady-state experiments and use it to iteratively design step perturbations for inference of the Snf1 signalling pathway in *S. cerevisiae*, starting from published *in vivo* steady-state data (Lorenz et al., 2009). The algorithm is based purely on the singular value decomposition (SVD), a classical tool in matrix algebra (see *e.g.* Horn and Johnson, 1990, p. 414-415) for which effective computational algorithms exist.

## 7.2 Published data with insufficient information

Lorenz *et al.* (2009) use steady-state transcriptional perturbation experiments to infer the interactions of an ageing and glucose repressed GRN consisting of 10 genes from the Snf1 signalling pathway in *S. cerevisiae*. They over-expressed each of the genes, one-by-one, using strains containing an integrated second copy of the perturbed gene under control of a doxy-cycline-inducible promoter and measured the expression of all genes by real-time quantitative polymerase chain reaction (qRT-PCR) relative to an isogenic control strain. When testing their interaction hypotheses, using promoter-reporter gene fusions in gene-deletion mutants, chromatin immunoprecipitation experiments and interactions reported in the literature, they found that their network model showed a mere 62% sensitivity and 69% precision. Sensitivity measures the percentage of known interactions that are successfully inferred, while precision measures the percentage of inferred interactions that are consistent with known interactions. Moreover, 24% of the predicted interactions had the wrong sign, *i.e.* activations were erroneously inferred as repressions and *vice versa*. When analysing if the weak agreement can be explained by poor data, we made the following discovery. For each of the 100 possible interactions we found at least one network model without it that could not be rejected based on a model goodness of fit test using their data (Lorenz *et al.*, 2009, supplementary Table S2-S4). Lorenz *et al.* assumed the measurement errors to be independent and zero mean normally distributed, and provided estimates of the standard deviation of each measured response so we used the  $\chi^2$  goodness of fit test with p-value 0.05 (see *e.g.* Jaqaman and Danuser, 2006; Cedersund and Roll, 2009). We present in Sections C.1 and C.2 18 non-rejectable network models that together proves that each possible interaction can be left out. Since many practically indistinguishable models with different structure obviously exist, none of the inferred interactions can be statistically significant. An example of two models with different network structure that cannot be rejected and hence are practically indistinguishable is given in Figure 7.1. The structure, *i.e.* signed topology, of the two networks differs considerably, as seen by the low number of identical links, *i.e.* interactions present in both models with same sign. More precisely, they only have 9 links with the same sign in common, while 7 links have the opposite sign and 15 links only exist in one of the networks. All 9 links with the same sign are self-loops, which due to mRNA degradation typically exist, so these networks are in principle as different as they in general can be.

Already the fact that both network models in Figure 7.1 can explain the data well despite being different proves that the data cannot be informative for robust network inference. Lorenz *et al.* perturbed each gene in an independent experiment, yielding a diagonal perturbation matrix  $\mathbf{P} \in \mathbb{R}^{n \times m}$ , and measured the steady-state expression change of all genes following each perturbation, yielding a response matrix  $\mathbf{Y} \in \mathbb{R}^{n \times m}$  with full rank and as many measurements as unknown parameters in the sought interaction matrix  $\mathbf{A} \in \mathbb{R}^{n \times n}$ . They solve the network inference problem



**Figure 7.1: Two different network models that fit well to the *in vivo* data of Lorenz *et al.* (2009), based on commonly used measures such as the unweighted and weighted residual sum of squares (Ashyraliyev *et al.*, 2009), prediction error (Ljung, 1999), and cannot be rejected based on a  $\chi^2$  goodness of fit test (Jaqaman and Danuser, 2006; Cedersund and Roll, 2009). Out of the large set of models that can explain the data of Lorenz *et al.*, we have selected to present these two network models, since they are sparse and the structural discrepancies despite good fit clearly illustrate that the data is not informative enough for inference of the structure of the GRN. Note that they are not intended to represent current biological knowledge. Here every link represents a direct causal interaction. An arrow shaped head on the link from *SNF1* to *SIP2* (left) indicates that *SNF1* upregulates (activates) the *SIP2* gene, while a (T) shaped head from *SNF1* to *HXK2* (left) indicates that *SNF1* downregulates (represses) the *HXK2* gene. The figure was generated in Cytoscape (Shannon *et al.*, 2003).**

by minimizing the sum of squared errors

$$\arg \min_{\mathbf{A}} \sum_{i=1}^n \sum_{k=1}^m \left( \sum_{j=1}^n a_{ij} y_{jk} + p_{ik} \right)^2, \quad (7.2.1)$$

using sparsity priors in terms of the number of nonzero elements  $a_{ij}$  per row (Lorenz *et al.*, 2009). A unique solution therefore exists for all reasonable objectives and it might appear surprising that alternative models with different structure can explain the data equally well. In other words, a closer examination of data requirements for robust network inference, which we provide next, is of essence.

### 7.3 Informative data for variable selection and network inference

Network inference is variable selection and we previously in Chapter 5 developed a theory for robust variable selection and network inference. We here recap the main results needed to explain the indistinguishability encountered in the data by Lorenz et al. (2009) and to specify the data requirements for robust variable selection and network inference. These are later used to introduce a principle for design of perturbation experiments in Section 7.5 and propose a simple design algorithm in Section 7.6.

The objective of network inference is, in short, to find the interactions that exists among a set of variables based on observations of them. Ability to distinguish between network models with different structure/topology/digraph based on recorded data is therefore essential. Indistinguishability can be caused for a considered set of parametrised model structures by structural/*a priori* unidentifiability of the model parameters, which is due to introduction of too many or linearly dependent parameters, or practical/*a posteriori* unidentifiability, which is due to measurement errors, noise, or uncertainty (see e.g. Ashyraliyev et al., 2009; Cedersund and Roll, 2009; Ljung, 1999, p. 105-113). In other words, the former is due to lack of observation or perturbation of some variables, which we investigated in Chapter 6, while the latter is due to lack of information in the recorded data, *i.e.* poor data, which is our focus here. Gene expression data is known to be scarce and noisy (Hecker et al., 2009; van Someren et al., 2002; Gardner et al., 2003), which strongly affects estimated parameters of network models and biases them (Fujita et al., 2009). Measurement errors, noise and uncertainty are in inference often dealt with by using an inference, selection, or estimation algorithm to search for the most likely model in some sense, but in robust network inference we, for the reasons given below, instead search for all models that cannot be rejected based on data. The ability to reject or invalidate a model is determined purely by the data, measurement assumptions and epistemology (Barlas, 1996). It is independent of how the model was obtained, *i.e.* the selection, inference, or estimation algorithm, which makes model rejection ideal for robust variable selection and network inference.

The set of all network models that cannot be rejected contains all alternative models that can explain data, so interactions that are present in all of these models are necessary to explain the data and can therefore be proven to exist at the given significance level. We therefore formally call these interactions existing and the corresponding variables practically selectable. In general, interactions and variables can, in layman terms, be classified as existing/present, non-existing/absent, non-evidential, and alternative. The data can always be explained without non-existing interactions, while the data contains no information about non-evidential and insufficient information about alternative ones. We formally call the corresponding variables practically excludable, practically negligible, and practically alternative. In other words, if the data is informative enough, then all possible interactions are

classified as existing or non-existing and the corresponding variables as practically selectable or excludable. We thus henceforth focus on data conditions for practical selectability and excludability. For further details see Chapter 5, where we developed the theory for classification of variables and interactions that forms the basis of robust variable selection and network inference. We next describe the connection from data to robust network inference via robust variable selection.

In accordance with the assumptions for our theory of robust network inference in Chapter 5, we restrict ourselves to inference of a linear map or transformation  $\mathbf{A}$ , consisting of the elements  $a_{ij}$  that constitutes the interactions of the network, and assume that the data can be described by the following data model

$$\phi_j = \check{\phi}_j + \mathbf{v}_j, \quad \xi_i = \check{\xi}_i + \epsilon_i, \quad \phi_j, \xi_i \in \mathbb{R}^m, \quad (7.3.1a)$$

$$\sum_{j=1}^n \check{\phi}_j \check{a}_{ij} = \check{\xi}_i, \quad \check{a}_{ij} \in \mathbb{R}, \quad \forall i, j \in \mathcal{V} \triangleq \{1, 2, \dots, n\}. \quad (7.3.1b)$$

In this data model the observed variables, which we call regressors  $\phi_j$  and regressands  $\xi_i$ , are equal to the latent “true” variables, marked by  $\check{\phantom{x}}$ , plus additive errors  $\mathbf{v}_j$  and  $\epsilon_i$ , respectively. We assume that the “true” regressors and regressands have been generated by a system of linear equations defined by the “true” interaction matrix  $\mathbf{A}$ , *i.e.* the “true” map representing the network model including the observed states of the system. This data model is identical to the one assumed by [Lorenz et al. \(2009\)](#), with each regressor  $\phi_j$  corresponding to the observed responses in gene  $j$  and each regressand  $\xi_i$  corresponding to the perturbations of gene  $i$ . In general, the errors  $\mathbf{v}_j$  and  $\epsilon_i$  are either assumed to be stochastic noise, *i.e.* random variables with known probability distributions, or unknown deterministic errors bounded by some known function. The former uncertainty description is standard in statistics, signal processing and system identification (see *e.g.* [Casella and Berger, 2001](#); [Kay, 1993](#); [Ljung, 1999](#)), while the latter typically is used in numerical analysis and robust control (see *e.g.* [Higham, 1996](#); [Skogestad and Postlethwaite, 1996](#)). [Lorenz et al.](#) estimated the standard error of each data point and assumed the measurement errors to be independent and normally distributed with zero mean.

Note that the dynamics of the underlying system may be nonlinear even though the model of the network is linear (see *e.g.* [Jörnsten et al., 2011](#); [Crampin et al., 2004](#)). From a model rejection perspective, however, it is sufficient to note that for each consistent realization of the errors,  $\mathbf{v}_j$  and  $\epsilon_i$ , (7.3.1) provides a system of linear equations

$$\begin{aligned} \sum_{j=1}^n a_{ij} \tilde{\phi}_j &= \tilde{\xi}_i, & \forall \text{ consistent } \tilde{\phi}_j \in \mathcal{U}_{\phi_j}^\alpha \subseteq \mathbb{R}^m \\ & & \text{and } \tilde{\xi}_i \in \mathcal{U}_{\xi_i}^\alpha \subseteq \mathbb{R}^m. \end{aligned} \quad (7.3.2)$$

Here we use the uncertainty sets  $\mathcal{U}_{\phi_j}^\alpha$  and  $\mathcal{U}_{\xi_i}^\alpha$  to represent the effect of the errors, because they in a precise manner capture the essential fact that we have a set

of realizations of each variable that are indistinguishable from the observed or “true” one. To obtain necessary conditions, we only consider solutions of (7.3.2) for combinations of the realizations of the regressors and regressand that cannot be rejected at the desired significance level  $\alpha$  based on the assumed error model, which for certain error models is a subset of all possible combinations in the uncertainty sets, and mark this by use of the word consistent, see Section 5.4. In practice, the “true” values and uncertainty are rarely known, but we can then instead use the observed values and estimates of the uncertainty and assume that the “true” values are within the uncertainty sets. Also, instead of using estimates of the uncertainty, we can ask how large the uncertainty of the observed data can be for rejection of a certain set of models. Nonetheless, for each row  $i$  of the interaction matrix we always get one system of linear equations for each consistent combination of realizations of the variables that cannot be rejected based on the data and assumed error model. To do robust variable selection we need to determine the regressors that are present in all of these systems and thus must be selected, *i.e.* practically selectable, or absent in some systems for all realizations of the regressand and thus always can be excluded, *i.e.* practically excludable. Note that all practically selectable regressors are true positives, assuming that the “true” model belongs to the set of indistinguishable models, while all practically excludable variables are true negatives, assuming that variables with weak relation should be excluded. For continuous variables the number of models is always infinite and checking all of them is computationally infeasible, but the data conditions for practical selectability and excludability that we established in Chapter 5 convert this into a rank problem that can be solved without solving or checking any model, so we next build upon them. Let us, however, first note that each row of the interaction matrix  $\mathbf{A}$  can be inferred independently of the other rows, because we have for each  $i$  in (7.3.2) a different set of systems of equations. In other words, each robust network inference problem consists of  $n$  robust variable selection problems.

The objective of robust variable selection is based on Definition 5.2.1 to find the variables that for all realizations  $\tilde{\xi}_i \in \mathcal{U}_{\xi_i}^\alpha$  either must be selected or always can be excluded for some realizations  $\tilde{\phi}_j \in \mathcal{U}_{\phi_j}^\alpha$  in (7.3.2), *i.e.* the subset of practically selectable regressors and the subset of practically excludable regressors. It is therefore logical to make the following definition of informative data for robust variable selection.

**Definition 7.3.1. Informative data.**

A data set is informative enough for robust variable selection if all regressors are either practically selectable (Definition 5.6.1) or practically excludable (Definition 5.7.2).

Consequently, if a data set is informative enough for all  $n$  variable selection problems, that each robust network inference problem consists of, then it is informative enough for robust network inference. Necessary and sufficient conditions for a data set to

be informative enough are best expressed using the practical span introduced in Section 5.5.

**Corollary 7.3.1. Informative data for robust network inference.**

A data set is informative enough for robust network inference if and only if it for all  $i \in \mathcal{V} = \{1, 2, \dots, n\}$  holds that each regressor  $\phi_k$  either is practically selectable,

$$\tilde{\xi}_i \notin \text{pspan } \Phi_{j \neq k} \quad \forall \text{ consistent } \tilde{\xi}_i \in \mathcal{U}_{\xi_i}^\alpha \subseteq \mathbb{R}^m, \quad (7.3.3)$$

or practically excludable,

$$\tilde{\xi}_i \in \text{pspan } \Phi_{j \neq k} \quad \forall \text{ consistent } \tilde{\xi}_i \in \mathcal{U}_{\xi_i}^\alpha \subseteq \mathbb{R}^m \quad (7.3.4)$$

and

$$\tilde{\phi}_k \notin \text{pspan } \Phi_{j \neq k} \quad \forall \text{ consistent } \tilde{\phi}_k \in \mathcal{U}_{\phi_k}^\alpha \subseteq \mathbb{R}^m, \quad (7.3.5)$$

with  $\Phi_{j \neq k} \triangleq [\phi_1, \dots, \phi_{k-1}, \phi_{k+1}, \dots, \phi_n]$ . Here  $\text{pspan}$  denotes the practical span (Definition 5.5.11).

**Proof.** Either practical selectability is ensured by (7.3.3) based on Definition 5.6.1 or practical excludability by (7.3.5) and (7.3.4) based on Definition 5.7.2 in Chapter 5 for all  $i \in \mathcal{V}$  and Definition 7.3.1 is therefore fulfilled for each of the  $n$  robust variable selection problems.  $\square$

The practical span is, however, in general not computable, so we next give sufficient conditions based on practical independence of data matrices, which computationally can be checked using the singular value decomposition or structured singular value as described in Section 5.10. In short, a matrix is practically independent if the set of all columns is linearly independent for all consistent realizations, while it is practically collinear if the set is linearly dependent for all consistent realizations.

**Corollary 7.3.2. Informative data for robust network inference.**

A data set is informative enough for robust network inference if  $\Phi \triangleq [\phi_1, \dots, \phi_n]$  is practically independent (Definition 5.5.13) and either  $\Psi_{ki} \triangleq [\phi_1, \dots, \phi_{k-1}, \phi_{k+1}, \dots, \phi_n, \xi_i]$  is practically independent or practically collinear (Definition 5.5.14) for all  $k, i \in \mathcal{V} = \{1, 2, \dots, n\}$ .

**Proof.** This follows trivially from Corollary 7.3.1 based on Theorem 5.6.2 and Theorem 5.7.1 in Chapter 5.  $\square$

The primary condition is practical independence of the regressor matrix  $\Phi$  and when it is fulfilled then we say that the regressors practically span the state space. By state space we refer to the space whose axes are the observed variables, which constitute the set of state variables of interest. Practical spanning is thus in practice necessary for robust variable selection and network inference, and we therefore later on stress its importance in design of perturbation experiments. In order for the data to be informative enough, the uncertainty sets  $\mathcal{U}_{\phi_j}$  may not contain the origin,

which implies that the norm of each regressor  $\|\phi_j\|$  needs to be sufficiently large relative to its uncertainty set, nor any consistent point of the uncertainty sets of the other regressors, which implies that the angle between any pair of regressors need to be sufficiently large relative to their uncertainty sets. Note that a data set that is not informative enough for robust variable selection within the class of linear functions, *i.e.* set of linear models, cannot be informative for the class of nonlinear functions, since the latter is a superset of the former.

Let us finally examine if the data by [Lorenz et al. \(2009\)](#) is poor based on practical selectability and practical spanning of the corresponding regressors. We use four quality indicators of data: existence of alternative models that cannot be rejected, the confidence score, the signal to noise ratio, and the condition number. First, the fact that each of the 100 possible interactions is missing in at least one of the 18 non-rejectable network models mentioned earlier proves that none of the regressors is practically selectable and none of the interactions  $a_{ij}$  can therefore be classified as existing. Second, the confidence score, defined as

$$\gamma(\Phi) \triangleq \sigma_n(\Phi(\chi)), \quad \text{with } \phi_{ij}(\chi) \triangleq \frac{\phi_{ij}}{\sqrt{\chi^{-2}(\alpha, nm)\lambda_{ij}}}, \quad (7.3.6)$$

needs to be larger than one in order for the regressor matrix  $\Phi = \mathbf{Y}^T$  to be practically independent, as proven in [Theorem 5.10.2](#) for independent and normally distributed errors. Here  $\sigma_n$  denotes the  $n$ th singular value with  $n$  being the number of variables, and  $\chi^{-2}(\alpha, nm)$  the inverse of the chi-square cumulative distribution with  $nm$  degrees of freedom at significance level  $\alpha$ . The variance  $\lambda_{ij}$  of each data point accounts for differences in the uncertainty. The confidence score is only 0.03 at significance level 0.05, so the regressors cannot be practically independent nor practically span the state space. Note that we assume the errors to be independent and normally distributed with variance equal to the square of the standard errors reported by [Lorenz et al.](#) for all data points except the unperturbed genes. For them we, in lack of a better estimate, assume the standard error 0.1, which is approximately one third to one quarter of the standard errors reported for perturbed genes, to account for variations in the cell cultures. Third, if the signal to noise ratio (SNR), defined as

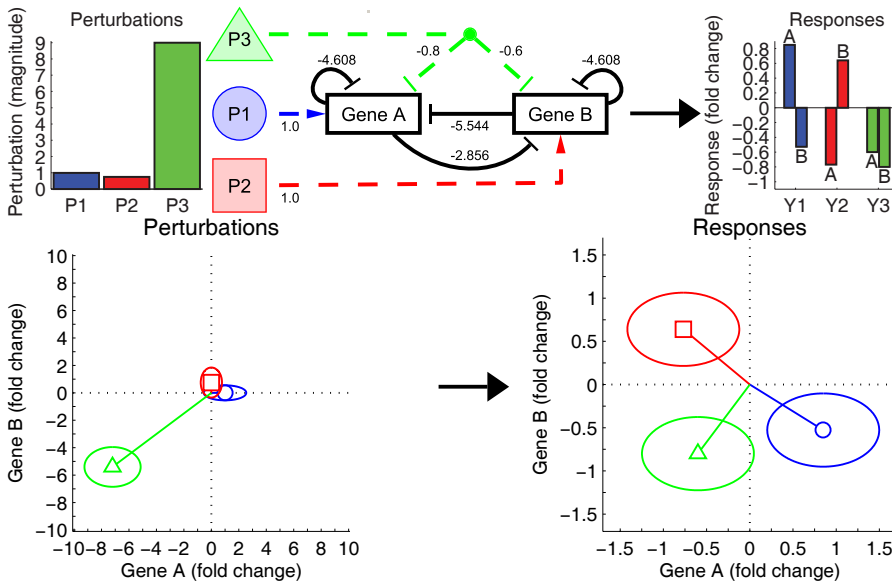
$$\text{SNR}(\phi_j) \triangleq \frac{\|\phi_j\|}{r_{\mathcal{U}_{\phi_j}^\alpha}}, \quad (7.3.7)$$

is smaller than one for any regressor, then the set of regressors is practically collinear and the data is poor. Here  $r_{\mathcal{U}_{\phi_j}^\alpha}$  denotes the radius of the uncertainty set at significance level  $\alpha$ . Using the standard errors reported by [Lorenz et al.](#) to construct the uncertainty sets as described in [Section 5.4](#), we find that the smallest SNR is only 0.33 at significance level 0.05. This implies that at least one realization of the corresponding regressor is parallel to any one realization of the other regressors and the data is poor. A problem known as near collinearity in regression, see *e.g.* [Belsley \(1991\)](#); [Alin \(2010\)](#), which hampers inference by any method. Fourth, the condition

number of the regressor matrix is 215. This value is high for a system with only 10 states, so the response matrix is ill-conditioned, which is well known to hamper network inference and parameter estimation, see Chapter 3. In conclusion, all four quality indicators show that the data is poor, which explains the severe practical distinguishability problems that we have pointed out earlier.

## 7.4 System properties necessitate iterative design

We have so far established practical spanning as a prerequisite for robust inference of GRNs and observed that the response matrix of Lorenz *et al.* does not practically span the gene space even though they perturbed all genes independently. To explain why their response matrix is ill-conditioned and show why iterative design of perturbation experiments is needed, we consider the two gene example in Figure 7.2, which we previously used in Chapter 3. The perturbations P1 and P2 are orthogonal, Figure 7.2 (bottom left), but the corresponding responses are



**Figure 7.2:** An interampatte GRN involving two genes A and B (top, middle) with three perturbations: P1, P2 and P3 (top left and bottom left) and their responses (top right and bottom right). An arrow shaped head on the link indicates up-regulation or activation, while a bar shaped head indicates down-regulation or repression, with the strength of the interaction given by the number on the link. The ellipsoids around the data vectors (bottom) marks the uncertainty set of each experiment at significance level 0.05. This two gene network was earlier used in Chapter 3, where it is described in more detail.

almost linearly dependent, Figure 7.2 (bottom right), due to the high degree of interampattiness of the two gene system. Interampattiness—strong INTERactions for simultaneous AMPLification and ATTEnuation of different signals—is a generic system property caused mainly by cascades and feedback loops within the GRN that leads to amplification and attenuation of different signals, *e.g.* P2 is amplified by a factor 1.3, while P3 is attenuated by a factor 9 in our example. Interampattiness explains the observation that most of the variation in microarray data can be accounted for by a few linear combinations of all genes, which previously have lead to the introduction of characteristic modes (Holter *et al.*, 2000, 2001) and eigengenes (Alter *et al.*, 2000; Nielsen *et al.*, 2002; Alter and Golub, 2006; Omberg *et al.*, 2007). We analysed both the cause and effect of interampattiness in detail, and postulated that GRNs are interampatte in Chapter 3. Inference of interampatte networks is hampered by the intrinsic tendency of the system to channel responses in the high gain directions where the signals are amplified, leading to ill-conditioned response matrices with weak signal in some direction of the gene space. To practically span the space, correlated perturbations that counteract the intrinsic attenuation, like P3 in our example, are needed. These correlated perturbations are specific to the unknown system and can therefore in practice only be designed iteratively based on information from previous perturbations and responses. The only alternative strategy is to lower the degree of interampattiness of the system, but this would require removal of unknown cascades and feedback loops, rendering it impracticable. In other words, it is necessary to excite the weak gain directions of the system to get informative data for robust variable selection, just as excitation of the weak gain is important for plant inversion in process control (Bruwer and MacGregor, 2006; Jacobsen, 1994). All perturbations used by Lorenz *et al.* are orthogonal and the magnitude of each perturbation and corresponding response is of similar order, so the weak signals in their data are due to the high degree of interampattiness of the Snf1 signalling network, which we established in Section 3.7.

## 7.5 Fundamental principle for design of experiments

The data requirements for robust variable selection in Section 7.3 lead us to propose the following principle for iterative design of perturbation experiments: design the next perturbation such that the expected response practically spans an additional dimension.

This principle should based on Corollary 7.3.2 first be used to make the regressor matrix  $\Phi$  practically independent, by designing the next perturbation such that the expected response practically spans an additional dimension of the state space. More precisely, such that the smallest dimension spanned by any of the consistent realizations of the obtained regressor matrix is expected to increase. The necessary perturbation experiments depend on the unknown “true” system and an iterative design that explores the gains of the system is therefore needed. In practice, this implies exploration of the gains of the system while striving towards excitation of

the weak gains, such that the intrinsic signal attenuation is counteracted, and the state space is practically spanned.

**Conjecture 7.5.1. DoE for practical spanning of the state space.**

Iterative design of the next perturbation experiment such that the expected response practically spans an additional dimension of the regressor matrix leads to practical spanning of the state space.

Proper application of this principle will, under mild assumptions, after a sufficient number of iterations yield data in which the state space is practically spanned. Each perturbation whose response does not practically span an additional dimension provides information that enables calculation of a better estimate of the perturbation needed to practically span the dimension. To formally prove convergence, however, requires specification of how the next perturbation is calculated and we therefore stated this as a conjecture in order to avoid technicalities that could dilute our message. Note that *e.g.* moiety conservation may give rise to linear dependencies among genes, so if it turns out to be impossible when designing perturbations for inference of a GRN to find a perturbation such that the regressor matrix becomes practically independent, then some gene and corresponding regressor is linearly dependent and should be excluded from the network model (see *e.g.* [Schmidt et al., 2005](#)). Every design algorithm should explore the state space such that the data practically spans it, since practical spanning of the state space for continuous variables is necessary for robust variable selection.

Practical spanning of the state space does, however, not guarantee that the data is informative enough for robust variable selection nor network inference. Even though the regressor matrix  $\Phi$  becomes practically independent and the magnitude of the regressand  $\xi$  increases with each experiment it does not guarantee that all  $\Psi_{ki}$  are either practically independent or collinear, as required in Corollary 7.3.2. The principle for iterative design of perturbation experiments should therefore, once the data is informative enough for practical independence of the regressor matrix, be used to design the next perturbation such that the expected response makes some  $\Psi_{ki}$  practically independent, until all  $\Psi_{ki}$  are either practically independent or collinear. Each time a perturbation experiment fails to make  $\Psi_{ki}$  practically independent, one should check if it is practically collinear, since it is possible that the corresponding interaction does not exist in the “true” network. If it is found to be practically collinear, then one should move on to the next  $\Psi_{ki}$  until all are either practically independent or collinear. The principle of designing such that the expected response practically spans an additional dimension is exactly the same in this case, but it is applied to make  $\Psi_{ki}$  practically independent instead of  $\Phi$ . Based on the discussion above we therefore state the following conjecture.

**Conjecture 7.5.2. DoE for robust selection and inference.**

Iterative design of the next perturbation experiment such that the expected response practically spans an additional dimension, first for the regressor matrix

$\Phi = [\phi_1, \dots, \phi_k, \dots, \phi_n]$  and then for all  $\Psi_{ki} = [\phi_1, \dots, \phi_{k-1}, \phi_{k+1}, \dots, \phi_n, \xi_i]$ , yield informative enough data for robust variable selection and network inference.

The principle is hence fundamental for DoE for robust variable selection and network inference, starting with no or incomplete *a priori* knowledge.

We propose the addition of an inner loop to the classical model construction cycle—design perturbations, perform the experiments, measure the response, model/variable selection, parameter estimation, model validation, and new cycle if needed. The idea of the inner loop is to ensure practical spanning, in order to avoid costly and misleading iterations of the classical cycle due to poor data. The proposed loop is illustrated in Figure 7.3 in a typical inference cycle, such as the one used by Lorenz et al. (2009). Implementation of our design principle does not require a model of the system, merely a way to predict the responses of considered perturbations based on previous data, so the classical cycle is not needed for generation of informative data. It is however needed to check modelling objectives.

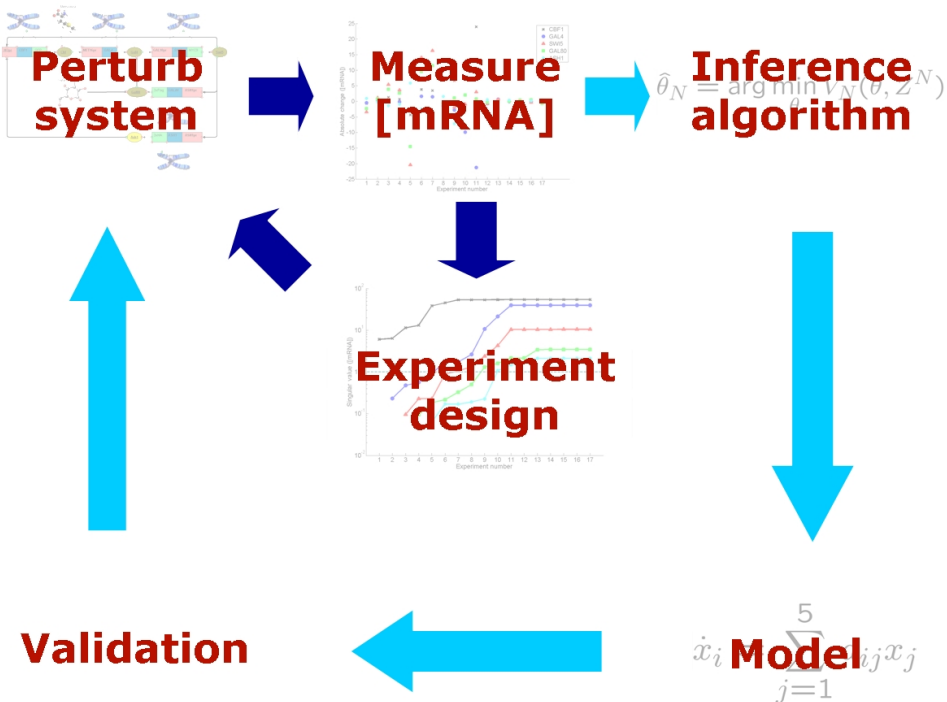


Figure 7.3: Illustration of the proposed inner design loop of the classical model construction cycle, exemplified by inference of a GRN. In this case the proposed inner loop consists of three steps: perturb system, measure [mRNA], and experiment design.

## 7.6 SVD design of steady-state experiments

To illustrate the use of the proposed principle for iterative design of perturbations we propose the following algorithm for designing steady-state experiments and use it to generate data that practically spans the gene space of the Snf1 signalling pathway studied by [Lorenz et al. \(2009\)](#).

**Algorithm 1.** *SVD design for practical spanning of the state space.*

First, perform as many different perturbation experiments as the number of variables/nodes in the network  $n$  and collect the observed responses as columns in the matrix  $\mathbf{Y}[t] \in \mathbb{R}^{n \times m}$ . The set of perturbations should be selected such that the columns of  $\mathbf{P}[t] \in \mathbb{R}^{n \times m}$  are linearly independent. Then, while  $\sigma_n(\Phi(\chi)) < 1$ , i.e. while the confidence score defined in (7.3.6) is smaller than one, use the four steps below to design a new perturbation based on all previous responses.

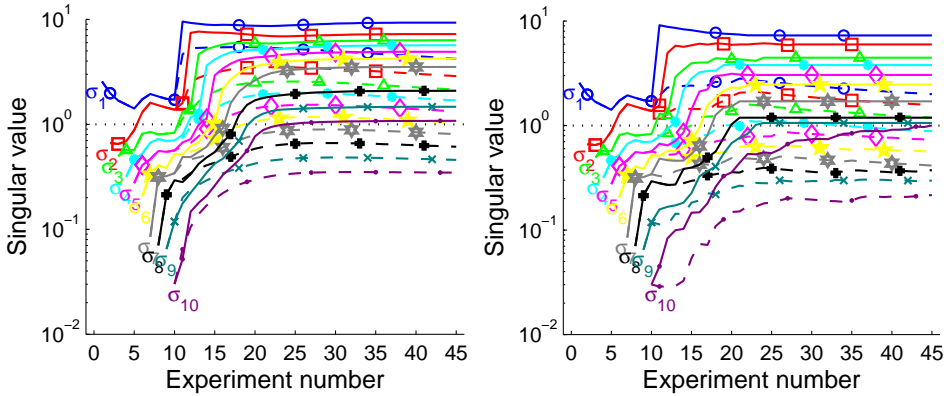
1. Perform an SVD of the scaled regressor matrix  $\Phi(\chi)$  defined in (7.3.6) with  $\Phi = \mathbf{Y}[t]^T$  and  $m = t$  using all previous responses. Introduce the set of indices corresponding to singular values  $\sigma_k$  of  $\Phi(\chi)$  below one:  $\mathcal{W} = \{k | \sigma_k(\Phi(\chi)) < 1\}$ .
2. Perform an SVD of the matrix with all previous responses  $\mathbf{U}_{\mathbf{Y}} \Sigma_{\mathbf{Y}} \mathbf{V}_{\mathbf{Y}}^T = \mathbf{Y}[t]$ .
3. Construct the next perturbation as a sum of linear combinations of previous perturbations with a magnitude compensating for the uncertainty level

$$\mathbf{p}_{t+1} = \mathbf{P}[t] \sum_{k \in \mathcal{W}} \frac{1}{\sigma_k(\Phi(\chi))} \mathbf{v}_k(\mathbf{Y}), \quad (7.6.1)$$

and add it to the perturbations  $\mathbf{P}[t+1] = [\mathbf{P}[t] \ \mathbf{p}_{t+1}]$ . Here  $\sigma_k(\Phi(\chi))$  denotes the  $k$ th singular value of  $\Phi(\chi)$ ,  $\mathbf{P}[t]$  the matrix with all previous perturbations, and  $\mathbf{v}_k(\mathbf{Y})$  the  $k$ th column of  $\mathbf{V}_{\mathbf{Y}}$ , which gives the combination.

4. Perform the experiment and add the obtained response to the matrix of responses  $\mathbf{Y}[t+1] = [\mathbf{Y}[t] \ \mathbf{y}_{t+1}]$ .

This design uses previous perturbations and responses to calculate a perturbation that is expected to practically span a new dimension of the state space. This is accomplished in (7.6.1) by summing the perturbation corresponding to each singular value after scaling by the inverse of the singular value. The singular values are therefore lifted one-by-one until they are all above one in accordance with Conjecture 7.5.1, as demonstrated on the Snf1 example in Figure 7.4. Note, however, that this algorithm does not account for experimental limitations nor nonlinearities that place a limit on how strong the response may be while maintaining the validity of the linear model. Likewise, no optimality properties are pursued, but this algorithm is simple, explores the gains of the system, counteracts the intrinsic signal attenuation, and ensures practical spanning of the state space. This algorithm is based purely



**Figure 7.4: Comparison of the iterative SVD design versus random perturbations.** The mean (left) and minimum (right) of each singular value of the scaled response matrix  $\Phi(\chi)$  defined in (7.3.6) as a function of the number of experiments based on 1000 Monte Carlo simulations starting from the *in vivo* data of Lorenz et al. (2009). The singular values (solid lines) are by Algorithm 1 lifted one-by-one to the threshold (dotted line), while the random perturbation strategy (dashed lines) fails to lift all the singular values till one.

on basic vector operations and the singular value decomposition (see *e.g.* Horn and Johnson, 1990, p. 414-415), so it is computationally as good as the best SVD implementation and can handle networks with tens of thousands of variables. More importantly, no model of the network or system is estimated in any step, implying that no *a priori* knowledge is needed, nor does the result in any way depend on inference or parameter estimation. The algorithm is, however, implicitly based on the assumption that a linear mapping exists from perturbations to responses. On the other hand, if *a priori* knowledge or a model is available, then an additional step can be added where it is used to simulate the constructed perturbation and based on this modify the perturbation.

Lorenz et al. (2009) performed as many experiments as the number of genes in their network, *i.e.*  $m = n = 10$ , using orthogonal perturbations where the genes are overexpressed one-by-one. So the prerequisite to start iterating over the four steps is fulfilled. We do not have experimental facilities, so we perform *in silico* experiments with the alternative network model in Table C.4, which is depicted in Figure 7.1 (left), as the “true” network, to demonstrate the effectiveness of Algorithm 1. We add independent and identical normally distributed measurement noise with the standard deviation reported by Lorenz *et al.* for each perturbation to the designed perturbations before simulating the response of each experiment; then we add to each simulated response independent and identical normally distributed measurement noise with standard deviation given by the mean value obtained by

Lorenz *et al.* for each gene. This ensures that the simulated data has the same error level in both the input and output as observed in real *in vivo* data. For details on the simulation see Section C.3. The SVD design lifts the smallest singular value to one, which ensures practical independence of the regressor matrix at the chosen significance level  $\alpha = 0.05$  for the specified errors, in all 1000 Monte Carlo simulations of 35 additional steady-state experiments. The confidence score was hence improved by a factor 33 from being a mere 0.03 in the original data and the condition number of the response matrix  $\mathbf{Y}$  decreased from 215 to 8 in all simulations, with a mean of 4. The mean number of additional experiments to reach a confidence score above one was 15. In other words, the quality and information content of the data was increased dramatically and practical spanning of the gene space achieved. For comparison we also performed random perturbation experiments with the same output energy, which in the best case only improved the confidence score by a factor 20 to 0.6 in 35 additional experiments. On average the random experiments improved it by a factor 11 and the condition number of  $\mathbf{Y}$  decreased merely to 66. The confidence score was in the best case still below one even after 100 experiments, so we warn against the use of random perturbations in inference of GRNs. The mean and minimum value of all singular values of the scaled regressor matrix over the 1000 Monte Carlo simulations are shown in Figure 7.4. This can be compared to five repetitions of the 10 experiments of Lorenz *et al.*, which barely improves the confidence score from 0.0301 to 0.0319. After 5000 repetitions, *i.e.* 50000 experiments, the confidence score would still only be 0.0335, so we therefore also warn against pure repetition of perturbation of the genes one-by-one. Both the random and one-by-one designs typically fail to generate perturbations that counteract the intrinsic signal attenuation of the system, while the proposed SVD design succeeds and therefore gives such a large improvement in data quality. Note that the data obtained by Algorithm 1 is informative enough for practical spanning of the state space, but in general not for robust inference, since only the first condition in Conjecture 7.5.2 is fulfilled. A modified version operating on  $\Psi_{ki}(\chi)$  instead of  $\Phi(\chi)$  is needed to fulfil the second condition and guarantee that the data is informative enough for robust inference.

In order to apply the proposed algorithm in practice one must be able to scale and combine the old perturbations into one new correlated multi-gene perturbation that counteracts the intrinsic signal attenuation of the system. This is illustrated well by the fifth perturbation designed in one of the experiment series  $\mathbf{p}_{11} = [2.1, -1.5, 36, 16, -1.7, 4.6, 0.66, -1, -0.86, 0.56]^T$  in which every gene is perturbed. Note that we use the fifth perturbation for illustration, because it is in the middle of the designed experiments needed in all 1000 cases. Expressed as a combination of the earlier perturbations used by Lorenz *et al.* this corresponds to  $[0.97, -0.69, 11, 7.3, -0.78, 2.2, 0.31, -0.7, -0.27, 0.23]$ , *i.e.* 0.97 times the first perturbation, a repression of 0.69 times the second instead of their activation and so forth. To combine and scale previous perturbations is challenging from an experimental point of view but worth pursuing, since one otherwise must compensate by performing many additional experiments. Constraints of the experimental

protocols could and should of course be incorporated in any design algorithm before the designed perturbations are applied *in vivo*, but it will complicate the algorithm and we here prioritized a clear and simple presentation. If it is *e.g.* only possible to perturb two genes in the same experiment, then the algorithm should be restricted to combinations of two perturbations. The more severe a restriction, the larger the number of experiments that will be needed to obtain informative data, but the data quality will improve as long as one is able to counteract the intrinsic signal attenuation.

## 7.7 Conclusions

We started by showing that a published gene expression data set used for inference of a Snf1 related gene regulatory network by [Lorenz et al. \(2009\)](#) lacks necessary information and ended by demonstrating how to iteratively design additional perturbation experiments for robust inference. To do this we established sufficient conditions for robust variable selection and network inference, the need for an iterative design that explores the multivariable gains of the underlying system, and the fundamental principle of designing the next perturbation such that the expected response practically spans an additional dimension. We focused on illustrating the principle and delivering the core message and have therefore left incorporation of experimental constraints and other details needed to use the principle in practice to later works. Data obtained by applying the principle is suitable both for hypothesis generation and testing in general, even though it in general is not optimal for discriminating among a set of models, because no objective function is specified and minimized. We conclude with a call for new experimental protocols and a discussion on the importance of analysing the data quality in systems biology.

### 7.7.1 Call for multi-gene perturbation techniques

[Lorenz et al. \(2009\)](#) used network inference based on gene expression data to generate physical interaction hypotheses, that they subsequently tested using promoter-reporter gene fusions in gene-deletion mutants and chromatin immunoprecipitation experiments. We want to go one step further and generate informative expression data for robust network inference, *i.e.* determine the interactions directly from the data with desired confidence, such that testing by other techniques is superfluous. Their work lacks three things to meet our goal: a method for robust inference, which we developed in Chapter 5, an iterative design that generates informative data, which we have presented here, and an experimental protocol for realizing multi-gene perturbations, *i.e.* scaled combinations of previous perturbations. The latter is needed to implement the precise correlated perturbations required to counteract the intrinsic signal attenuation of GRNs. We therefore call for development of multi-gene perturbation techniques, which as far as we can see is the last thing missing before the power of *in vivo* data based robust network inference can be proven experimentally.

### 7.7.2 Importance of data quality and analysis

This work not only illustrates the importance of having informative data but also the importance of analyzing the information content of a data set with regard to the purpose of the model. The sum of squared residuals of the final network inferred by Lorenz *et al.* (2009, supplemental Table S9) is in agreement with the expectation based on the reported standard errors (Table S3) and it only includes interactions that differs significantly from zero based on a t-test. In other words, the classical statistical estimation and analysis show that the inferred network is a good model that explains the recorded input-output data well. However, many other networks with different structure do it equally well, due to the poor quality of the data, as we have shown here. In particular, collinearity is problematic and in no way accounted for by the statistical analysis of Lorenz *et al.* We therefore recommend analysis of the information content of data, in particular analysis of the dimensions practically spanned by data. In general, we concur with Jaqaman and Danuser (2006) when they state that the systems biology community should require a minimum set of tests on models before they are published.

When two regressors are nearly collinear, small errors will cause large correlated changes in the values of both parameters and the parameter confidence region will be ellipsoidal. An ellipsoidal confidence region implies that changes to certain linear combinations of parameters will have a much smaller effect on the fit to data than changes to other combinations, *i.e.* the parameters are sloppy according to Brown *et al.* (2004). Gutenkunst *et al.* (2007), who analysed 17 previously published nonlinear models of biological systems and found that the parameter spectrum in all of them was sloppy. They observe that the possible values of individual parameters typically span several orders of magnitude, which implies that the predicted response of the model to the perturbations is robust to changes in parameter values. Robustness of biological systems to parameter changes has also been noted by *e.g.* Barkai and Leibler (1997); von Dassow *et al.* (2000); Prinz *et al.* (2004); Piazza *et al.* (2008). The observation that good data prediction is achieved even though the parameter uncertainty is large led Gutenkunst *et al.* to conclude that computational modelling should focus on prediction and not parameter estimation. The fitted parameters are highly dependent on the model and data, so they cannot be used separately in another model anyway. We offer a different conclusion, namely poor data. The sloppy parameters show that the data is not informative for estimation of them. One can say that the model has more degrees of freedom than the data practically spans and the good prediction merely reflects the good fit to the poor data. In general, any model that fits well to data will predict similar experiments well. The concept of core predictions, introduced by Cedersund and Roll (2009), therefore reflects the information content of the data. If the model has a purpose beyond data prediction, then the experiments should be designed such that they are informative for the purpose of the model. This often forgotten fact, is here illustrated by our purpose–inference of direct causal influences between genes—which requires correlated multi-gene perturbations that counteracts

the intrinsic signal attenuation of the system. We want to, in particular, highlight the difference between data prediction and mechanistic understanding. Both network models in Figure 7.1 are good for data prediction but none of them provide any mechanistic understanding, since their very existence implies that the experimental data does not contain necessary information to uniquely determine the structure of the network. Considering that analysis of the informativeness of data is rare and current data sets are characterized by few data points compared to the high number of genes and possible interactions, large measurement uncertainty both in the perturbations and responses, and redundant nearly collinear variables (Holter et al., 2000; Alter et al., 2000; Tegnér and Björkegren, 2007; Faith et al., 2007; Cosgrove et al., 2010; Nordling and Jacobsen, 2009a; Wu and Wu, 2010), we predict that the literature contains many falsely identified networks that both contain non-existing interactions and lack existing ones.

We believe that lack of proper data analysis caused Krishnan et al. (2007) to wrongly conclude that reverse engineering of GRNs from expression data is an indeterminate problem. The network inference case presented here fulfils the conditions for reconstruction of networks established by Goncalves and Warnick (2008), which are sufficient albeit not necessary, see Section 6.5, so GRNs can be reverse engineered from steady-state expression data generated by transcriptional perturbations. Moreover, the same approach can and should be used to infer all types of biological networks, including metabolic and protein-protein interactions.



---

## Summary and future work

---

*“L’avenir, tu n’as point à le prévoir mais à le permettre.”*

Antoine Marie Jean-Baptiste Roger de Saint-Exupéry,  
Citadelle, 1948.

### 8.1 Our contribution towards solving the network inference puzzle

Inside each cell thousands of different molecules interact. Together they form a dynamic system that integrates both genomic information and environmental cues (Wolkenhauer et al., 2005b; Sontag, 2005; Csete and Doyle, 2002). Depending on the level of abstraction, different graphical network representations are commonly used to describe this system; *e.g.* gene regulatory networks (GRNs) capture the regulatory interactions resulting from changes in gene expression (He et al., 2009; Hecker et al., 2009; Karlebach and Shamir, 2008; Christensen et al., 2007). The structure of these networks reveals the feedback loops, feedforward loops and cascades that are fundamental for the normal function and behaviour of the system. The system and its network representation are altered by disease, *e.g.* cancer arises from multiple spontaneous and/or inherited mutations of the DNA that affects the control of cell growth and division (Hood et al., 2004; Roukos, 2010a). The molecular interactions therefore need to be studied from a systems perspective to understand how genetic information and environmental cues give rise to function, and to enable development of predictive, personalized, preventive, and participatory medicine (Barabási et al., 2011; Hood and Friend, 2011; Roukos, 2010b; Wolkenhauer et al., 2009; Ideker and Sharan, 2008). Most genes are regulated by a small number of other genes, implying that GRNs are sparse, *i.e.* only a subset of all possible links exists (McAdams and Arkin, 1998; Thieffry et al., 1998; Arnone and Davidson, 1997; Jeong et al., 2000, 2001; Lima-Mendez and van Helden, 2009). Network inference offers the ability to reverse engineer sparse network models of the type required for these studies and provide new biological knowledge based on *in vitro*

or *in vivo* measurements of changes in gene expression, copy numbers, metabolite concentrations, protein abundance and phosphorylation, *etc.* (Tegnér and Björkegren, 2007; Cho et al., 2007; Crampin, 2006; Gardner and Faith, 2005; Goncalves and Warnick, 2008). Inference of gene regulatory networks has the potential to reveal thousands of regulatory interactions from a series of perturbation experiments that could efficiently complement the knowledge acquired so far from traditional experimental approaches.

Network inference has been intensively studied within the past decade, but several theoretical and practical problems remain (Hendrickx et al., 2011; He et al., 2009; Hecker et al., 2009). Network inference is today mainly a method for filtering or data mining; *e.g.* selection of a subset of interactions or molecules, or construction of a predictive model, a valuable use, as recently demonstrated in *e.g.* Cacciottolo et al. (2011); Julià et al. (2007); Jörnsten et al. (2011). Nonetheless, this is far from the potential and envisioned future as a field of research (He et al., 2009; Bonneau, 2008; Styczynski and Stephanopoulos, 2005; Cassman et al., 2005; Filkov, 2005). Despite the promising performance shown in the original publication of most inference algorithms, few algorithms are actually used, except in comparisons of algorithms. In a recent evaluation of six published reverse engineering algorithms, Hache et al. (2009) conclude that the performance of the tested methods is not good enough for inference of large GRNs in practice. The set of algorithms, ARACNe (Basso et al., 2005), ParCorA (de la Fuente et al., 2004), GNRevealer (Hache et al., 2007), Banjo (Yu et al., 2004), LDST (Rangel et al., 2004), GeneNet (Schäfer and Strimmer, 2005), is only a small subset of all the inference algorithms that exist, but it represents several of the major model formalisms used to describe biological systems (Hecker et al., 2009; Karlebach and Shamir, 2008; de Jong, 2002). Better inference methods are clearly needed, but as shown in this thesis, it is not enough to realize the potential of network inference. We focused on inference of biologically relevant causal interactions in subsystems responsible for some function or behaviour and asked why previous studies failed to deliver on their promise, as well as, how to reach the potential.

To model the subsystem that generates a function of interest, three basic problems must be solved: First, one needs to determine the components of the system that are essential for generation of the function, *i.e.* identify the nodes of the subnetwork behind the function. Second, one needs to design experiments that yield informative data for network inference. Third, one needs to infer a model with the correct structure of the network based on the recorded data. Aspects of the first and third problem have been widely studied, albeit not solved to our satisfaction, while the second one has largely been neglected. For further details, see Chapter 1 and the review articles Hecker et al. (2009); Karlebach and Shamir (2008); Bonneau (2008); Tegnér and Björkegren (2007); Styczynski and Stephanopoulos (2005); Filkov (2005); D'haeseleer et al. (2000). We have studied four issues related to the second and third problem: properties of biological systems, robust network inference, inference of subnetworks, and design of perturbation experiments. These issues turned out to be essential for explaining why network inference has not delivered and how to fulfil

its potential. Below we summarize our results and put them in a wider perspective.

Existing data sets for reverse engineering of GRNs have the following characteristics (see *e.g.* Holter et al., 2000; Alter et al., 2000; Tegnér and Björkegren, 2007; Faith et al., 2007; Cosgrove et al., 2010; Wu and Wu, 2010). The data points are few compared to the number of genes and possible interactions, *i.e.* the dimensionality is high in terms of variables but low in samples. The measurement uncertainty is large both in perturbations and responses, *i.e.* the signal to noise ratio is low and the inference problem is an errors-in-variables problem. The variables are nearly collinear, *i.e.* the response matrices are ill-conditioned. Previous works have shown that most of the variation in gene expression can be explained by a few “characteristic modes” or “eigengenes”, *i.e.* linear combinations of variables, and stressed their importance, but neglected the equally important weak modes (Holter et al., 2000; Alter et al., 2000; Kuruvilla et al., 2002; Wu and Dewey, 2006). We postulated in Chapter 3 that these are reflections of a generic property of biological systems, that we call interampattiness. An interampatte system is characterised by strong INTERactions enabling simultaneous AMPLification and ATTEnuation of different signals, which leads to ill-conditioned data with both characteristic and weak modes. We showed that existence of multiple time-scales and feedback loops increase the degree of interampattiness, which has strong implications for the dynamics and reverse engineering of the network. The change in gene expression in response to external perturbations is highly correlated, even in the absence of common transcription factors, implying that interampatte GRNs erroneously may be assumed to have co-expressed/co-regulated genes. The weak modes enable data compression and reduction of the system dimensionality, using clustering, singular value decomposition, principal component analysis, subspace identification, *etc.*, but then essential information for reconstruction of the underlying network is lost. To illustrate how weak modes hampers network inference, we showed an example of network models with different structure that equally well explains ill-conditioned data from an interampatte GRN. Existing ill-conditioned data sets are in other words not informative enough for network inference, which is one reason for why network inference has failed to deliver. To overcome this obstacle perturbation experiments that counteract the intrinsic signal attenuation are needed, and we elaborated on this in Chapter 7.

Traditional estimation techniques are not suitable for data with mentioned characteristics and a variety of inference algorithms have therefore been developed and tested during the last decade (He et al., 2009; Hecker et al., 2009). In particular, inference based on data sets with fewer data points than variables, corresponding to solution of an underdetermined problem, has been investigated. Many authors claim to have demonstrated how to infer regulatory interactions from gene expression data, but their methods have two essential shortcomings that we brought forward in Chapter 4. Firstly, to yield sparse networks, with only a subset of all possible links, the proposed methods rely on regularization, model selection, or thresholding techniques. As a consequence of the employed techniques, the inferred models are largely determined by algorithm specific details and measurement noise, especially,

for data sets that do not contain enough information for network inference. This partly explains why benchmarking studies have shown that inference of GRNs usually results in a large fraction of false positives and negatives, even among the links deemed most significant (Marbach et al., 2010; Stolovitzky et al., 2009). Secondly, employed measures of confidence of inferred interactions/links are misleading, in particular for data sets with fewer observations than variables. We proved that no reliable measure of confidence exists when the data set contains fewer observations than variables. By showing that many of the inferred interactions deemed to be the most significant by Lorenz et al. (2009) are not needed to explain their data, we demonstrated that the employed t-test is misleading, even though the data set contains as many observations as variables. To infer interactions with confidence from expression data, at least as many samples as genes of interest must be collected. Considering the rapid development of techniques for measuring changes in gene expression, we anticipate that collection of sufficient numbers of data points for inference of subnetworks responsible for a function of interest will become mainstream in the near future. Instead of following the direction of previous works and focusing on the inference of the most likely network in some sense by solving an underdetermined problem, we focused on how to find the interactions that can be shown to exist with desired confidence. For this, reliable measures of confidence must be developed, which we address next.

The desire to find the subset of all possible interactions that exists *in vitro* or *in vivo* based on observations of their effect makes network inference a variable selection problem. Variable selection is also known as feature selection, feature reduction, attribute selection, subset selection, or model structure selection (Guyon and Elisseeff, 2003; Fan and Lv, 2010; Hara and Sillanp, 2009; George, 2000; Hong et al., 2008; Stoica and Selen, 2004). Despite at least half a century of active research, variable/model selection is today recognized as an open and hard problem. Selection by brute force is proven to be NP-hard (Yusta, 2009; Cotta et al., 2004; Kohavi, 1995; Cover and Van Campenhout, 1977) and previous studies have shown that a small prediction error or residual does not, in general, imply selection of the correct subset (Fu and Desmarais, 2010; Guyon and Elisseeff, 2003; Kohavi and John, 1997; Lukacs et al., 2009; Freedman, 1983). For further information see Sections 1.2 and 5.1. This partly explains why it is difficult to solve the network inference problem and implies that one to infer interactions with confidence needs to select variables with confidence. We therefore in Chapter 5 developed a theory of variable selection and network inference with confidence, which we termed robust variable selection and robust network inference, respectively. Please note that the term robust in the literature in general is used in different ways, see *e.g.* Aelst et al. (2008), while we strictly use it to mark that a property holds for all models in a set—in accordance with the tradition in robust control. It is essential to distinguish between robust and non-robust variable selection, since they are conceptually different. The objective of the former is to determine the variables that must be included in or always can be excluded from models explaining data at a desired confidence level, while selection of the variables that optimize some quantity, *e.g.* maximize the likelihood of the

observed data or minimize the prediction or classification error, typically is the objective of the latter. To the best of our knowledge, this is the first theory of robust selection and inference, as well as the first one that can accommodate the combination of low signal to noise ratios, errors-in-variables, near collinearity, and few data points—seen in current gene expression data sets.

By considering all possible models of a specified class that can explain data at a desired significance/confidence level for the associated error model, we in Chapter 5 showed that variables and interactions, loosely speaking, can be classified as either present/existing, absent/non-existing, non-evidential, or alternative. The data cannot be explained without the variables and links that are present in all of these models, so they are, under mild assumptions, true positives. We thus achieve truly robust selection and inference by only classifying them as present/existing. We find these variables and links without estimating a single parameter or fitting any model to the data by reformulating the variable selection problem as a rank problem. This decouples the variable selection problem from parameter estimation, which ensures that our selection only depends on the observed data and specified class of models. All previous methods for variable/model selection that we know are based on comparison of different model structures by fitting models to data, either explicitly, as in likelihood ratio and Akaike information tests, or implicitly, as in regularization (see *e.g.* Cedersund and Roll, 2009; Casella and Berger, 2001; Stoica and Selen, 2004; Akaike, 1973). The decoupling removes all interdependencies between selection and estimation that, in general, weaken the conclusions that can be drawn. Our variable selection method therefore offers new opportunities that need to be investigated beyond inference of biological networks. Philosophically, a model structure must be selected before any parameter can be estimated, and any property of the parameters in general only hold when the selected model structure is assumed correct (Burnham and Anderson, 2002, p. 14). The t-test, used by Lorenz et al. (2009) to calculate p-values of inferred links, is based on the ratio between parameter value and variance, so this explains why it fails to account for models with alternative structure, as mentioned above. In Section 5.10, we proved that the structured singular value method, which is a classical tool in robust control (see *e.g.* Zhou and Doyle, 1998, p. 189), in general can be used to solve the rank problem and enables us to assign confidence to individual interactions in a reliable way. This is in Section 4.6 demonstrated by inferring the synthetic IRMA GRN, that recently was engineered in yeast for benchmarking of inference methods (Cantone et al., 2009), from published *in vivo* data. The resulting model proves, under mild assumptions, that a previously unknown activation of transcription of *SWI5* by *CBF1* exists in yeast. This illustrates that robust inference reveals novel biological knowledge and contributes towards fulfilling the potential of inference.

Although the variables and interactions that are classified as present/existing are arguably most interesting, in particular in any setting where a small subset out of possibly thousands of variables are sought, the other three classes—absent/non-existing, non-evidential, and alternative—also provide valuable knowledge. Real measurement data is always corrupted by errors, noise, or uncertainty and a set of models can in

general explain the data, *i.e.* these models cannot be rejected at a desired significance level. The variables and interactions classified as present/existing must be included in the models to explain data and are thus, under mild assumptions, true positives. Their counterpart, the variables and interactions classified as absent/non-existing can always be excluded from the models without losing the ability to explain data and are thus, under mild assumptions, true negatives. Variables and interactions classified as non-evidential can be included in and excluded from every model without affecting its ability to explain data, so the data contains no information about them. Based on the principle of parsimony, they should always be excluded, but it is important to know that they are excluded based on lack of information, contrary to absent/non-existing ones that are excluded based on information. Any variable and interaction that cannot be proven to belong to one of the former three classes is called alternative. Strictly speaking, a variable and interaction is only classified as alternative if it is needed in some model to explain data and can be excluded in some other model. Additional data is needed to determine which of the non-evidential and alternative variables and links that exist. To enable this classification, in Chapter 5 we defined new concepts with necessary and/or sufficient conditions for each case. These concepts are used in rank based tests that only require the observed data, but implicitly check all models within a chosen class that cannot be rejected based on the assumed error model and observed data. In addition to proving that the structured singular value in general can be used to test if an interaction exists, as mentioned above, we prove that the singular value decomposition provides sufficient and necessary conditions for data with normally distributed errors. This implies that existing numerically stable and computationally efficient algorithms for singular value decomposition enable robust selection and inference even for cases with thousands of variables. In conclusion, we have proposed a solution to the core problem of how to infer regulatory interactions with confidence from expression data that is informative enough.

Our classification of all variables and interactions does not provide a model useful for prediction *per se*, which regularization based algorithms such as LASSO does (see *e.g.* Tibshirani, 1996; Friedman et al., 2010). However, once the classification is known then a predictive model can in general be obtained by solving an optimization problem with constraints on the structure. Moreover, a statistically efficient estimator can be used if the data is informative enough, because the correct number of degrees of freedom and parameters with value zero are then given by the classification. If the data is not informative enough, then some variables and interactions are classified as alternative and one should typically include some of these in the model to decrease the prediction error. Previously published algorithms could be used to select among the alternative variables, because typically, one of the objectives is the selection of the variables that optimize some quantity, *e.g.* minimize the prediction or classification error (see *e.g.* Dash and Liu, 1997; Peng et al., 2005; Molina et al., 2002; Guyon and Elisseeff, 2003; Fan and Lv, 2010; Hara and Sillanp, 2009; George, 2000). Thus, they should also be used if a predictive model or the most likely model in some sense is sought. By aiming to find the best subset of

variables and links in some sense, non-robust methods generate hypotheses and should be used for exploratory analysis. Our robust methods, on the other hand, perform testing of implicitly generated hypotheses, which generates knowledge. The robust methods developed here therefore offers automated discovery, hypothesis generation and testing of the type sought in systems biology (see *e.g.* Ideker et al., 2001). Regularization methods have been shown to select the correct subset, under certain conditions, even for underdetermined problems, *i.e.* when fewer data points than variables are available (Candès and Plan, 2009; Fan and Lv, 2010; Candès and Wakin, 2008; Filkov, 2005). At least as many data points as variables are needed to assign confidence to individual variables and links, so in general our robust methods require more data, but this is the price of knowledge. It is one thing to calculate the answer correctly and another to know that the answer is correct.

When inferring GRNs it is common practice to only observe the change in mRNA abundance, neglecting all proteins, metabolites, and other molecules, and only perturb selected genes or quantities with unknown effect on the observed genes (Hecker et al., 2009; Tegnér and Björkegren, 2007; Brazhnik et al., 2002). Clearly the neglected components also affect the state of the cell. In Chapter 6 we, therefore, analysed how these unobserved and unperturbed latent state variables affect the requirements of the set-up of the perturbation experiments for inference. Necessary and sufficient conditions for reconstruction of linear time invariant systems from knowledge of the input-output behaviour have previously been established by Goncalves and Warnick (2008). However, as we showed in Section 6.4, counter examples exist that show that their conditions are not necessary, unless hidden states are considered that have no effect on the input-output behaviour. We therefore performed a similar analysis and showed that it is necessary to know a linearly independent subset of as many relations between the states of interest and responses or perturbations as there are states of interest. Otherwise the states of interest, which constitute nodes in the subnetwork to be inferred, are not uniquely defined, and consequently not the structure of the subnetwork either. In addition, a minimal realization with the states of interest needs to exist, which is observable and controllable per definition. Loosely speaking, each state of interest should be affected by a perturbation and should trigger a response. These requirements are when the set of states of interest contains all states of the system necessary and sufficient for inference of the full network. This implies that the fundamental information requirements for inference of GRNs scale linearly with the number of genes of interest. In steady-state experiments, it is necessary to independently both observe and perturb each state of interest, which implies that at least as many experiments as variables of interest are needed. The previously discussed conditions for robust inference can therefore be met by focusing on subnetworks with fewer genes of interest when then number of experiments cannot be increased. When necessary, some observed genes, whose response is linearly dependent on the set of genes of interest and therefore do not provide any information useful for robust inference, should be neglected. We also showed in Section 6.6 that bounds on the number of latent states and different paths between the states of interest in principle can be

determined from time-series data. However, the physical quantities corresponding to the latent states cannot be determined from perturbation experiments alone, unless they are directly perturbed or observed. Even though we cannot infer the links between latent states, we at least obtain information about how many intermediary latent states to look for if needed. The existence of latent states implies that in general a subnetwork of pseudo-direct causal influences, accounting for all environmental effects, is inferred. This is essential for interpretation of inferred networks and planning of validation experiments, because indirect links do not correspond to physical binding between two proteins. In conclusion, by focusing on subnetworks the required number of samples for robust inference can be achieved in practice.

Experimental protocols for perturbation of individual genes are today standard in inference of GRNs, see *e.g.* Gardner et al. (2003); Lorenz et al. (2009); Cantone et al. (2009). They however typically provide poor excitation of interampatte systems and ill-conditioned data, as discussed above, which in part explains the large number of false positives and negatives that has been documented (Stolovitzky et al., 2009; De Smet and Marchal, 2010; Fujita et al., 2009). This is illustrated in Chapter 7 through analysis of a reverse engineering study of the Snf1 signalling pathway in yeast (Lorenz et al., 2009). The data lacks necessary information for inference even though the authors perturbed all 10 genes one-by-one in separate experiments. From a control theory perspective, an interampatte system operates in closed loop and one either needs to open up the unknown feedback loops or counteract the intrinsic signal attenuation to get informative data. Even if one could open the loops, it would alter the very part of the system that one wants to identify, so the only foreseeable remedy is iterative design of perturbation experiments that counteract the signal attenuation. Design of experiments has been studied for a long time, but the focus of the branch considering model selection has been on discrimination among a small set of specified models (Michalik et al., 2010; Schwaab et al., 2008; Pronzato, 2008; Fisher, 1935). These methods require *a priori* information that typically is not available in reverse engineering of GRNs, see Sections 1.3 and 7.1 for details. Therefore, in Chapter 7 we proposed a principle—select the next perturbation such that the expected response practically spans an additional dimension—that can be applied without having any model of the system. It is based on the sufficient conditions on data for robust variable selection and network inference that we established in Section 7.3. We numerically demonstrated this principle by designing additional steady-state experiments for inference of the Snf1 signalling pathway. On average 15 and in the worst case 35 additional experiments were needed to increase the strength of the weakest signal 33 times and obtain informative data in all 1000 Monte Carlo simulations that we performed. On the contrary, designs using only random or single gene perturbations yielded poor data in all simulations even after 100 additional experiments. Large improvements in data quality are hence obtainable by replacing the random and single gene designs used today. The lack of informative data that currently hampers network inference can, in other words, be solved through iterative design of correlated multi-gene perturbation experiments. Ability to combine and scale previous perturbations is essential for the iterative

experiment design and we therefore call for development of improved experimental protocols.

Earlier works have demonstrated how to perturb individual genes experimentally and how to measure the response with acceptable accuracy, but failed to infer the correct network due to poor excitation of the system and lack of a robust inference method. In this thesis, we have demonstrated how to design informative perturbation experiments, how to robustly infer the underlying network from data, and how to assign confidence to individual interactions.

## 8.2 Suggested work to realize the potential of network inference

As discussed above, the central parts for robust inference of direct causal interactions existing between genes are now in place. Techniques for perturbing individual genes and measuring the resulting expression changes exist (see *e.g.* Gardner et al., 2003; Lorenz et al., 2009; Cantone et al., 2009). In Chapter 6 we established what to perturb and measure to obtain a model of a subsystem and how to interpret it. The theory and tools needed for robust inference of existing interactions were established in Chapter 5 and demonstrated on a published *S. cerevisiae* data set in Chapter 4. Publicly available data sets, however, suffer from a lack of information, so additional data is needed to demonstrate the power of robust inference and automated knowledge discovery that it enables. The strategy for iterative design of perturbation experiments proposed in Chapter 7, which counteracts the intrinsic signal attenuation reported in Chapter 3, is needed. However, we believe that it is experimentally challenging to scale and combine previous perturbations, as required for iterative design, so new experimental protocols for multi-gene perturbations need to be developed and tested. The future of this technique depends on a successful demonstration based on experimental data, which therefore should have the highest priority in future research. These techniques should then be used to generate much sought mechanistic knowledge about, in particular, intracellular regulation and dis-regulation that is needed to understand complex diseases like cancer.

In Chapter 5, we developed the theoretical basis for robust classification of variables that are not needed to explain data for different reasons by introducing practical excludability and negligibility, as well as sufficient and necessary conditions for them, but we left the development of computational tools for future work. In particular, a method for testing of practical collinearity that can be used to prove that a variable is practically excludable, *i.e.* always can be excluded while explaining data, should be developed as soon as possible. Practically excludable variables are the counterpart of practically selectable variables, *i.e.* the ones that must be selected to explain data. Conditions and methods for determining the sign of an interaction, which correspond to our concept of practical assignability of vectors, also need to be developed. Statistical methods for analysis of the information content of data sets with regard to both robust and non-robust network inference are needed. Also

the connection between existing non-robust inference algorithms and our robust techniques, both in terms of required information, resulting network estimate, and computational cost is of interest, already for the following reason. Existing inference algorithms that do not recover existing interactions, corresponding to practically selectable variables, do not utilize the information of the data set properly and should therefore in general not be used at all.

Future research on design of experiments should in general focus on design of perturbations that counteract the intrinsic signal attenuation of the system under study, since, according to our analysis in Chapter 3, this is the key to obtaining informative data for network inference. Our focus in Chapter 7 was on the fundamental principle for iterative design of informative experiments for robust inference and we left the derivation of optimal algorithms to others. We therefore hereby challenge other computational biologists to propose new algorithms for design of perturbation experiments that minimize the number of experiments while meeting both experimental and quality constraints. In particular, practical design algorithms for a mixture of steady-state and time-series experiments, incorporating experimental constraints and accounting for uncertainty, should be developed. The effect of interampattiness and signal attenuation on estimation of parameters in nonlinear models should also be analysed in future works.

Robust variable selection for network  
inference in depth

---

Additional information that supports Chapter 4 and 5 is presented in this appendix. The effective rank, which we used in Nordling and Jacobsen (2011), and its shortcomings are discussed in Section A.1. Common ways of graphically illustrating parameter estimation and linear programming is discussed in relation to our illustrations of variable selection in Section A.2. Additional theorems that are necessary to prove our claims in Chapter 4 and 5 are presented with proofs in Section A.3. Generation of the *in silico* data for the IRMA example used in Section 4.2 is described in Section A.4. The error model of the IRMA data that we used for robust inference of the IRMA network in Sections 4.6 and A.6 is derived in Section A.5. Details on how we robustly inferred the IRMA network and estimated the strength of included interactions are given in Section A.6. The TOEL example that we used for graphical illustrations of robust variable selection is specified in Section A.7. In Section A.8, we show in detail how to use the structured singular value and singular value decomposition to determine if a set of regressors is practically independent, as described in Section 5.10, for the TOEL example with three different error models.

## A.1 Use of effective rank in network inference

We here discuss the use of effective rank (Konstantinides and Yao, 1988; Roy and Vetterli, 2007; Aksasse et al., 2006) in network inference. Effective rank is a previous extension of the rank concept and therefore interesting to compare to practical rank, which we introduced in this work, see Definition 5.5.15. Existing definitions of effective rank are based on the singular values of an observed matrix  $\Phi$ , since the  $\text{rank}(\Phi)$  is equal to the number of nonzero singular values  $k \leq \min\{m, n\}$ . These singular values are obtained from the singular value decomposition (SVD) defined in (5.10.5). Assuming that the observed matrix contain additive errors

$$\Phi = \check{\Phi} + \Upsilon, \quad (\text{A.1.1})$$

as in our data model (5.3.2a), the notion of effective rank has been introduced to find the rank of the latent “true” matrix  $\check{\Phi}$  from the SVD of  $\Phi$ . The singular values of  $\check{\Phi}$  are assumed to be larger than the singular values of the error matrix  $\Upsilon$  and the number of large singular values of  $\Phi$ , according to some measure, is defined as the effective rank  $\text{erank}(\Phi)$ , implying that

$$\text{erank}(\Phi) \leq \text{rank}(\Phi). \quad (\text{A.1.2})$$

Several measures, with references, are presented in Konstantinides and Yao (1988) and the effect of the errors on the singular values of  $\Phi$  is analysed in Stewart (1990). We recently used the concept of effective rank to classify interactions as existing, non-existing, alternative, and non-evidential, based on data with a practically collinear set of regressors but fairly high signal to noise ratio (SNR), with some success (Nordling and Jacobsen, 2011). However, the employed method depends on a sufficient separation between the singular values of  $\check{\Phi}$  and  $\Upsilon$ , similar to other

methods for calculating the effective rank, see *e.g.* Theorem 3 in [Konstantinides and Yao \(1988\)](#). We have earlier postulated that gene regulatory networks are interampatte, which means that the “true” regressor matrix in practice has some small singular values, see Chapter 3. Expression data is, as mentioned earlier, known to be noisy with a poor SNR, so some singular value of  $\mathbf{Y}$  is in general larger than some singular value of  $\check{\Phi}$ . Methods based on effective rank will therefore in general not work as intended. Moreover, it is clear based on the dyadic expansion of the pseudo inverse of the regressor matrix

$$\Phi^\dagger = \sum_{j=1}^k \frac{1}{\sigma_j} \mathbf{v}_j \mathbf{u}_j^T, \quad \text{with } k = \min\{m, n\}, \quad (\text{A.1.3})$$

that small singular values have a large influence on the parameter estimates, since their reciprocal dominates and the ordinary least squares estimate can be expressed as  $\hat{\theta} = \Phi^\dagger \xi$ . Qualitatively different parameter estimates and network models are obtained depending on which of the small singular values that are included or excluded. Inference methods that use effective rank to obtain a model of the network can therefore be misleading.

## A.2 Illustration of variable selection and parameter estimation

Our illustrations are intended to concretize the theoretical results and provide an intuitive understanding of the data conditions for variable selection. Our ability to understand new concepts is largely dependent on connecting them to previous knowledge by seeing similarities. To foster this process we here comment on the connection between our illustrations and common representations used to illustrate, in particular, parameter estimation by least squares.

Figures similar to Figure A.7 (bottom) and Figure 5.1 are commonly used to explain *e.g.* least squares estimation ([Casella and Berger, 2001](#), p. 542), the orthogonality principle of least squares regression ([Kay, 1993](#), p. 228), and collinearity ([Belsley, 1991](#), p. 15-16). Estimation of the parameter  $\theta$ , under assumption of the data model  $\phi_i \theta + v_i = \xi_i$  for each row  $i$ , is in most textbooks depicted in a two dimensional Euclidean space with  $\phi$  on the x-axis,  $\xi$  on the y-axis, and  $m$  points representing the  $m$  observations of the  $n = 1$  regressor  $\phi_i$  and regressand  $\xi_i$ , see *e.g.* [Casella and Berger \(2001, p. 542\)](#), *i.e.* the equations/rows are seen as vectors in a  $n + 1$  dimensional space. The least squares estimate  $\hat{\theta}$  is, in such a figure, given by the slope of the line that minimizes the sum of squared residuals, *i.e.* errors  $\hat{v}_i = \phi_i \hat{\theta} - \xi_i$ . In Figure A.7 (bottom) we used an  $n$  dimensional space having only the  $n$  regressors as axes and call it the state space. If each variable corresponds to a gene, then we also call it the gene space. The orthogonality principle of least-squares regression is, however, commonly depicted in a  $m$  dimensional Euclidean space with an observation on each axis and two vectors corresponding to the regressor  $\phi$  and

regressand  $\xi$ , see *e.g.* Kay (1993, p. 228), *i.e.* the columns are seen as  $n + 1$  vectors in a  $m$  dimensional dual space that we call the data space and used in Figure 5.1. The least squares estimate is now found by projecting the regressand onto the subspace spanned by the regressor. The merits of both geometrical interpretations are discussed in Belsley (1991, p. 15-16), where they are used to illustrate collinearity. In a third representation of estimation problems, which is commonly used in linear programming, see *e.g.* Reeb and Leavengood (1998); Nash and Sofer (1996, p. 67-83), each axis corresponds to a parameter and each equation/row constitutes a hyperplane that limits the set of feasible solutions. We selected to primarily use the data space, in order to easily depict the uncertainty sets of each variable.

### A.3 Additional theorems with proofs

Here we provide additional theorems and lemmas that are needed in the proofs of the main results.

The following lemma is used in the proof of Theorem A.3.2 to show that no confidence can be assigned to the existence of any interaction if  $m < n$  and noise exists.

**Lemma A.3.1.** *For any  $\Phi \in \mathbb{R}^{m \times n}$  with  $m \leq n$  and  $\varepsilon > 0$  there exists a  $\Delta \in \mathbb{R}^{m \times n}$  such that  $\text{span}(\Phi + \Delta) = \mathbb{R}^m$  and  $\|\Delta\|_\alpha < \varepsilon$  for any seminorm  $\|\cdot\|_\alpha$ .*

**Proof.** *The number of dimensions spanned by a matrix  $\Phi \in \mathbb{R}^{m \times n}$  equals the number of nonzero singular values (Horn and Johnson, 1990, p. 414). We can always construct a matrix*

$$\Delta \triangleq \sum_{i=1}^n \varsigma_i \mathbf{u}_i \mathbf{v}_i^T \in \mathbb{R}^{m \times n}, \quad (\text{A.3.1})$$

with  $\varsigma_i > 0$  if  $\sigma_i = 0$  and  $\varsigma_i = 0$  if  $\sigma_i > 0$  in the SVD (Horn and Johnson, 1990, p. 414-415)

$$\Phi = \sum_{i=1}^n \sigma_i \mathbf{u}_i \mathbf{v}_i^T, \quad (\text{A.3.2})$$

which ensures that  $\Phi + \Delta$  spans  $m$  dimensions and hence the space  $\mathbb{R}^m$ . The only remaining thing is to show that we for every  $\varepsilon > 0$  can find a  $\delta > 0$  such that if all  $\varsigma_i \leq \delta$  then  $\|\Delta\|_\alpha < \varepsilon$  for any seminorm  $\|\cdot\|_\alpha$ . Let  $\mathcal{S} \triangleq \{i | \sigma_i = 0\}$  and pick  $\delta = \frac{1}{2} \left\| \sum_{i \in \mathcal{S}} \mathbf{u}_i \mathbf{v}_i^T \right\|_\alpha^{-1} \varepsilon$ , then using the multiplicative property and triangle inequality of seminorms (Horn and Johnson, 1990, p. 259) we can show that

$$\|\Delta\|_\alpha = \left\| \sum_{i \in \mathcal{S}} \varsigma_i \mathbf{u}_i \mathbf{v}_i^T \right\|_\alpha \leq \max_i \varsigma_i \left\| \sum_{i \in \mathcal{S}} \mathbf{u}_i \mathbf{v}_i^T \right\|_\alpha \leq \frac{1}{2} \varepsilon < \varepsilon, \quad (\text{A.3.3})$$

which completes our proof.  $\square$

**Theorem A.3.2.** *For every possible interaction, an alternative network model without it always exists, that can explain noisy data, if the number of experiments  $m$  is smaller than the number of variables  $n$ , implying that no confidence can be assigned to any inferred interaction.*

**Proof.** *The data model (4.2.5) and (5.5.1) connects the data (regressors and regressand) to the network model. For  $m < n$  the system of equations specified by the data model is underdetermined and has infinitely many solutions, see e.g. Anton and Rorres (2000); Friedberg et al. (2003); Horn and Johnson (1990), but some regressor could be necessary to explain the data (as shown for noise-free data in Section 4.2). However, if a regressor  $\phi_k$  can be replaced by a linear combinations of the remaining regressors in the matrix  $\Phi_{j \neq k} \triangleq [\phi_1, \dots, \phi_{k-1}, \phi_{k+1}, \dots, \phi_n]$ , then an alternative model without the interaction from variable  $k$  exists, since the solution of a system of equations only is unique if the regressors are linearly independent, see e.g. Anton and Rorres (2000); Friedberg et al. (2003); Horn and Johnson (1990). Based on Lemma A.3.1 a realization of  $\Phi_{j \neq k}$  that spans  $\mathbb{R}^m$  exists within a distance  $\varepsilon$  from the observed  $\Phi_{j \neq k}$  measured by any norm when  $m < n$ . Since  $\phi_k \in \mathbb{R}^m$  and this holds for any  $\varepsilon > 0$  it implies that an alternative model without the interaction from variable  $k$  that can explain the data always exists when each data point contains noise or uncertainty. This holds for each variable  $k \in \{1, \dots, n\}$ .  $\square$*

The following lemma is needed in the proof of Theorem 5.6.3. While Lemma A.3.1 concerned the existence of a matrix spanning the space, the following lemma concerns the ability to represent a vector as a linear combination of the columns of a matrix.

**Lemma A.3.3.** *If a linearly dependent subset  $\mathcal{S}$  of the vectors in  $\Phi \in \mathbb{R}^{m \times n}$  with at most  $\min\{m, n\}$  elements exists, then for any  $\phi \in \mathbb{R}^m$  a matrix  $\tilde{\Phi} \in \mathcal{B}_\rho(\Phi)$  exists such that the subset is linearly independent and  $\phi \in \text{span}(\tilde{\Phi}_{\mathcal{S}})$ , with the neighbourhood of a matrix  $\mathcal{B}_\rho(\Phi)$  defined in (5.6.6) and  $\tilde{\Phi}_{\mathcal{S}}$  being a matrix that only contains the vectors indexed in  $\mathcal{S}$ .*

**Proof.** *Linear dependence implies that a non-trivial solution  $\theta \neq \mathbf{0}$  exists such that  $\Phi_{\mathcal{S}}\theta = \mathbf{0}$  (Friedberg et al., 2003, p. 36-37). To have  $\phi \in \text{span}(\tilde{\Phi}_{\mathcal{S}})$  is equivalent to existence of a vector  $\tilde{\theta}$  such that  $\tilde{\Phi}_{\mathcal{S}}\tilde{\theta} = \phi$ , based on the definition of span (Friedberg et al., 2003, p. 30), and the set of vectors in  $\tilde{\Phi}_{\mathcal{S}}$  is linearly independent when  $\tilde{\Phi}_{\mathcal{S}}\theta \neq \mathbf{0}$  for every solution  $\theta \neq \mathbf{0}$ . Any vector  $\phi \in \mathbb{R}^m$  can now be constructed from  $\tilde{\Phi}\theta$  by adding  $\tilde{\Phi}\tilde{\theta}$ , which gives*

$$\phi = \Phi_{\mathcal{S}}\theta + \tilde{\Phi}_{\mathcal{S}}\tilde{\theta}. \quad (\text{A.3.4})$$

*Intuitively this can be understood by considering that we can make small changes to each element of  $\Phi$ , which translate into small changes in each element of the zero vector and thereby produce a tiny nonzero vector that points in the wished direction. This tiny vector is then amplified by changing the elements of  $\theta$ . If we introduce*

$\Delta = \tilde{\Phi}_S - \Phi_S$ , then we can rewrite this as

$$\begin{aligned}\phi &= \Phi_S \theta + (\Delta + \Phi_S) \tilde{\theta} && \Leftrightarrow \\ \phi &= \Phi_S (\theta + \tilde{\theta}) + \Delta \tilde{\theta} && \Leftrightarrow \\ \phi - \Phi_S (\theta + \tilde{\theta}) &= \Delta \tilde{\theta}. && \text{(A.3.5)}\end{aligned}$$

Now we can take the norm and use the submultiplicative property ([Horn and Johnson, 1990](#), p. 290-293)

$$\begin{aligned}\|\phi - \Phi_S (\theta + \tilde{\theta})\| &= \|\Delta \tilde{\theta}\| \leq \|\Delta\| \|\tilde{\theta}\| && \Leftrightarrow \\ \frac{\|\phi - \Phi_S (\theta + \tilde{\theta})\|}{\|\tilde{\theta}\|} &\leq \|\Delta\|. && \text{(A.3.6)}\end{aligned}$$

We do not have any constraint on  $\tilde{\theta}$ , so for any  $\rho > 0$  there exists a  $\delta$ , such that if  $\|\tilde{\theta}\| > \delta$  then  $\|\Delta\| < \rho$ . Only the norm of  $\Delta$  is bounded so we have the freedom to select the tiny changes of the elements of  $\Phi$  such that the subset becomes linearly independent, since the number of elements in  $\mathcal{S}$  at most is  $\min\{m, n\}$  and each basis of  $\mathbb{R}^m$  consists of a linearly independent set of  $m$  vectors ([Friedberg et al., 2003](#), p. 46-48). This proves that a matrix  $\tilde{\Phi} \in \mathcal{B}_{\rho_{\Phi}}(\Phi)$  exists, such that  $\phi \in \text{span}(\tilde{\Phi}_S)$  for any  $\phi \in \mathbb{R}^m$ , and  $\tilde{\Phi}$  can always be chosen such that the subset  $\mathcal{S}$  of the vectors are linearly independent.  $\square$

The following lemma is needed in the proof of Theorems [5.6.3](#) and [5.6.7](#).

**Lemma A.3.4.** *If a solution  $\theta$  exists for the system of equations  $\Phi \theta = \xi$ , with  $\Phi \in \mathbb{R}^{m \times n}$  and  $\xi \in \mathbb{R}^m$ , then another matrix  $\tilde{\Phi} \in \mathcal{B}_{\rho_{\Phi}}(\Phi)$  and another vector  $\tilde{\xi} \in \mathcal{B}_{\rho_{\xi}}(\xi)$  exists such that a solution  $\tilde{\theta} \in \mathcal{B}_{\rho_{\theta}}(\theta)$  exists for  $\tilde{\Phi} \tilde{\theta} = \tilde{\xi}$ . The neighbourhood of a matrix  $\mathcal{B}_{\rho}(\Phi)$  is defined in [\(5.6.6\)](#) and the neighbourhood of a vector  $\mathcal{B}_{\rho}(\xi)$  in [\(5.6.5\)](#).*

**Proof.** Addition of a nonzero matrix  $\Delta \in \mathbb{R}^{m \times n}$  with  $\|\Delta\| < \rho_{\Phi}$  to  $\Phi$  in  $\Phi \theta = \xi$  gives

$$\tilde{\Phi} \tilde{\theta} = (\Phi + \Delta) \theta = \underbrace{\Phi \theta}_{=\xi} + \underbrace{\Delta \theta}_{=\delta} = \tilde{\xi}. \quad \text{(A.3.7)}$$

Here  $\tilde{\Phi}$  is by construction a matrix different from  $\Phi$  and  $\tilde{\xi}$  is within a neighbourhood of  $\xi$ , since

$$\|\tilde{\xi} - \xi\| = \|\delta\| = \|\Delta \theta\| \leq \|\theta\| \|\Delta\| < \|\theta\| \rho_{\Phi} = \rho_{\xi}. \quad \text{(A.3.8)}$$

This neighbourhood can be made arbitrarily small by selecting  $\rho_{\Phi}$  sufficiently small. Similar constructs can be made by adding a nonzero vector to  $\xi$  or  $\theta$ .  $\square$

The proof of Theorem [5.6.7](#) depends on a parameter in the solution of a system of equations being positive, which we establish in the following theorem. To do this, we first establish the following projection lemma, which is needed to prove the theorem.

**Lemma A.3.5.** *The projection  $(\mathbf{I} - \mathbf{T}_{j \neq k})\phi_k = \mathbf{0}$  with  $\mathbf{T}_{j \neq k}$  defined in (5.6.8) if and only if  $\phi_k$  can be expressed as a linear combination of the regressors in  $\Phi_{j \neq k}$ , which contains a subset of all regressors except  $\phi_k$ .*

**Proof.** *The projection*

$$(\mathbf{I} - \mathbf{T}_{j \neq k})\phi_k = \mathbf{0} \quad \Leftrightarrow \quad \phi_k = \mathbf{T}_{j \neq k}\phi_k. \quad (\text{A.3.9})$$

If  $\phi_k$  can be expressed as a linear combination of the regressors in  $\Phi_{j \neq k}$ , then  $\phi_k = \Phi_{j \neq k}\theta$ , which inserted in (A.3.9) gives

$$\begin{aligned} \mathbf{T}_{j \neq k}\phi_k &= \Phi_S(\Phi_S^T\Phi_S)^{-1}\Phi_S^T\Phi_{j \neq k}\theta \\ &= \Phi_{j \neq k}\theta = \phi_k. \end{aligned} \quad (\text{A.3.10})$$

Note that the inverse of  $\Phi_S^T\Phi_S$  exists since  $\Phi_S$  per definition consists of a linearly independent subset of regressors that constitute a basis of the subspace spanned by the  $n - 1$  regressors in the matrix  $\Phi_{j \neq k} = [\phi_1, \dots, \phi_{k-1}, \phi_{k+1}, \dots, \phi_n]$ . The projection  $\mathbf{T}_{j \neq k}$  does not affect vectors that lie in the subspace spanned by  $\Phi_{j \neq k}$ , since they already are in the subspace that it projects vectors onto, which explains why it has no effect on  $\Phi_{j \neq k}$  in (A.3.10). If  $\phi_k$  cannot be expressed as a linear combination of the regressors in  $\Phi_{j \neq k}$ , then  $\phi_k = \Phi_{j \neq k}\theta + z$  with  $\Phi_{j \neq k}^T z = \mathbf{0}$ , which inserted in the projection gives

$$\mathbf{T}_{j \neq k}\phi_k = \Phi_{j \neq k}\theta \neq \phi_k. \quad (\text{A.3.11})$$

□

**Theorem A.3.6.** *A parameter  $\theta_k$  is positive in all solutions of a system of linear equations  $\Phi\theta = \xi$  if and only if  $\mathbf{h}_k^T\xi > 0$ , with the normal of the hyperplane of the other regressors  $\mathbf{h}_k$  defined in (5.6.10).*

**Proof.** *The system of linear equations*

$$\Phi\theta = \xi \quad \Leftrightarrow \quad \Phi_{j \neq k}\theta_{j \neq k} + \phi_k\theta_k = \xi, \quad (\text{A.3.12})$$

with  $\theta_{j \neq k} \triangleq [\theta_1, \dots, \theta_{k-1}, \theta_{k+1}, \dots, \theta_n]^T$ . The condition  $\mathbf{h}_k^T\xi > 0$  is equivalent to

$$\begin{aligned} \phi_k^T(\mathbf{I} - \mathbf{T}_{j \neq k})\xi &> 0 && \Leftrightarrow \\ \phi_k^T(\mathbf{I} - \mathbf{T}_{j \neq k})(\Phi_{j \neq k}\theta_{j \neq k} + \phi_k\theta_k) &> 0 && \Leftrightarrow \\ \phi_k^T(\mathbf{I} - \mathbf{T}_{j \neq k})\phi_k\theta_k &> 0, && (\text{A.3.13}) \end{aligned}$$

since  $\mathbf{I} - \mathbf{T}_{j \neq k}$  is an orthogonal projection and fulfils (Friedberg et al., 2003, p. 400)

$$\begin{aligned} \mathbf{T}_{j \neq k}^T \mathbf{T}_{j \neq k} &= \Phi_S (\Phi_S^T \Phi_S)^{-1} \Phi_S^T \\ \Phi_S (\Phi_S^T \Phi_S)^{-1} \Phi_S^T &= \mathbf{T}_{j \neq k} = \mathbf{T}_{j \neq k}^T \end{aligned} \quad (\text{A.3.14})$$

$$\begin{aligned} (\mathbf{I} - \mathbf{T}_{j \neq k})^2 &= (\mathbf{I} - \mathbf{T}_{j \neq k})^T (\mathbf{I} - \mathbf{T}_{j \neq k}) \\ &= \mathbf{I} - \mathbf{T}_{j \neq k}^T - \mathbf{T}_{j \neq k} + \mathbf{T}_{j \neq k}^T \mathbf{T}_{j \neq k} \\ &= \mathbf{I} - \mathbf{T}_{j \neq k} = (\mathbf{I} - \mathbf{T}_{j \neq k})^T \end{aligned} \quad (\text{A.3.15})$$

$$\begin{aligned} (\mathbf{I} - \mathbf{T}_{j \neq k}) \Phi_{j \neq k} &= \Phi_{j \neq k} - \\ \Phi_S (\Phi_S^T \Phi_S)^{-1} \Phi_S^T \Phi_{j \neq k} &= \mathbf{0}. \end{aligned} \quad (\text{A.3.16})$$

Here  $\mathbf{I} - \mathbf{T}_{j \neq k}$  is a positive semidefinite matrix, since we from (A.3.15) have  $(\mathbf{I} - \mathbf{T}_{j \neq k})^2 = \mathbf{I} - \mathbf{T}_{j \neq k}$ , see e.g. (Horn and Johnson, 1990, p. 396). So in (A.3.13)  $\phi_k^T (\mathbf{I} - \mathbf{T}_{j \neq k}) \phi_k \geq 0$  and it is based on Lemma A.3.5 equal to zero if and only if  $\phi_k$  can be expressed as a linear combination of the regressors in  $\Phi_{j \neq k}$ . Therefore  $\mathbf{h}_k^T \xi > 0$  implies that  $\theta_k > 0$  and  $\phi_k$  cannot be expressed as a linear combination of the other regressors. On the other hand, if  $\theta_k$  is positive in all solutions, then  $\mathbf{h}_k^T \xi > 0$ , since a parameter only can be positive in all solutions if it is unique and  $\phi_k$  cannot be expressed as a linear combination of the other regressors. Note that the inverse of  $\Phi_S^T \Phi_S$  exists since  $\Phi_S$  per definition consists of a linearly independent subset of regressors that constitute a basis of the subspace spanned by the  $n - 1$  regressors in the matrix  $\Phi_{j \neq k} = [\phi_1, \dots, \phi_{k-1}, \phi_{k+1}, \dots, \phi_n]$ .  $\square$

#### A.4 Generation of synthetic data for the IRMA example

We simulated the *in silico* perturbation experiments in the introductory example in Chapter 4 using unit strengths of all interactions in the engineered IRMA network, since the real strengths are unknown and it simplifies our illustration. In other words, we used the interaction matrix (2.1.1). The “true” regressor matrix is the transpose of the “true” response matrix  $\check{\mathbf{Y}}$ , which for steady-state experiments is given by the data model

$$\mathbf{Y} = -\check{\mathbf{A}}^{-1} \mathbf{P} = \check{\mathbf{G}} \mathbf{P}, \quad (\text{A.4.1})$$

i.e. the data model in (2.3.3) with  $\mathbf{E} = \mathbf{F} = \mathbf{0}$ . The simulated data in (4.2.4) is therefore obtained by

$$\check{\Phi} = -\check{\mathbf{P}}^T \check{\mathbf{A}}^{-T}, \quad (\text{A.4.2})$$

using the perturbation matrix

$$\check{P} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad (\text{A.4.3})$$

and the regressand  $\check{\xi}_3$  is equal to the third row of the perturbation matrix with opposite sign.

## A.5 Development of an error model for the IRMA data

All measurement data contains errors and these must be accounted for when inferring properties of the system from the data, in order to distinguish true findings from artefacts generated by noise or to assign a measure of confidence or significance to the properties. To do this, we need an error model that describes the typical errors of the data at hand. We have only found two studies—[Peccoud and Jacob \(1996\)](#) and [Bengtsson et al. \(2008\)](#)—that specifically investigate the errors in qRT-PCR data, even though it is generally known that RT-PCR contains many error sources, see [Section 2.2](#). The actual error is poorly characterized and no generally accepted error model exists for qRT-PCR data. We therefore next analyse the error in the IRMA data based on the  $C_T$  values reported by [Cantone et al. \(2009, supplemental\)](#). They performed 3 to 9 clones, *i.e.* biological replicates, of each experiment and reported the average  $C_T$  and standard error from two technical replicates of the PCR for each clone.

We started our error analysis by first calculating  $\Delta C_{Tic} \triangleq C_{ic} - C_{rc}$  for the three clones, *i.e.* three biological replicates of the unperturbed control experiment, and  $\Delta C_{Titp} \triangleq C_{itp} - C_{rtp}$  for the 3 to 9 clones of the perturbed experiments. Here  $C_{ic}$  is the mean  $C_T$  value for gene  $i$  in the two technical replicates of the PCR of the unperturbed control sample and  $C_{itp}$  of the sample in which gene  $t$  is perturbed, and  $C_{rc}$  is the mean  $C_T$  value for the housekeeping gene in the two technical replicates of the PCR of the unperturbed control sample and  $C_{rtp}$  of the sample in which gene  $t$  is perturbed. We here use  $i$  as the index for the observed gene and  $t$  as the index of the perturbed gene, *i.e.*  $i, t \in \{1 = \text{CBF1}, 2 = \text{GAL4}, 3 = \text{SWI5}, 4 = \text{ASH1}, 5 = \text{GAL80}\}$ , while  $r$  denotes the housekeeping gene, *i.e.*  $r = \text{ACT1}$ . Similarly,  $c$  is used as the index for the clone of the control experiment, *i.e.*  $c \in \{1, 2, 3\}$ , and  $p$  as the index for the clone of the perturbed experiment, *i.e.*  $p \in \{1, 2, \dots, 9\}$  for *ASH1* and  $p \in \{1, 2, 3\}$  for all other genes. We then calculated the log-fold changes

$$y_{itcp} \triangleq \Delta C_{Tic} - \Delta C_{Titp} \quad (\text{A.5.1})$$

for all possible combinations of one control clone and one perturbed clone. This is done since the control and perturbation experiments are independent and no

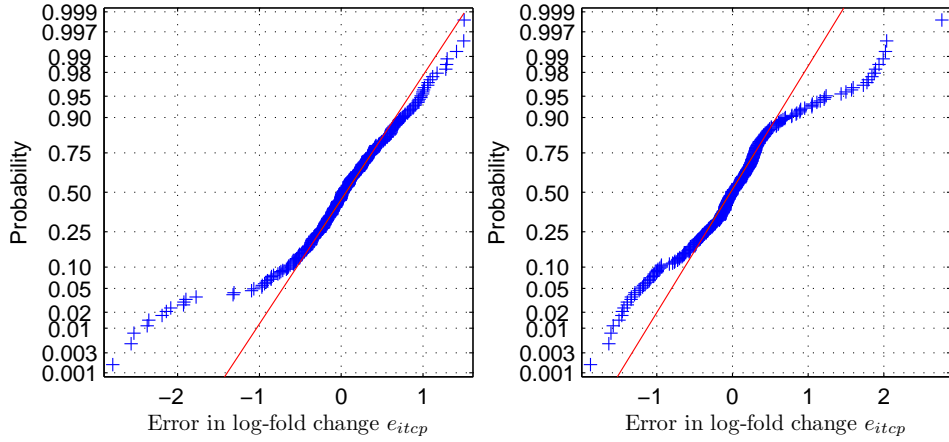
particular pairing can be motivated. The error of each possible log-fold change is then estimated as

$$e_{itcp} \triangleq y_{itcp} - y_{it}, \quad (\text{A.5.2})$$

with  $y_{it}$  being the log-fold change for gene  $i$  defined in (2.2.4), *i.e.* the mean log-fold change

$$\begin{aligned} y_{it} &= \frac{1}{\max c} \frac{1}{\max p} \sum_{c=1}^{\max c} \sum_{p=1}^{\max p} (\Delta C_{Tic} - \Delta C_{Titp}) \\ &= \underbrace{\frac{1}{\max c} \sum_{c=1}^{\max c} \Delta C_{Tic}}_{=\Delta C_{TiC}} - \underbrace{\frac{1}{\max p} \sum_{p=1}^{\max p} \Delta C_{Titp}}_{=\Delta C_{Tit}}. \end{aligned} \quad (\text{A.5.3})$$

We chose to neglect the standard errors of the technical replicates, which were reported by Cantone *et al.*, since they are negligible compared to the calculated errors and the use of error propagation formulas merely would complicate the calculations. Several authors, listed in Steibel *et al.* (2009), have assumed that the log-fold change is normally distributed, which in Bengtsson *et al.* (2008) is supported by quantile-quantile plots of the  $C_T$  values, while Peccoud and Jacob (1996) showed that the distribution has several maxima for certain amplification efficiencies and only approaches a normal distribution in the limit as the initial copy number goes to infinity. We therefore tested the hypothesis that the log-fold change is normally distributed by a Lilliefors test (Lilliefors, 1967) and studied the normal probability plots for the errors  $e_{itcp}$  in both the Glucose and Galactose data using Matlab R2007b ([www.mathworks.com](http://www.mathworks.com)). Note that the log-fold change and errors have the same distribution since we only subtract the mean to get the errors, but the mean value is in general different for all genes in all perturbation experiments so by subtracting it we can compare all values and thereby increase the statistical power. The normal hypothesis was rejected at the 5% significance level, with p-values smaller than 0.001, based on all 315 data points for both the *S. cerevisiae* cells cultured in Glucose and Galactose. The data points are clearly not on the line as they should be if the data came from a normal distribution in the normal probability plots shown in Figure A.1. When comparing all values, we assume that the variance is the same for all genes in all experiments, which in general is not true. We therefore also tested the normal hypothesis for each gene individually and found that it was rejected at the 5% significance level for three of the five genes, with p-values below 0.013, for both the Glucose and Galactose data using the 63 data points that then were available. The data points in all normal probability plots exhibit some curvature or deviation from the line (data not shown). And for each experiment individually, it was rejected at the 5% significance level only for the experiment in which *ASH1* was perturbed, with p-values below 0.009, for both the Glucose and Galactose data. Note that we only had 45 data points for the experiments when the hypothesis was not rejected, compared to 135 for the

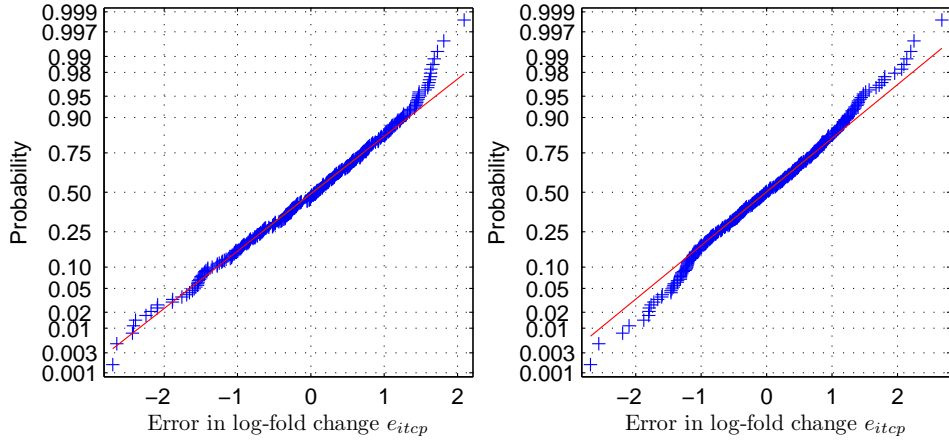


**Figure A.1: Normal probability plots.** The error of the log-fold change of the steady-state data recorded by Cantone et al. (2009) when *S. cerevisiae* is grown in Glucose (left) and Galactose (right) is plotted against a theoretical normal distribution, such that the data points should lay approximately on a line if the data stems from a normal distribution. Significant curvature and deviations from the line are seen and the log-fold change of the data is therefore not normally distributed.

experiment when it was rejected. Lack of statistic power could explain the failure to reject the hypothesis for the other experiments. For the Glucose data only the data points in the normal probability plot of the *ASH1* experiment exhibit strong curvature, but for the Galactose data the data points in all normal probability plots exhibit some curvature or deviation from the line (data not shown). The variance could also depend on both the gene and experiment. It is not in this case possible to reliably test the normal hypothesis for each gene and experiment individually, since we only have nine data points for all genes, except for the experiment in which *ASH1* was perturbed for which we have 27 data points. We therefore standardized each error by dividing with the standard deviation

$$e_{itcp}^s = \frac{e_{itcp}}{\sqrt{\frac{1}{\max c \max p-1} \sum_{c=1}^{\max c} \sum_{p=1}^{\max p} e_{itcp}^2}} \quad (\text{A.5.4})$$

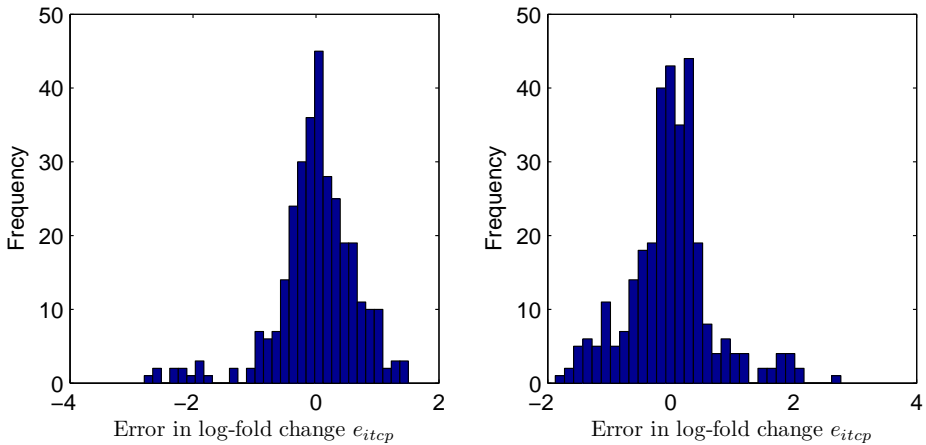
and tested the hypothesis using all standardized errors. Now the normal hypothesis was not rejected at the 5% significance level based on the Lilliefors test for the Glucose data (p-value 0.1), nor was it for the Galactose data (p-value 0.3). The normal probability plot of the Glucose data exhibits clear curvature at the tail, while the plot of the Galactose data points almost form a line, see Figure A.2. Failure to reject a hypothesis does not mean that it is true, merely that we cannot at the desired 5% significance level rule out that the log-fold change is normally



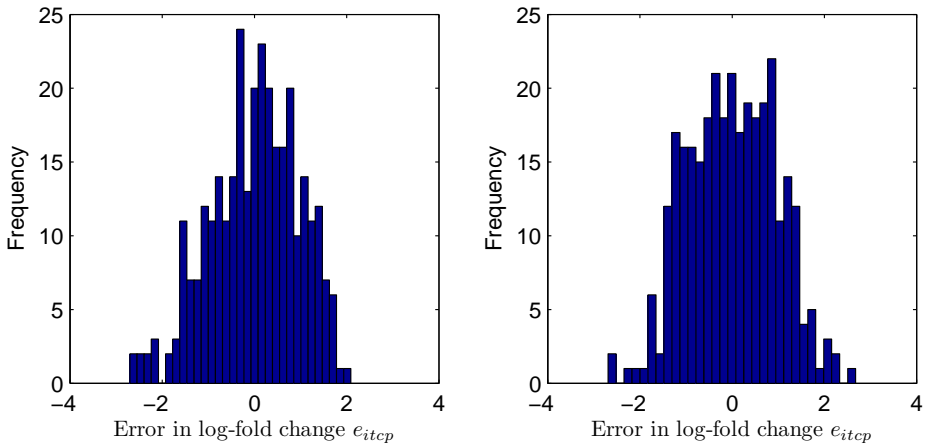
**Figure A.2: Normal probability plots.** The standardized error of the log-fold change of the steady-state data recorded by Cantone et al. (2009) when *S. cerevisiae* is grown in Glucose (left) and Galactose (right) is plotted against a theoretical normal distribution, such that the data points should lay approximately on a line if the data stems from a normal distribution. Clear curvature is seen for the Glucose data, while the Galactose data points almost lies on the line, so the log-fold change could be normally distributed with a variance that depends on the experiment and gene.

distributed with a variance that depends on the experiment and gene. We later study the histogram to see if we recognize any well known distribution. When combined the results of our investigation so far suggests that the log-fold change of the steady-state data recorded by Cantone et al. (2009) is not normally distributed.

The log-fold change of the IRMA data is unlikely to be normally distributed and we therefore continue our analysis of the errors in order to find a reasonable error model. We plotted the histogram of the errors of the log-fold change using Matlab R2007b ([www.mathworks.com](http://www.mathworks.com)) in order to see if we can recognize any well known probability distribution. The histograms for all 315 data points of both the Glucose and Galactose data is shown in Figure A.3. Unfortunately we are not able to recognize any well known distribution, but we note that both histograms have multiple peaks, indicating a multimodal distribution in agreement with the theoretical study by Peccoud and Jacob (1996). When we use all data points then we assume that the distribution does not depend on the gene or experiment, which in general does not hold. We therefore also studied the histogram of each gene and each experiment independently and they all appear to have multiple peaks (data not shown). We also studied the histogram of the standardized errors (A.5.4) to see if it resembles the probability density function of a normal distribution and found that it too appears to have multiple peaks, see Figure A.4, which implies that it is unlikely to be normally distributed. Previous studies of the absolute



**Figure A.3: Histograms.** A histogram of the error of the log-fold change of the steady-state data recorded by [Cantone et al. \(2009\)](#) when *S. cerevisiae* is grown in Glucose (left) and Galactose (right). Both histograms have multiple peaks, which indicates a multimodal distribution.



**Figure A.4: Histograms.** A histogram of the standardized error of the log-fold change of the steady-state data recorded by [Cantone et al. \(2009\)](#) when *S. cerevisiae* is grown in Glucose (left) and Galactose (right). Neither of the histograms look like the probability density function of a normal distribution. Both histograms appear to have two peaks, implying that the data is unlikely to be normally distributed.

quantification error have found that the error mainly depends on the initial copy number and thereby on the measured gene (Peccoud and Jacob, 1996; Bengtsson et al., 2008). We therefore investigated if the maximum, minimum, mean, or median absolute error, as well as, the standard error calculated by Cantone *et al.* correlates with the gene, experiment, growth medium, log-fold change (2.2.4), or absolute value of the mean log-fold change in the combined Glucose and Galactose data. We calculated the correlation coefficient and p-value using `corrcoef` in Matlab R2009b ([www.mathworks.com](http://www.mathworks.com)). The p-value is the probability of getting a correlation as high as the observed one by chance when the true correlation is zero. To calculate the correlation with the gene and experiment, we assigned a number to each of the five genes or experiments. The correlation depends on which number that was assigned to which gene or experiment, so we calculated the correlation for all,  $5! = 120$ , possible assignments and report the most significant correlation, *i.e.* the one with the lowest p-value, in order to avoid dependence on our numbering. Similarly, numbers were assigned to the two different growth mediums, Glucose and Galactose, to investigate if the error correlates with them. All calculated correlation coefficients are presented in Table A.1. The maximum absolute error and mean absolute error are clearly correlated with the perturbation experiment, correlation coefficient 0.75 and p-values  $5 \times 10^{-10}$  and  $5 \times 10^{-8}$ , which are at least five orders of magnitude smaller than the p-values of any other investigated correlation. In Bengtsson et al. (2008) the RT and PCR error depended on the gene that was observed and we therefore expected that the error would correlate with the gene. A weak but significant correlation indeed exists between the mean absolute error and the gene, correlation coefficient 0.29 and p-value 0.004, but the correlation with the experiment is both stronger and more significant, which suggests that an unknown error source that depends on the perturbation experiment dominates. No significant correlation exists between the error and the log-fold change or growth medium. This is a strong argument against using a multiplicative error model in which the error depends on the observed log-fold change and any model in which the error depends on the growth medium. Taken together, these observations points towards an additive error model in which the magnitude of the error mainly depends on the perturbation experiment and to a lesser degree on the gene. The earlier part of our analysis pointed towards an unknown multimodal distribution and the main error source is unknown. To prioritize robustness of the results, we therefore decide to use a deterministic error model in which the absolute error of each element is bounded, *i.e.* the observed log-fold change  $y_{it}$  is assumed to consist of the “true” log-fold change  $\check{y}_{it}$  plus some error  $v_{it}$

$$y_{it} = \check{y}_{it} + v_{it} \quad \text{with } |v_{it}| \leq \varpi_{it}. \quad (\text{A.5.5})$$

To use this error model we need to determine the bounds of the elements  $\varpi_{it}$ , which we do next and present in Table A.2 for the Glucose data and in Table A.3 for the Galactose data.

We will now determine upper bounds on the magnitude of the errors which we can use in the selected deterministic error model. The absolute error estimates of

**Table A.1: Correlation between the error and certain variables.** The correlation coefficient of the error and variable with associated p-value in parenthesis for the data in Cantone et al. (2009). The most significant correlation exists between the experiment and maximum absolute error, as well as the mean absolute error. The error  $e_{itcp}$  is defined in (A.5.2) and the log-fold change  $y_{it}$  in (2.2.4).

Variable \ Error	$\max_{c,p}  e_{itcp} $	$\min_{c,p}  e_{itcp} $	$\text{mean}_{c,p}  e_{itcp} $
Experiment	0.75 ( $5 \times 10^{-10}$ )	0.44 (0.1)	0.75 ( $5 \times 10^{-8}$ )
Gene	0.29 (0.04)	0.13 (0.01)	0.29 (0.004)
Growth medium	0.01 (0.9)	0.23 (0.1)	0.07 (0.6)
$y_{it}$	-0.09 (0.5)	-0.13 (0.4)	-0.04 (0.8)
$ y_{it} $	-0.04 (0.8)	-0.11 (0.4)	0.03 (0.8)

Variable \ Error	$\text{median}_{c,p}  e_{itcp} $	standard error
Experiment	0.75 (0.00002)	0.45 (0.001)
Gene	0.25 (0.004)	0.29 (0.006)
Growth medium	0.08 (0.6)	0.03 (0.8)
$y_{it}$	0.003 (1)	0.18 (0.2)
$ y_{it} $	0.09 (0.6)	0.24 (0.09)

**Table A.2: Upper bounds for error in the Glucose data.** Our estimates of the upper bound on the error in the log-fold changes  $\varpi_{it}$ , defined in (A.5.5), in the Glucose steady-state experiments recorded by Cantone et al. (2009), which we obtained by calculating the maximum absolute error.

Observed \ Perturbed gene	<i>CBF1</i>	<i>GAL4</i>	<i>SWI5</i>	<i>ASH1</i>	<i>GAL80</i>
<i>CBF1</i>	0.5	0.6	1.3	2.8	0.8
<i>GAL4</i>	0.4	0.8	0.6	2.4	0.4
<i>SWI5</i>	0.5	1.0	0.8	2.5	0.5
<i>ASH1</i>	0.3	0.3	0.7	2.1	0.3
<i>GAL80</i>	0.2	0.4	0.4	1.3	0.3

each gene for each experimental condition is only based on nine data points for all genes, except *ASH1* for which it is based on 27 data points. These data points are, for each gene, all derived from qRT-PCR measurements of only three unperturbed control experiments. Therefore, our estimates of the error are likely to be sensitive, *i.e.* if Cantone *et al.* would have done an additional biological replicate of each perturbation experiment then our estimates would probably have changed. To check how sensitive they are, we left one error  $e_{itcp}$  at a time out and recalculated the maximum, minimum, mean, and median absolute error. We chose this procedure since the number of data points is low, the variance may be a poor estimate of

**Table A.3: Upper bounds for error in the Galactose data.** Our estimates of the upper bound on the error in the log-fold changes  $\varpi_{it}$ , defined in (A.5.5), in the Galactose steady-state experiments recorded by Cantone *et al.* (2009), which we obtained by calculating the maximum absolute error.

Observed \ Perturbed gene	<i>CBF1</i>	<i>GAL4</i>	<i>SWI5</i>	<i>ASH1</i>	<i>GAL80</i>
<i>CBF1</i>	0.1	0.6	0.8	1.9	0.8
<i>GAL4</i>	0.5	1.1	0.3	1.2	0.2
<i>SWI5</i>	0.7	1.0	1.6	2.8	0.7
<i>ASH1</i>	0.3	0.6	0.6	2.0	0.4
<i>GAL80</i>	0.3	0.4	0.6	1.9	0.4

variation for a multimodal distribution, and the estimated maximum and minimum absolute errors, in particular, may change considerably when one data point is left out. We found the largest change in the maximum absolute error to be 0.3 for an error that is 2.53 in the Glucose data, while it for the minimum, mean, and median is 0.2 (0.2), 0.08 (0.76), 0.09, (0.71), with the value of the error in parenthesis. For the Galactose data the numbers are 0.7 (2.8), 0.1 (0.4), 0.1 (0.8), 0.3 (0.2), respectively. For Glucose the minimum absolute error is most sensitive, since the first digit is affected, while the second digit is affected for the other error estimates. The mean absolute error is the least sensitive both in absolute and relative numbers, but the difference compared to the maximum absolute error is small. For Galactose, on the other hand, the median absolute error is most sensitive and the maximum absolute error is the least sensitive, being the only one where only the second digit is affected. We therefore conclude that little can be gained by trying to estimate the upper bound by any other estimator than the maximum absolute error and use it. Our estimates of the upper bounds  $\varpi_{it}$  are presented in Table A.2 for the Glucose data and in Table A.3 for the Galactose data. For convenience of the reader, we also present the log-fold changes  $y_{it}$ , which were calculated by Cantone *et al.*, in Table A.5 for the Glucose data and in Table A.6 for the Galactose data. The log-fold changes  $y_{it}$  are means of the data points based on three unperturbed control experiments, while our estimates of the upper bounds  $\varpi_{it}$  apply to a single data point based on a single experiment. The upper bounds should therefore be a conservative estimate of the error in the reported log-fold changes. We later use ellipsoidal uncertainty sets with the length of the semi-principal axes equal to the bounds because it is unlikely that the error in all elements is at its maximum value at the same time.

Following Cantone *et al.* we assume the magnitude of each perturbation to be one, which makes the perturbation matrix  $\mathbf{P} = \mathbf{I}$ . This is the logical choice, since Cantone *et al.* perturbed the system by over-expressing each of the five genes one at a time in separate experiments by inserting a plasmid containing a copy of the gene under control of the strong constitutive GPD promoter. They did not measure the strength of the perturbations so no data for estimation of the perturbation errors

**Table A.4: Upper bounds for error in the perturbations.** The upper bounds we assume on the error in the perturbations  $\varpi_{it}$  used in the steady-state experiments recorded by [Cantone et al. \(2009\)](#).

Affected \ Perturbed gene	<i>CBF1</i>	<i>GAL4</i>	<i>SWI5</i>	<i>ASH1</i>	<i>GAL80</i>
<i>CBF1</i>	0.2	0.1	0.1	0.1	0.1
<i>GAL4</i>	0.1	0.2	0.1	0.1	0.1
<i>SWI5</i>	0.1	0.1	0.2	0.1	0.1
<i>ASH1</i>	0.1	0.1	0.1	0.2	0.1
<i>GAL80</i>	0.1	0.1	0.1	0.1	0.2

**Table A.5: Log-fold change for the Glucose data.** The log-fold changes  $y_{it}$ , defined in (2.2.4), for the Glucose steady-state experiments recorded by [Cantone et al. \(2009\)](#).

Observed \ Perturbed gene	<i>CBF1</i>	<i>GAL4</i>	<i>SWI5</i>	<i>ASH1</i>	<i>GAL80</i>
<i>CBF1</i>	8.79	3.69	3.32	0.53	-0.33
<i>GAL4</i>	0.52	5.05	0.31	-0.59	0.04
<i>SWI5</i>	4.39	3.71	5.83	-0.16	0.46
<i>ASH1</i>	1.68	0.55	0.96	5.16	-0.42
<i>GAL80</i>	0.05	0.43	0.55	0.20	6.13

**Table A.6: Log-fold change for the Galactose data.** The log-fold changes  $y_{it}$ , defined in (2.2.4), for the Galactose steady-state experiments recorded by [Cantone et al. \(2009\)](#).

Observed \ Perturbed gene	<i>CBF1</i>	<i>GAL4</i>	<i>SWI5</i>	<i>ASH1</i>	<i>GAL80</i>
<i>CBF1</i>	4.54	1.80	2.10	-0.43	-2.26
<i>GAL4</i>	-0.02	4.08	0.06	-0.42	-0.22
<i>SWI5</i>	2.32	2.41	2.57	-1.34	-3.57
<i>ASH1</i>	2.05	1.73	1.53	4.01	-1.45
<i>GAL80</i>	0.28	0.15	0.00	-0.26	6.63

exist. The measured responses are, however, averages for a culture and variations in the cell cultures should affect the actual average of the perturbations, so we consider it unreasonable to assume no error. For the perturbations, we therefore also assumed an additive deterministic error model with bounds  $\varpi_{it}$  on the absolute error of each data point. In lack of better guidance we set the bounds to 0.1 for unperturbed genes and 0.2 for the perturbed ones. These bounds are summarized in Table A.4.

## A.6 Robust inference of the IRMA network

We here describe how to robustly infer the IRMA network from the *in vivo* steady-state data sets recorded by Cantone et al. (2009). We reported these results for the data recorded when growing their *S. cerevisiae* strain in Galactose free medium in our concluding example in Section 4.6. Additionally, we here report the result for the Glucose data and the signs that can be inferred robustly for both networks. We also show how to use constrained optimization to estimate the parameters of these networks, including how we obtained the network estimate reported in (A.6.3). For details on IRMA see Section 2.1 or Cantone et al. (2009).

Robust inference of the existing interactions (Definition 5.5.6) is done by calculating the confidence score  $\gamma$  for practical independence of  $\Psi_{ki} = [\phi_1, \dots, \phi_{k-1}, \phi_{k+1}, \dots, \phi_n, \xi_i]$  using the structured singular value (SSV) as described in Theorem 5.10.1. If  $\Psi_{ki}$  is practically independent, then the regressor  $\phi_k$  is practically selectable based on Theorem 5.6.2 and the corresponding interaction  $a_{ik}$  exists. We assume that the error in each regressor  $\phi_j$  is deterministic and that the uncertainty sets are ellipsoidal with the length of the semi-principal axes equal to the upper bounds  $\varpi_{jt}$  in Table A.2 for the Glucose data and in Table A.3 for the Galactose data, and  $\varpi_{it}$  in Table A.4 for the perturbations. The weight matrices of the ellipsoidal sets defined in (5.10.4) are hence

$$\mathbf{Q}_{\phi_j} = \text{diag} \left( \frac{1}{\varpi_{j1}^2}, \frac{1}{\varpi_{j2}^2}, \dots, \frac{1}{\varpi_{jm}^2} \right) \text{ and } \mathbf{Q}_{\xi_i} = \text{diag} \left( \frac{1}{\varpi_{i1}^2}, \frac{1}{\varpi_{i2}^2}, \dots, \frac{1}{\varpi_{im}^2} \right). \quad (\text{A.6.1})$$

We then use the `mussv` implementation in Matlab ([www.mathworks.com](http://www.mathworks.com)) to calculate the upper bound on the SSV, which equals the lower bound on  $\gamma(\Psi_{ki})$ . An example of how to use `mussv` to determine practical independence is presented in Section A.8. The confidence score  $\gamma(\Psi_{ki})$  is the smallest factor by which these ellipsoidal uncertainty sets need to be scaled such that rank deficient realizations of  $\Psi_{ki}$  only exist on the boundary of the sets and  $a_{ik}$  therefore exists up to the corresponding significance level. In practice, the higher the  $\gamma(\Psi_{ki})$  value is the more likely is the  $a_{ik}$  interaction to exist, since the error model itself is uncertain and the interaction then can be shown to exist for larger uncertainty. The obtained lower bounds are presented in Table A.7 for the Glucose data and Table A.8 for the Galactose data. We here use two digits, even though only one is significant, in order to be able to rank the interactions.

Robust inference of the activating and repressing interactions is done by, in addition to  $\gamma(\Psi_{ki})$ , calculating the confidence score  $\gamma(\Phi)$  for practical independence of the regressor matrix  $\Phi = [\phi_1, \phi_2, \dots, \phi_n]$  using the SSV as described in Theorem 5.10.1. If both  $\Psi_{ki}$  and  $\Phi$  are practically independent, then the regressor  $\phi_k$  is practically assignable based on Theorem 5.6.8, since the number of experiments equals the number of variables and the uncertainty sets are connected set, and the corresponding interaction  $a_{ik}$  is either activating or repressing depending on the sign in any model that explains the data. We again use the `mussv` implementation

**Table A.7: Practical selectability for the Glucose data.** Lower bounds on the confidence score  $\gamma$  at which the  $j$ th regressor is practically selectable for row  $i$ , calculated by the structured singular value method using the upper bounds in Table A.2 and Table A.4 to bound the uncertainty ellipsoid of each variable recorded in the Glucose steady-state data by Cantone et al. (2009). In practice, the higher the  $\gamma$  value is the more likely is the  $a_{ij}$  interaction to exist. The lower bound on the confidence score  $\gamma$  at which the regressor matrix is practically independent is 1.5.

Gene $i \setminus j$	<i>CBF1</i>	<i>GAL4</i>	<i>SWI5</i>	<i>ASH1</i>	<i>GAL80</i>
<i>CBF1</i>	1.6	0.61	0.89	0.14	0.26
<i>GAL4</i>	0.27	1.6	0.11	0.2	0.0083
<i>SWI5</i>	1.2	0.89	1.7	0.039	0.37
<i>ASH1</i>	0.59	0.24	0.23	1.8	0.38
<i>GAL80</i>	0.33	0.14	0.42	0.11	1.5

**Table A.8: Practical selectability for the Galactose data.** Lower bounds on the confidence score  $\gamma$  at which the  $j$ th regressor is practically selectable for row  $i$ , calculated by the structured singular value method using the upper bounds in Table A.3 and Table A.4 to bound the uncertainty ellipsoid of each variable recorded in the Galactose steady-state data by Cantone et al. (2009). In practice, the higher the  $\gamma$  value is the more likely is the  $a_{ij}$  interaction to exist. The lower bound on the confidence score  $\gamma$  at which the regressor matrix is practically independent is 0.58.

Gene $i \setminus j$	<i>CBF1</i>	<i>GAL4</i>	<i>SWI5</i>	<i>ASH1</i>	<i>GAL80</i>
<i>CBF1</i>	0.94	0.08	0.95	0.14	0.17
<i>GAL4</i>	0.012	0.58	0.12	0.17	0.043
<i>SWI5</i>	0.95	0.84	1.3	0.21	1.1
<i>ASH1</i>	0.22	0.11	0.35	0.55	0.17
<i>GAL80</i>	0.17	0.1	0.064	0.099	0.59

in Matlab to calculate the upper bound on the SSV, which equals the lower bound on  $\gamma(\Phi)$ . These are presented in the caption of Table A.7 for the Glucose data and Table A.8 for the Galactose data. For consistency we here also use two digits, even if only one is significant. The interaction  $a_{ik}$  is activating or repressing at the significance level corresponding to  $\min(\gamma(\Psi_{ki}), \gamma(\Phi))$ . We mark the interactions that at  $\gamma = 1$  are classified as activating by blue and repressing by red in Table A.9, while interactions that are classified as existing but cannot be classified as activating or repressing due to  $\gamma(\Phi) < 1$  are marked by green.

Inference of a single model is standard in the systems biology community today and has the advantage of being easier to comprehend, analyse, and simulate than a set of many models, in particular if they are infinitely many. We therefore next use constrained optimisation to estimate parameters in a few models with chosen

**Table A.9: Inferred interaction matrices.** The value of each coefficients in a few chosen representatives of the set of interaction matrices that can explain the Glucose or Galactose steady-state data recorded by [Cantone et al. \(2009\)](#). The lower bound on the confidence score for practical selectability ( $\gamma$ ) in [Table A.7](#) and [Table A.8](#) is also included since it is a measure of the confidence of each interaction. Coefficients classified at  $\gamma = 1$  as existing and activating are coloured blue, exiting and repressing red, and only existing green.

Coeff.	Glucose				Galactose	
	$\gamma$	$\hat{A}_{\text{Glu}}$	$\hat{A}_{\text{Gla}}$	$\hat{A}_{\text{Glb}}$	$\gamma$	$\hat{A}_{\text{Gal}}$
$a_{11}$	1.6	-0.143	-0.153	-0.149	0.94	-0.349
$a_{12}$	0.61	0.043	0	0	0.08	0
$a_{13}$	0.89	0.073	0.103	0.099	0.95	0.237
$a_{14}$	0.14	0	0	0	0.14	0
$a_{15}$	0.26	0	0	0	0.17	0
$a_{21}$	0.27	0	0	0.004	0.012	0
$a_{22}$	1.6	-0.183	-0.173	-0.174	0.58	-0.227
$a_{23}$	0.11	0	0	0	0.12	0
$a_{24}$	0.2	0	0	0	0.17	0
$a_{25}$	0.0083	0	0	0	0.043	0
$a_{31}$	1.2	0.103	0.103	0.098	0.95	0.196
$a_{32}$	0.89	0.091	0.133	0.134	0.84	0.179
$a_{33}$	1.7	-0.219	-0.232	-0.229	1.3	-0.407
$a_{34}$	0.039	0	0	0	0.21	0
$a_{35}$	0.37	0	-0.001	-0.001	1.1	-0.137
$a_{41}$	0.59	0.036	0.038	0	0.22	0.073
$a_{42}$	0.24	0	0	-0.004	0.11	0
$a_{43}$	0.23	0	-0.002	0.044	0.35	0.052
$a_{44}$	1.8	-0.177	-0.178	-0.155	0.55	-0.205
$a_{45}$	0.38	0	0	0	0.17	0
$a_{51}$	0.33	0	0	0.006	0.17	0
$a_{52}$	0.14	0	0	0	0.1	0
$a_{53}$	0.42	0.009	0	0	0.064	0
$a_{54}$	0.11	0	0	0	0.1	0
$a_{55}$	1.5	-0.152	-0.151	-0.148	0.59	-0.141

structure. We chose the structure based on the confidence scores reported in [Table A.7](#)

and Table A.8, such that all interactions with  $\gamma > 1$  and a few with  $\gamma$  a bit below one are present, and use the constrained optimisation to ensure that the parameters are feasible and the model belongs to the set of models that cannot be rejected, defined in (5.5.3). The idea is to include the interactions that are most likely to exist based on  $\gamma$  or show that a model with a specific structure cannot be rejected. Our parameter estimation, in other words, consists of two steps. First, we chose a structure for the model, more precisely a signed adjacency matrix for the network graph, that we here, following the notation in Julius et al. (2009), encode in a matrix  $\mathbf{S} \in \{0, +, -, ?\}^{n \times n}$ . Each element  $s_{ij}$  implies that the corresponding element of the estimated interaction matrix  $\hat{a}_{ij}$  should be zero if  $s_{ij} = 0$ , positive if  $s_{ij} = +$ , negative if  $s_{ij} = -$ , and free to take any real value if  $s_{ij} = ?$ . These correspond to forcing the interaction to be absent, activating, repressing, and no constraint. Second, we solve the following constrained optimization problem:

$$\{\hat{\mathbf{A}}, \hat{\mathbf{P}}\} = \arg \min_{\hat{\mathbf{A}}, \hat{\mathbf{P}}} \sum_{t=1}^m \boldsymbol{\eta}_t^T \mathbf{R}_t \boldsymbol{\eta}_t + \varrho \quad (\text{A.6.2a})$$

$$\text{s.t. } \boldsymbol{\eta} = \tilde{\mathbf{A}}\mathbf{Y} + \tilde{\mathbf{P}}, \quad (\text{A.6.2b})$$

$$|\tilde{p}_{it} - p_{it}| \leq \varrho \varpi_{it}, \quad (\text{A.6.2c})$$

$$\varrho \leq 0.3 \quad (\text{A.6.2d})$$

$$\tilde{a}_{ij} = 0 \quad \text{if } s_{ij} = 0, \quad (\text{A.6.2e})$$

$$\tilde{a}_{ij} \geq 0.001 \quad \text{if } s_{ij} = +, \quad (\text{A.6.2f})$$

$$\tilde{a}_{ij} \leq -0.001 \quad \text{if } s_{ij} = -, \quad (\text{A.6.2g})$$

$$\tilde{a}_{ij} \in \mathbb{R} \quad \text{if } s_{ij} = ?. \quad (\text{A.6.2h})$$

Here  $\hat{\mathbf{A}}$  denotes the estimated interaction matrix,  $\hat{\mathbf{P}}$  the estimated perturbation matrix,  $\mathbf{P} = \mathbf{I}$  the observed perturbation matrix,  $\mathbf{Y}$  the observed response matrix in Table A.5 for the Glucose data and Table A.6 for the Galactose data, and  $\varpi_{it}$  the upper bounds on the perturbation errors in Table A.4. The weight  $\mathbf{R}_t = \left( \hat{\mathbf{A}}_{ols} \text{diag}(\mathbf{w}_t) \hat{\mathbf{A}}_{ols}^T \right)^{-1}$ , with  $\hat{\mathbf{A}}_{ols} = -\mathbf{P}\mathbf{Y}^\dagger$  and  $w_{it} = \varpi_{it} / \sum_{i,t} \varpi_{it}$  using the upper bounds  $\varpi_{it}$  in Table A.2 for the Glucose data and in Table A.3 for the Galactose data. The Moore-Penrose generalized inverse  $\mathbf{Y}^\dagger$  is used to calculate the ordinary least-squares estimate  $\hat{\mathbf{A}}_{ols}$ . This optimization problem is convex and we solve it using CVX ([cvxr.com/cvx](http://cvxr.com/cvx)) in Matlab ([www.mathworks.com](http://www.mathworks.com)). The bound on  $\varrho$  was tuned such that a reasonable trade off between perturbation errors and response errors were achieved and the estimated parameters were feasible. The estimated parameters of the chosen representatives of the set of interaction matrices, which can explain the Glucose or Galactose steady-state data recorded by Cantone et al. (2009), are presented in Table A.9. Note that we had to tweak the weights  $w_{it}$  a bit, more precisely divide  $w_{41}$  by 2.5 and  $w_{42}$  by 1.5, to obtain the parameter estimates in  $\hat{\mathbf{A}}_{\text{Glb}}$ .

The network that we presented in the concluding example in Section 4.6 has an

interaction matrix,

$$\hat{\mathbf{A}}_{\text{Glu}} = \begin{bmatrix} -0.143 & 0.043 & 0.073 & 0 & 0 \\ 0 & -0.183 & 0 & 0 & 0 \\ 0.103 & 0.091 & -0.219 & 0 & 0 \\ 0.036 & 0 & 0 & -0.177 & 0 \\ 0 & 0 & 0.009 & 0 & -0.152 \end{bmatrix}, \quad (\text{A.6.3})$$

containing the 11 interactions that are most likely to exist based on our confidence scores in Table A.7. The total residual sum of squares for responses and perturbations is for this model 0.82, which is small compared to the total sum of squares for the response and perturbation matrices, 270, so the model explains the data well. The data is not informative enough to show that any interaction with  $\gamma < 1$  exists, but we selected to include these interactions since it is unlikely that the real network only has six interactions. This model has 8 links in common with the engineered model, three additional links not found in the engineered one, and lacks 5 links present in the engineered one. Two of the lacking interactions correspond to the Gal4-Gal80 protein interactions (Bhat and Murthy, 2001; Sellick et al., 2008), which strictly speaking should not have been included in the engineered model in the first place, since these protein interactions do not affect transcription of the *GAL4* or *GAL80* gene. To illustrate that it is only one of several alternative models with different selections of links having  $\gamma < 1$  that can explain the data we estimated two alternative networks. First, one with interaction matrix

$$\hat{\mathbf{A}}_{\text{Glb}} = \begin{bmatrix} -0.149 & 0 & 0.099 & 0 & 0 \\ 0.004 & -0.174 & 0 & 0 & 0 \\ 0.098 & 0.134 & -0.229 & 0 & -0.001 \\ 0 & -0.004 & 0.044 & -0.155 & 0 \\ 0.006 & 0 & 0 & 0 & -0.148 \end{bmatrix}. \quad (\text{A.6.4})$$

Then one with interaction matrix

$$\hat{\mathbf{A}}_{\text{Gla}} = \begin{bmatrix} -0.153 & 0 & 0.103 & 0 & 0 \\ 0 & -0.173 & 0 & 0 & 0 \\ 0.103 & 0.133 & -0.232 & 0 & -0.001 \\ 0.038 & 0 & -0.002 & -0.178 & 0 \\ 0 & 0 & 0 & 0 & -0.151 \end{bmatrix} \quad (\text{A.6.5})$$

to show that the same structure can be used both explain the Glucose and

Galactose data, as evident by the following interaction matrix

$$\hat{\mathbf{A}}_{\text{Gal}} = \begin{bmatrix} -0.349 & 0 & 0.237 & 0 & 0 \\ 0 & -0.227 & 0 & 0 & 0 \\ 0.196 & 0.179 & -0.407 & 0 & -0.137 \\ 0.073 & 0 & 0.052 & -0.205 & 0 \\ 0 & 0 & 0 & 0 & -0.141 \end{bmatrix}, \quad (\text{A.6.6})$$

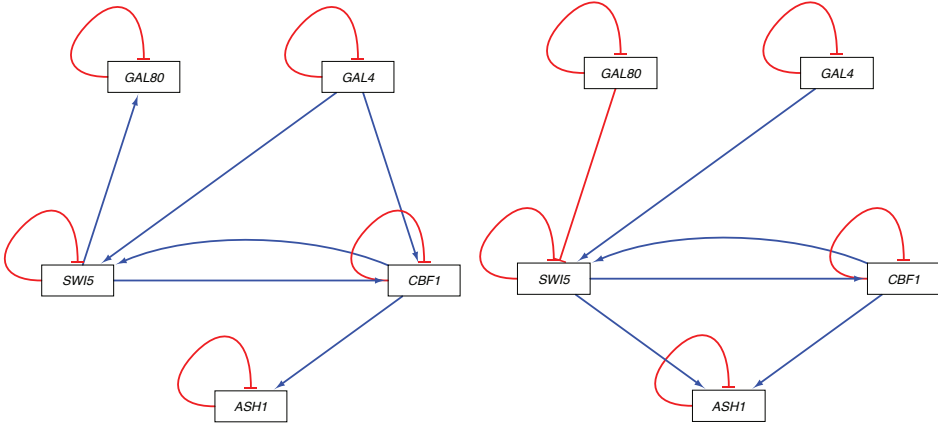
containing the 11 interactions that are most likely to exist based on our confidence scores in Table A.8. It is worth noting that the eigenvalues of all four matrices are negative, so these systems are stable. The degree of interampattiness of all of them are between 3.8 and 6.5. Also note that *ASH1* and *GAL80* does not have any influence on any of the other genes, which makes sense considering that the other genes shows almost no expression change when these two are perturbed, see Tables A.5 and A.6. The network graphs of  $\hat{\mathbf{A}}_{\text{Glu}}$  and  $\hat{\mathbf{A}}_{\text{Gal}}$  are shown in Figure A.5 and the graph of the network Cantone *et al.* intended to engineer in Figure A.6 for comparison.

## A.7 Additional information on the TOEL example

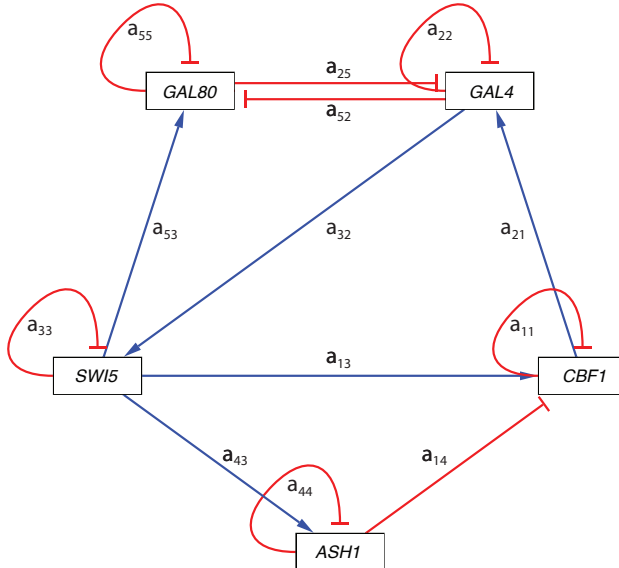
We here describe the two gene *in silico* example, which is used for illustration of interampattiness, uncertainty sets, partly informative data, robust variable selection for network inference, and the importance of designing additional perturbation experiments throughout this thesis. Consider a gene regulatory network with two genes, *A* and *B* in Figure A.7 (top, middle), which we call the TwO gene ExampLe (TOEL). These genes are connected in a positive feedback loop, consisting of two negative interactions, *i.e.* both genes *A* and *B* down-regulates each other. Both genes also have negative self-loops accounting for self degradation. The interactions are for simplicity here described using a linear model, since we only are interested in the behaviour of the system near a steady-state, *i.e.* a particular physiological condition. The “true” interaction matrix of the network is

$$\check{\mathbf{A}} = \begin{bmatrix} -4.608 & -5.544 \\ -2.856 & -4.608 \end{bmatrix} \quad (\text{A.7.1})$$

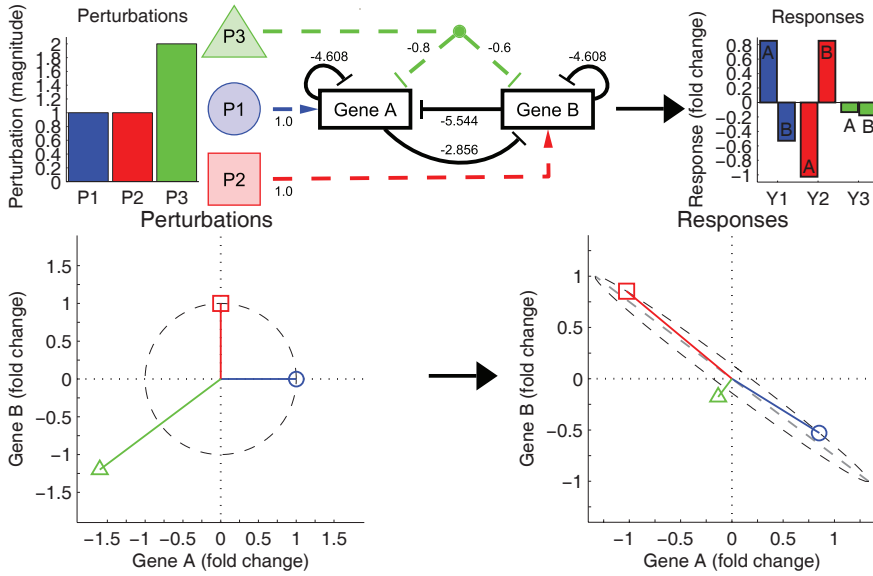
and the network is depicted together with the state-space model in Figure 3.3. The degree of interampattiness of the two gene system is 15, implying that a perturbation in the direction of maximum amplification, *i.e.* exciting the strong mode, will yield a response that is 15 times stronger than that of a perturbation in the direction of maximum attenuation, *i.e.* the weak mode, see Chapter 3. If we, *e.g.*, perform three steady-state experiments where we apply the external perturbations P1, P2 and P3, shown in Figure A.7 top left, then we obtain the responses Y1, Y2 and Y3, shown top right. The corresponding regressor and regressand matrices are



**Figure A.5: The inferred networks.** The structure of the network inferred from the Glucose steady-state data (left) and Galactose steady-state data (right) recorded by Cantone et al. (2009) using our theory for robust variable selection followed by estimation of the nonzero parameters by constrained optimisation. The strength of the interactions, *i.e.* estimates of the nonzero parameters, are presented in Table A.9 for Glucose ( $\hat{A}_{\text{Glu}}$ ) and Galactose ( $\hat{A}_{\text{Gal}}$ ). This figure was generated in Cytoscape (Shannon et al., 2003).



**Figure A.6: The engineered IRMA network.** The structure of the network with the corresponding element of the interaction matrix  $\mathbf{A}$  marked on each edge. An arrow shaped head on the link indicates up-regulation or activation, while a bar shaped head indicates down-regulation or repression. This network graph was generated in Cytoscape (Shannon et al., 2003).



**Figure A.7: Illustration of steady-state perturbation experiments.** The two gene TOEL network (top, middle). An arrow shaped head on the link indicates up-regulation or activation, while a bar shaped head indicates down-regulation or repression. The dashed lines from the external perturbations P1, P2 and P3 represent different signal pathways. The bar diagram (top, left) shows the strength of the external perturbations, while the bar diagram (top, right) shows the steady-state response of both genes in each of the three perturbation experiments. The same perturbations, denoted by the same marker and colour, is shown in the gene space (bottom, left), with corresponding steady-state responses (bottom, right).

$$\Phi = \begin{bmatrix} 0.8533 & -0.5289 \\ -1.027 & 0.8533 \\ -0.1333 & -0.1778 \end{bmatrix} \quad \text{and} \quad \Xi = \begin{bmatrix} -1 & 0 \\ 0 & -1 \\ 1.6 & 1.2 \end{bmatrix}. \quad (\text{A.7.2})$$

These are not corrupted by addition of noise, but we use them as the observed ones in order to keep our illustrations simple and pedagogical. Note that the third perturbation has twice the strength of the one used in Box 1 (Section 3.1). Here the responses are fold changes relative to the steady-state mRNA concentration of the unperturbed system in the physiological condition of interest. These are given by multiplying the perturbation matrix by the static gain matrix, *i.e.* the inverse of the interaction matrix, which yields the response matrix (2.3.3). To concretize, let us assume that genes *A* and *B* regulate two alternative metabolic pathways yielding the same product but based on different nutrients, then P1 can be viewed as an increase in the concentration of one of the nutrients, which through a signal cascade up-regulates gene *A*. After some time, when the system reaches steady-state, the expression of gene *A* is increased, while that of gene *B* is decreased, and the production has adapted to the change of nutrients. We are not interested in the signal pathways here, so we could instead represent the external perturbation by its direct effect on the two genes, which is best illustrated in what we call the gene space, shown in Figure A.7 (bottom, left: circle, square, triangle). In the gene space, the fold change of each gene is represented on a different coordinate axis, so that each experiment corresponds to a point in the figure. Similarly, we can study the response of the system directly in the gene space (bottom, right). An advantage of the gene space is that the magnitude and direction of each perturbation and response is directly visible as the length and direction of the corresponding vector. In this example perturbation P2 is amplified 1.3 times, such that the expression of gene *B* is increased, while that of gene *A* is decreased. Perturbation P3 is attenuated 9 times, *i.e.* the magnitude of the response is one ninth of the applied perturbation. The first two responses are almost on a line, which the third is orthogonal to. The third response is significantly weaker despite corresponding to a perturbation of twice the magnitude of the first two. All of these are due to the interampatteness of the system.

The system contains only two genes so in theory the first two perturbation experiments are sufficient for generating informative data for network inference. However, if we consider uncertainty in the applied perturbations and measurements of the response, then already modest levels of uncertainty hamper the inference, as we show next. The effect of uncertainty is best illustrated in the data space, in which each variable, *i.e.* gene, is a point and the fold change in each experiment is represented on a different coordinate axis. The data for inference of the first row of the interaction matrix from the first two experiments is depicted in Figure 5.1. The regressors  $\phi_1$  and  $\phi_2$  contains the responses in genes *A* and *B*, respectively. While the regressand  $\xi_1$  contains the perturbations of gene *A*. Note that two different error models, a deterministic and a stochastic model, associated with different uncertainty

sets, described in (5.4.1), (5.4.2), (5.4.3), and (5.4.4), are depicted in the figure. Both error models, however, give rise to uncertainty sets of the same size. The effect of this uncertainty is further investigated for the deterministic example in Figure 5.2, where we show that only one of the interactions can be classified as existing using the conditions summarized in Table 5.1. The data from these two perturbation experiments is in other words not informative enough for inference of the structure of the TOEL network since three of the four interactions are classified as alternative, meaning that they are both present and absent in some models that can explain the data. This is due to the overlap of the uncertainty cones of the regressand and regressors in Figure 5.2. Also the data from the third perturbation experiment is needed to robustly infer all interactions of the TOEL network. This is demonstrated in Figure A.8, where neither of the uncertainty cones of the regressands overlap with neither of the uncertainty cones of the regressors, except for the intersection at the origin, and all interactions are therefore shown to exist. For the third perturbation experiment we assume that the uncertainty in the measurement of gene *A* decreased to 0.1 and gene *B* to 0.04, while the uncertainty in the perturbation of both genes increased to 0.8, reflecting improvements in the measurements and the difficulty to perturb both genes simultaneously. This implies that the uncertainty set of both regressors are ellipsoids instead of spheres, which they would have been if the uncertainty had remained the same as in the first two experiments. The nominal uncertainty sets of the regressors corresponding to significance level  $\alpha = 0$  are hence

$$\mathcal{U}_{\phi_j} = \{\mathbf{v}_j \mid \mathbf{v}_j^T \mathbf{Q}_{\phi_j} \mathbf{v}_j \leq 1\}, \quad (\text{A.7.3})$$

with

$$\mathbf{Q}_{\phi_1} = \begin{bmatrix} 11.111 & 0 & 0 \\ 0 & 11.111 & 0 \\ 0 & 0 & 100 \end{bmatrix} \quad \text{and} \quad \mathbf{Q}_{\phi_2} = \begin{bmatrix} 156.25 & 0 & 0 \\ 0 & 156.25 & 0 \\ 0 & 0 & 625 \end{bmatrix}. \quad (\text{A.7.4})$$

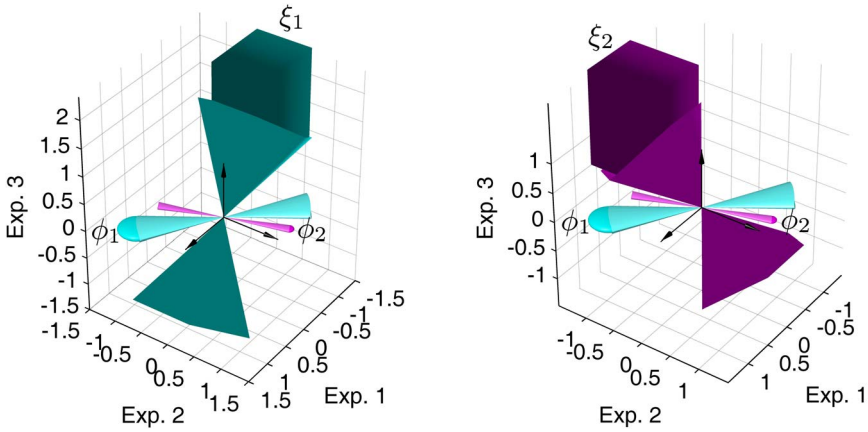
The nominal uncertainty sets of the regressands are

$$\begin{aligned} \mathcal{U}_{\xi_i} &= \{\boldsymbol{\epsilon}_i \mid \boldsymbol{\epsilon}_i \triangleq [\epsilon_{1i}, \dots, \epsilon_{ki}, \dots, \epsilon_{mi}]^T, \\ &\quad \text{with } |\epsilon_{ki}| \leq \varpi_{ki}, \quad k \in \{1, 2, \dots, m\}\}, \end{aligned} \quad (\text{A.7.5})$$

with  $\varpi_{11} = \varpi_{22} = 0.6$ ,  $\varpi_{21} = \varpi_{12} = 0.5$ , and  $\varpi_{31} = \varpi_{32} = 0.8$ .

## A.8 Robust inference illustrated on the TOEL example

Here we show in detail how to use the structured singular value method described in Section 5.10 to determine if the set of regressors is practical independent for the TOEL example in Appendix A.7 with three different error models. In the first case, we assume that the error in each regressor is deterministic with bounds on the



**Figure A.8: Illustration of informative data for network inference.** The three steady-state perturbation experiments that yielded the data are illustrated in Figure A.7, the deterministic uncertainty sets are explained in Figure 5.1 (left). Data for inference of the first row of the interaction matrix  $\mathbf{A}$  is shown left, while data for the second row is shown right. The cyan coloured ellipsoid at the end of the cyan cone depicts the uncertainty set of regressor  $\phi_1$  and the magenta coloured ellipsoid at the end of the magenta cone depicts the uncertainty set of regressor  $\phi_2$ , with the cones representing their uncertainty cones. While the rectangular cuboid at the end of both the teal coloured cone (left) and purple coloured cone (right) depicts the uncertainty set of the regressand  $\xi_1$  (left) and  $\xi_2$  (right). None of the cones overlap, so both regressors are practically selectable and the corresponding interactions are classified as existing based on the results in Table 5.1, which implies that the data is informative for robust inference of the adjacency matrix of the two gene TOEL network. Actually this data is even informative for correct inference of the signed adjacency matrix, since the regressors are practically assignable and the corresponding interactions classified as repressing. As a comparison, note that the data obtained in the first two experiments, Figure 5.2, only contained enough information to show that the  $a_{11}$  interaction exists.

induced 2-norm. In the second case, we also assume that the errors are deterministic, but that the absolute value of the error in each element of the regressors is bounded. In the third case, we assume a stochastic error model in which the errors in each regressor are jointly normally distributed.

The set of regressors is practically independent (Definition 5.5.13) if no realization exists within the uncertainty sets such that the regressor matrix is of lower rank. According to Theorem 5.10.1, it is for any error model possible to show that the set of regressors is practically independent at a desired significance level by finding ellipsoidal supersets of the uncertainty sets, containing all realizations that cannot be rejected by the error model at the desired significance level, such that the confidence score is greater or equal to one. The confidence score  $\gamma(\Phi)$  is the smallest factor by which these ellipsoidal uncertainty sets need to be scaled such that rank deficient realizations of  $\Phi$  only exist on the boundary of the sets. This test of practical independence is however in general conservative, so failure to do so does not in general imply that a rank deficient realization is consistent with the error model. We therefore for the second and third case also use an exact test.

In the first case the uncertainty sets are circular and contain the “true” realization, since a deterministic error model with bounds on the induced 2-norm of the error in each regressor is assumed. The uncertainty circles are specified in (5.4.1) and depicted in Figure 5.1 left. From Figure 5.2, we already know that many rank deficient realizations exist within the uncertainty sets, since the uncertainty cones overlap and many parallel realizations of the two regressors exist. The confidence score is therefore of course smaller than one when the structured singular value method is used to calculate the distance to rank deficiency from the regressor matrix in (A.7.2) with column structure on the uncertainty. The uncertainty sets are circular, which is a special case of an ellipsoid in two dimensions, but of different size, so scaling of the problem is needed to convert them into unit circles before the `mussv` function in Matlab ([www.mathworks.com](http://www.mathworks.com)) can be used to calculate the structured singular value and  $\gamma$  bounds. The purpose of this case is to demonstrate this scaling and show that the uncertainty cones only will intersect at the origin when the uncertainty circles are multiplied by  $\gamma$ . This scaling is done by multiplying the pseudo inverse of the observed regressor matrix in (A.7.2) by the right transformation matrix, defined in (5.10.3),

$$\mathbf{W}_1 = \begin{bmatrix} 0.3 & 0 & 0.08 & 0 \\ 0 & 0.3 & 0 & 0.08 \end{bmatrix}. \quad (\text{A.8.1})$$

The block diagonal structure of  $\Delta$  is specified by `blk = [2 1; 2 1]`, *i.e.* two real column matrices of size  $2 \times 1$ , when calling `[mu,muinfo]=mussv(Phi†W1,blk)`. The first returned variable `mu=[2.1990, 2.1990]` contains the upper and lower bound of the structured singular values, while `muinfo` is a structure containing among other things a  $\Delta$  matrix such that  $\det(\mathbf{I} - \Phi^\dagger \mathbf{W}_1 \Delta) = 0$ . The lower bound on confidence score  $\gamma(\Phi)$  is 0.4547 and the error realization that makes the regressor

matrix rank deficient is

$$\tilde{\mathbf{Y}}_s = \mathbf{W}_1 \mathbf{\Delta} = \begin{bmatrix} -0.1133 & -0.03021 \\ -0.07602 & -0.02027 \end{bmatrix}. \quad (\text{A.8.2})$$

The uncertainty sets obtained by scaling the nominal ones by the lower bound on  $\gamma$  and the closest rank deficient realization are shown in Figure A.9. The uncertainty cones almost touch and the closest rank deficient realization is just outside the uncertainty cones, illustrating that the confidence score for deterministic error models with ellipsoidal bounds on the error in each regressor is tight. In other words, if we scale the uncertainty sets by any factor larger than the confidence score, then a rank deficient realization will be contained in the scaled sets. The confidence score is hence for deterministic error models with ellipsoidal bounds on the error in each regressor proportional to the significance level  $\alpha$  that the scaled uncertainty sets correspond to. Note that the uncertainty cones, however, contain realizations belonging to the nominal uncertainty sets that are outside of the scaled sets but practically independent, so the actual p-value is smaller than  $\alpha$ .

In the second case, we also assume that the errors are deterministic, but that the absolute value of the error in each element of the regressors is bounded instead of the induced 2-norm of the errors in each regressor. More specifically, we assume the following uncertainty sets

$$\begin{aligned} \mathcal{U}_{\phi_j} = \{ \mathbf{v}_j \mid \mathbf{v}_j \triangleq [v_{1j}, \dots, v_{kj}, \dots, v_{mj}]^T, \\ \text{with } |v_{kj}| \leq \varpi_{kj}, k \in \{1, 2, \dots, m\} \}, \end{aligned} \quad (\text{A.8.3})$$

with  $\varpi_{11} = \varpi_{21} = 0.0983$  and  $\varpi_{12} = \varpi_{22} = 0.0262$ . To test if the set of regressors is practically independent by the structured singular value method, according to Theorem 5.10.1, we first need to determine an ellipsoidal superset of each uncertainty square. In principle, any such supersets can be used, but the bigger the sets are the more conservative the obtained confidence score is. We therefore find the minimal supersets, which have a radius equal to half the length of the diagonal of the corresponding uncertainty square,

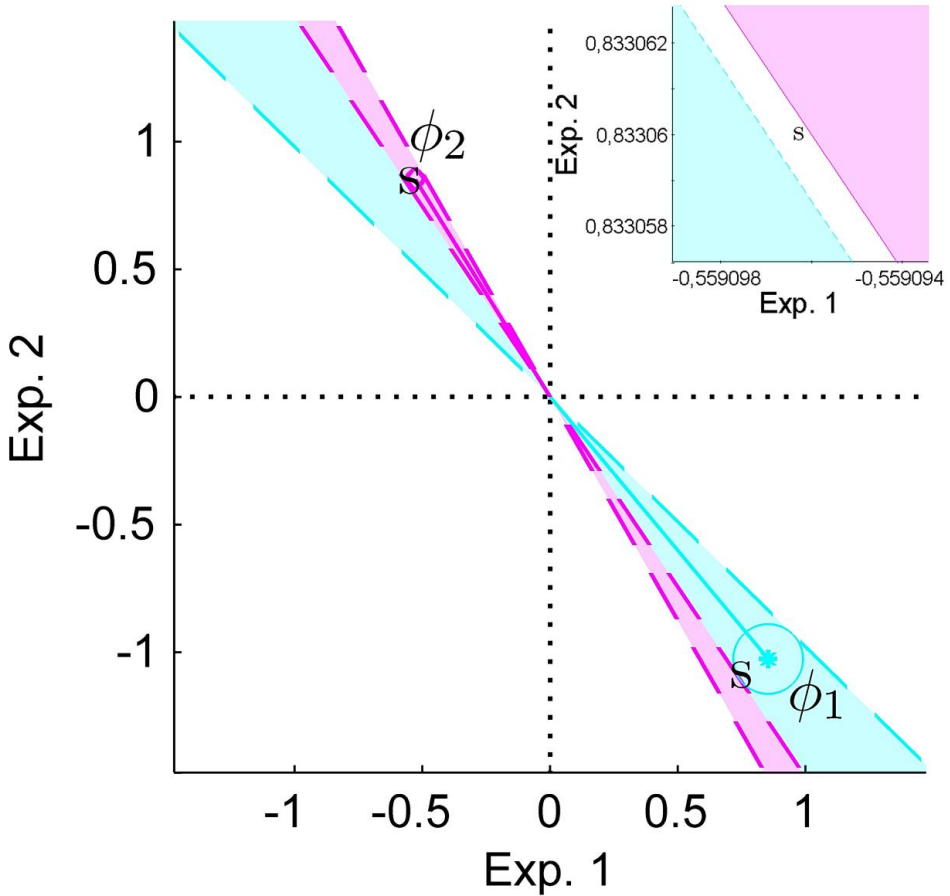
$$\mathcal{U}_{\phi_j}^{\text{ell}} = \{ \mathbf{v}_j \mid \mathbf{v}_j^T \mathbf{Q}_{\phi_j} \mathbf{v}_j \leq 1 \}, \quad (\text{A.8.4})$$

with

$$\mathbf{Q}_{\phi_1} = \begin{bmatrix} 51.74 & 0 \\ 0 & 51.74 \end{bmatrix} \quad \text{and} \quad \mathbf{Q}_{\phi_2} = \begin{bmatrix} 728.4 & 0 \\ 0 & 728.4 \end{bmatrix}. \quad (\text{A.8.5})$$

Scaling of the pseudo inverse of the observed regressor matrix in (A.7.2) by the right transformation matrix, defined in (5.10.3),

$$\mathbf{W}_1 = \begin{bmatrix} 0.139 & 0 & 0.03705 & 0 \\ 0 & 0.139 & 0 & 0.03705 \end{bmatrix} \quad (\text{A.8.6})$$



**Figure A.9:** Illustration of scaling of the uncertainty sets by the confidence score  $\gamma$  for deterministic errors bounded in each regressor. The nominal uncertainty sets of both regressors, depicted in Figure 5.2, have been scaled by  $\gamma(\Phi) = 0.4547$  to obtain the circular uncertainty sets shown here. No rank deficient realization of  $\Phi$  exists within these uncertainty sets, since the corresponding uncertainty cones only intersect in the origin. The set of regressors is hence practically independent at the significance level corresponding to this confidence score. The closest rank deficient realization is marked by  $s$  and lies just outside of the uncertainty sets (visible in the enlargement of the boundary of the uncertainty circle of  $\phi_2$  in the upper right corner), illustrating that the confidence score is tight for deterministic error models with ellipsoidal uncertainty bounds on each variable.

is also in this case needed to convert this problem into one with unit circles before using the `mussv` function. The lower bound  $\gamma(\Phi) = 0.9815$  is obtained by calling `[mu,muinfo]=mussv(Phi^dagger*W_1,blk)` with the same block diagonal structure on  $\Delta$  as in the first case above and taking the reciprocal of the returned upper bound on the structured singular value  $\mu$ . This value is below one and we used the minimal ellipsoidal superset, so we cannot show that the set of regressors is practically independent based on Theorem 5.10.1. Many rank deficient realizations of the regressor matrix are contained within the constructed circular uncertainty sets, *e.g.* the following error realization that besides numerical round-off is equal to (A.8.2) makes the regressor matrix rank deficient

$$\tilde{\Upsilon}_s = \mathbf{W}_1 \Delta = \begin{bmatrix} -0.1133 & -0.0302 \\ -0.07603 & -0.02027 \end{bmatrix}. \quad (\text{A.8.7})$$

Visual inspection of Figure A.10 where the nominal uncertainty sets and their uncertainty cones are plotted, however, reveals that the set of regressors is practically independent since these uncertainty cones only intersect at the origin. This illustrates that the confidence score obtained by the structured singular value method with column uncertainty is conservative unless the errors in each regressors are bounded by an ellipsoid in the assumed error model. A less conservative confidence score  $\gamma_e$  can, however, at least when the number of experiments and variables is equal and the errors are assumed to be bounded in each element, be calculated by the structured singular value method with element-wise uncertainty instead of column uncertainty. This confidence score is obtained by solving

$$\gamma_e(\Phi) \triangleq \frac{1}{\mu_e(\Phi)} \triangleq \min_{\Delta} \{ \bar{\sigma}(\Delta) \mid \det(\mathbf{I} - \mathbf{W}_2 \Phi^\dagger \mathbf{W}_1 \Delta) = 0 \\ \text{for } \Delta = \text{diag}(\delta_{11}, \delta_{21}, \dots, \delta_{kj}, \dots, \delta_{mn}) \text{ with } \delta_{kj} \in \mathbb{R} \}. \quad (\text{A.8.8})$$

Here the right transformation matrix of  $\Phi^\dagger$  is defined as

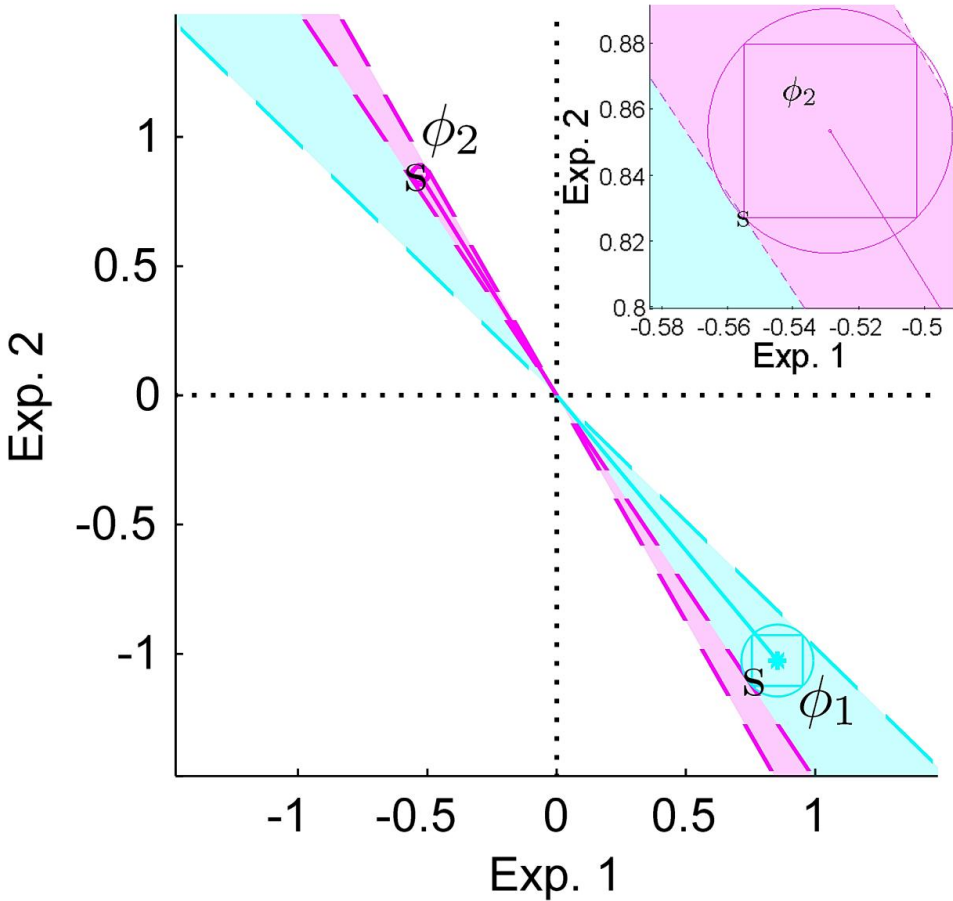
$$\mathbf{W}_1 \triangleq [\mathbf{I}_{m \times m}, \dots, \mathbf{I}_{m \times m}] \text{diag}(\varpi_{11}, \varpi_{21}, \dots, \varpi_{kj}, \dots, \varpi_{mn}) \quad (\text{A.8.9})$$

and the left one as

$$\mathbf{W}_2 \triangleq \text{diag}(\mathbf{1}_{m \times 1}, \dots, \mathbf{1}_{m \times 1}), \quad (\text{A.8.10})$$

with  $\mathbf{I}_{m \times m}$  be an  $m \times m$  identity matrix and  $\mathbf{1}_{m \times 1}$  a column vector of length  $m$  with all elements one. These are for this case

$$\mathbf{W}_1 = \begin{bmatrix} 0.0983 & 0 & 0.0262 & 0 \\ 0 & 0.0983 & 0 & 0.0262 \end{bmatrix} \quad (\text{A.8.11})$$



**Figure A.10: Illustration of practical independence testing for deterministic errors bounded in each element.** The uncertainty cones corresponding to the nominal uncertainty squares of each regressor only intersects at the origin, so this set of regressors is practically independent. But the practical independence test in Theorem 5.10.1 cannot be used to show this since the minimal ellipsoidal supersets of the nominal uncertainty sets contain realizations outside of these uncertainty cones and their corresponding cones would therefore overlap. This is seen in the enlargement of the uncertainty set of the second regressor in the upper right, where the circular superset extends beyond the uncertainty cone. The confidence score  $\gamma$  given by the structured singular value method with column uncertainty is conservative due to the use of supersets, but the same method with element-wise uncertainty in  $\Delta$  gives a confidence score  $\gamma_e$  that is tight. This is illustrated by the closest rank deficient realization based on element-wise uncertainty, marked by s, being just outside of the uncertainty sets.

and

$$\mathbf{W}_2 = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}. \quad (\text{A.8.12})$$

A lower bound on this confidence score is obtained by calling `[mu,muinfo]=mussv(W2Φ†W1,blk)` with the element-wise diagonal structure of  $\mathbf{\Delta}$  specified by `blk = [1 1;1 1;1 1;1 1]` and taking the reciprocal of the upper bound on the structured singular value  $\mu$ . In this case, the lower bound on the confidence score  $\gamma_e(\Phi)$  is 1.0003 and the error realization that makes the regressor matrix rank deficient is

$$\tilde{\mathbf{Y}}_s = \mathbf{W}_1 \mathbf{\Delta} \mathbf{W}_2 = \begin{bmatrix} -0.09833 & -0.02621 \\ -0.09833 & -0.02621 \end{bmatrix}. \quad (\text{A.8.13})$$

This shows that the set of regressors is practically independent, since  $\gamma_e(\Phi) > 1$ .

In the third case, we assume a stochastic error model in which the errors in each regressor are jointly normally distributed with covariance matrix  $\mathbf{\Lambda}_\mathbf{Y} = \text{diag}(0.00024, 0.00315, 0.00024, 0.00315)$ . In other words, roughly 3.6 times larger error in the observations of both genes in the second experiment than in the first experiment. We do this in order to illustrate scaling of the experiments for normally distributed errors. To test if the set of regressors is practically independent by the structured singular value method according to Theorem 5.10.1, we first select the significance level  $\alpha = 0.05$  and generate ellipsoidal uncertainty set for each regressor as defined in (5.4.3). These will contain all realizations that cannot be rejected based on a  $\chi^2$  hypothesis test at significance level  $\alpha = 0.05$ , but also realizations that are rejected as shown in Section 5.4. Together these uncertainty sets form a superset of the set of realizations that are consistent with the multivariate normal distribution. We use them since the consistent realization of one regressor depends on the realization of the other, which is difficult to illustrate graphically in the data space. The generated sets,

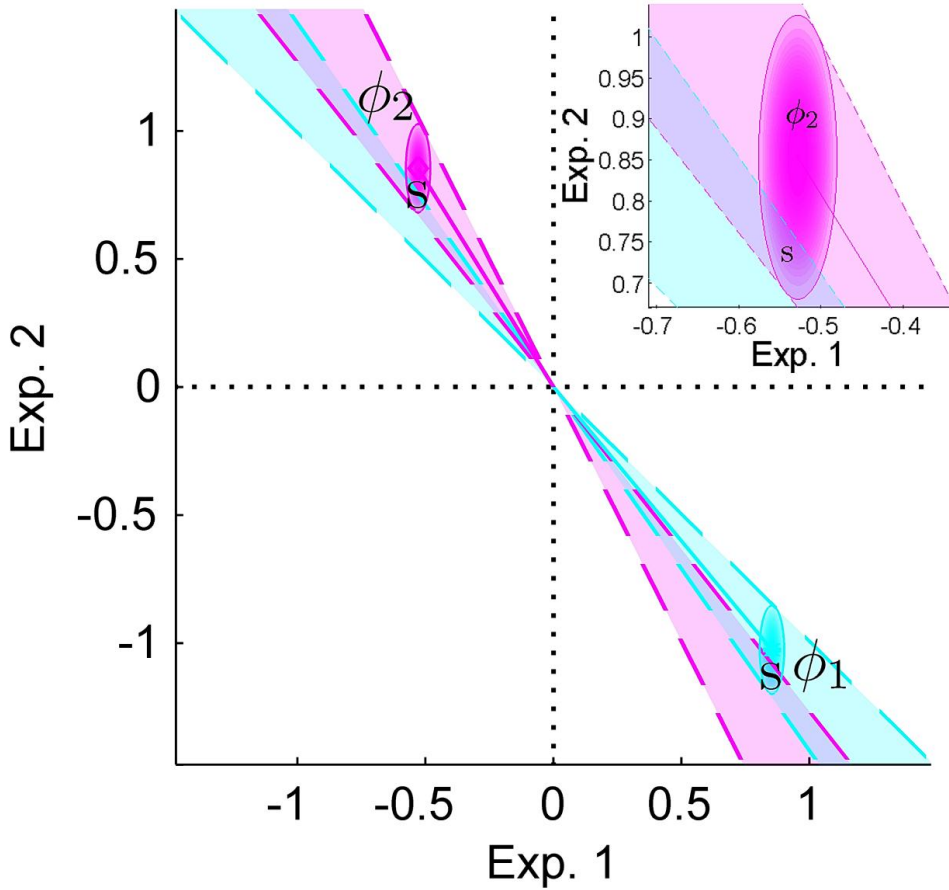
$$\mathcal{U}_{\phi_j}^{0.05} = \{\mathbf{v}_j \mid \mathbf{v}_j^T \mathbf{Q}_{\phi_j} \mathbf{v}_j \leq 1\}, \quad (\text{A.8.14})$$

with

$$\mathbf{Q}_{\phi_1} = \mathbf{Q}_{\phi_2} = \frac{1}{\chi^2(0.05, 4)} \mathbf{\Lambda}_{v_1}^{-1} = \begin{bmatrix} 439.2 & 0 \\ 0 & 33.46 \end{bmatrix}, \quad (\text{A.8.15})$$

are depicted in Figure A.11. Scaling of the Moore-Penrose generalized inverse of the observed regressor matrix in (A.7.2) by the right transformation matrix, defined in (5.10.3),

$$\mathbf{W}_1 = \begin{bmatrix} 0.04772 & 0 & 0.04772 & 0 \\ 0 & 0.1729 & 0 & 0.1729 \end{bmatrix} \quad (\text{A.8.16})$$



**Figure A.11: Illustration of practical independence testing for stochastic errors assumed to be jointly normally distributed.** The uncertainty cones corresponding to the uncertainty ellipses of each regressor overlap, so the practical independence test in Theorem 5.10.1 cannot be used to show that this set of regressors is practically independent at significance level  $\alpha = 0.05$  even though it is. The closest rank deficient realization based on column uncertainty, marked by  $s$ , is just one of many rank deficient realizations within these uncertainty sets, as evident in the enlargement in the upper right corner.

is also in this case needed to convert this problem into one with unit circles before using the `mussv` function. The lower bound  $\gamma(\Phi) = 0.7255$  is obtained by calling `[mu,muinfo]=mussv(Phi^dagger*W_1,blk)` with the same block diagonal structure on  $\Delta$  as in the first case above and taking the reciprocal of the returned upper bound on the structured singular value  $\mu$ . This value is below one and we used the minimal ellipsoidal supersets that contains all non-rejectable realizations, so we failed to show that the set of regressors is practically independent at  $\alpha = 0.05$  based on Theorem 5.10.1. The uncertainty cones overlap so many rank deficient realizations of the regressor matrix are contained within the generated uncertainty sets, *e.g.* the following error realization that is marked by `s` in Figure A.11 makes the regressor matrix rank deficient

$$\tilde{\mathbf{Y}}_s = \mathbf{W}_1 \Delta = \begin{bmatrix} -0.01217 & -0.01217 \\ -0.1174 & -0.1174 \end{bmatrix}. \quad (\text{A.8.17})$$

This rank deficient realization is however rejected based on the  $\chi^2$  hypothesis test at the selected significance level despite clearly belonging to the generated uncertainty sets, so it illustrates that this confidence score is conservative. A less conservative confidence score  $\gamma_f$  can, however, at least when the errors are assumed to be jointly normally distributed with covariance matrix  $\mathbf{I}$ , be calculated by the singular value decomposition based on Theorem 5.10.2. In this case, however, we have different variance in each experiment, so we need to scale the regressor matrix. Therefore, we next modify the definitions in Theorem 5.10.2 to suit cases with the same variance in each variable for independent experiments with different variance. The smallest nonzero singular value  $\sigma_n$  gives the distance to the closest matrix of lower rank without imposing any structure on the uncertainty, so we define the confidence score as

$$\gamma_f(\Phi) \triangleq \sigma_n(\mathbf{W}_1^{-1} \Phi), \quad (\text{A.8.18})$$

based on the SVD of a matrix defined in (5.10.5). Here the scaling of the regressor matrix by

$$\mathbf{W}_1 \triangleq \sqrt{\chi^{-2}(\alpha, nm)} \mathbf{\Lambda}_{v_1} \quad (\text{A.8.19})$$

is needed to account for the difference in variance between the experiments and to normalize the confidence score to one for all realizations  $\tilde{\mathbf{Y}}$  such that  $\tilde{\mathbf{Y}}^T \mathbf{\Lambda}_{\tilde{\mathbf{Y}}}^{-1} \tilde{\mathbf{Y}} = \chi^{-2}(\alpha, nm)$ . This weight corresponds to the right transformation matrix of  $\Phi^\dagger$  in the previous cases and this confidence score can alternatively be calculated by solving

$$\gamma_f(\Phi) = \frac{1}{\mu_f(\Phi)} \triangleq \min_{\Delta} \{ \bar{\sigma}(\Delta) \mid \det(\mathbf{I} - \Phi^\dagger \mathbf{W}_1 \Delta) = 0 \text{ for } \Delta \in \mathbb{R}^{m \times n} \}, \quad (\text{A.8.20})$$

e.g. by calling `[mu,muinfo]=mussv(Phi^+W_1,blk)` with the full block structure of  $\Delta$  specified by `blk = [m, n]`. The scaling matrix is in this case

$$\mathbf{W}_1 = \begin{bmatrix} 0.04772 & 0 \\ 0 & 0.1729 \end{bmatrix}, \quad (\text{A.8.21})$$

the confidence score  $\gamma_f(\Phi)$  is 1.0027, and the corresponding error realization that makes the regressor matrix rank deficient is

$$\tilde{\mathbf{Y}}_s = \mathbf{W}_1 \gamma_f \mathbf{u}_n \mathbf{v}_n^T = \begin{bmatrix} -0.008855 & -0.01378 \\ -0.08804 & -0.137 \end{bmatrix}. \quad (\text{A.8.22})$$

This rank deficient realization is with  $\tilde{\mathbf{Y}}^T \Lambda_{\tilde{\mathbf{Y}}}^{-1} \tilde{\mathbf{Y}} = 9.5381 > 9.4877 = \chi^2(0.05, 4)$  rejected based on the  $\chi^2$  hypothesis test at the selected significance level, which shows that the set of regressors is practically independent, since this is the closest matrix of lower rank measured by the induced 2-norm and

$$\begin{aligned} \tilde{\mathbf{Y}}^T \Lambda_{\tilde{\mathbf{Y}}}^{-1} \tilde{\mathbf{Y}} &= \tilde{\mathbf{Y}}^T \Lambda_{\tilde{\mathbf{Y}}}^{-\frac{1}{2}T} \underbrace{\Lambda_{\tilde{\mathbf{Y}}}^{-\frac{1}{2}} \tilde{\mathbf{Y}}}_{\triangleq \tilde{\mathbf{Y}}_t} = \|\mathbf{Y}_t\|_{\text{Fro}}^2 \\ &\geq \|\mathbf{Y}_t\|_2^2 = \left\| \sqrt{\chi^2(\alpha, nm)} \mathbf{W}_1^{-1} \mathbf{Y} \right\|_2^2 = \chi^2(\alpha, nm) \gamma_f^2(\Phi). \end{aligned} \quad (\text{A.8.23})$$

In general the Frobenius norm of a matrix is greater than or equal to the induced 2-norm (Horn and Johnson, 1990, p. 314), but for a rank one matrix such as  $\tilde{\mathbf{Y}}_s$  it is always equal, since  $\|\mathbf{Y}\|_{\text{Fro}} = \sqrt{\sum \sigma_j^2(\mathbf{Y})}$  (Skogestad and Postlethwaite, 1996, p. 556), implying that all other rank deficient matrices have a larger Frobenius norm and consequently are rejected based on the  $\chi^2$  hypothesis test. The confidence score  $\gamma_f(\Phi)$  is hence tight when the errors are assumed to be jointly normally distributed and the variance in each variable is the same for independent experiments. Note that to obtain uncertainty ellipses with this rank deficient realization  $\tilde{\mathbf{Y}}_s$  on the boundary, we have to scale the nominal ones in Figure A.11 by different factors, more precisely  $\mathcal{U}_{\phi_1}^{0.05}$  by 0.5420 and  $\mathcal{U}_{\phi_2}^{0.05}$  by 0.8435.

We conclude by noting that this example has shown that the condition for testing practical independence in Theorem 5.10.1, in general, is conservative and that tight conditions exist for certain cases. In general, the distance from the observed regressor matrix to the closest rank deficient realization is for column uncertainty the same for each regressor and for element-wise uncertainty the same for each element, while it for full matrix uncertainty is different for each regressor and element.



---

# Inference of subnetworks in depth

---

Our extended IRMA example, which we use in Section 6, is an extension of the IRMA network engineered in *S. cerevisiae* by Cantone *et al.* (2009). We here explain why and how we constructed it, as well as give a short introduction to transfer functions for readers unfamiliar with them. The construction of the extended IRMA example is presented in Section B.1, which we have divided into subsections in the following way. We first give a state-space model of the IRMA network that we then extend by introducing 7 additional states. In Section B.1.1, we explain what the additional state variables represent and motivate why we added them as well as give the resulting interaction matrix. In Section B.1.2, we discuss published evidence for the interactions that we introduce together with the additional states. In Section B.1.3, we derive the transfer functions between the five original states that the additional latent states give rise to. In Section B.1.4, we identify the coefficients of the full interaction matrix of the extended IRMA model based on the inferred Glucose and Galactose model in Section 4.6. We finally give a short introduction to transfer functions in Section B.2.

## B.1 Construction of the extended IRMA example

To illustrate the effect of latent states, we extended the IRMA network model, described in Section 2.1, to also contain state variables representing proteins and bound mRNA. Typically models of gene regulatory networks only contain state variables representing the abundance of mRNA of the considered genes, but this is even based on the central dogma of molecular biology clearly a simplification (Brazhnik *et al.*, 2002; Crick, 1970). At least some of the regulatory interactions are mediated by proteins, which can form complexes, be phosphorylated, transported *etc.*, which motivates our introduction of intermediary states. Completion of the human genome project has revealed that many of the transcribed non-coding sequences also have a regulatory function at the translational level (Lander, 2011; Venter *et al.*, 2001; IHGSC, 2001, 2004). The translation of some genes may therefore be regulated by molecules binding to the mRNA and blocking translation. We next explain how the example was constructed and motivate those parts that can be motivated based on existing biological knowledge. Note however that our emphasis when selecting the latent states was on construction of a simple and pedagogical example which can explain the steady-state data recorded by Cantone *et al.* (2009) and not on representation of all current biological knowledge.

Before we extend the IRMA network, consider the following system of linear ordinary differential equations, *i.e.* state-space model, which describes the IRMA network as engineered by Cantone *et al.* (2009). Let us introduce one state variable representing the log-fold change of each gene ( $x_1=CBF1$ ,  $x_2=GAL4$ ,  $x_3=SWI5$ ,  $x_4=ASH1$ ,  $x_5=GAL80$ ) and the steady-state perturbations used by Cantone *et al.*, which directly over-express each gene ( $p_1=CBF1$ ,  $p_2=GAL4$ ,  $p_3=SWI5$ ,  $p_4=ASH1$ ,  $p_5=GAL80$ ). In state-space, the IRMA system in Figures 2.1 and 4.1 can now be

described as

$$\begin{aligned} \begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \\ \dot{x}_3(t) \\ \dot{x}_4(t) \\ \dot{x}_5(t) \end{bmatrix} &= \begin{bmatrix} a_{11} & 0 & a_{13} & a_{14} & 0 \\ a_{21} & a_{22} & 0 & 0 & a_{25} \\ 0 & a_{32} & a_{33} & 0 & 0 \\ 0 & 0 & a_{43} & a_{44} & 0 \\ 0 & a_{52} & a_{53} & 0 & a_{55} \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \\ x_4(t) \\ x_5(t) \end{bmatrix} \\ &+ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} p_1(t) \\ p_2(t) \\ p_3(t) \\ p_4(t) \\ p_5(t) \end{bmatrix}. \end{aligned} \quad (\text{B.1.1})$$

The type of the interaction is determined by the signs of the coefficients  $a_{ij}$  of the interaction matrix  $\mathbf{A}$  and the strength by the absolute value. All state variables and perturbations represent changes ( $x_i(t) = \log_2 \underline{x}_{it}(t) - \log_2 \underline{x}_{iC}(t)$ ) from the steady-state values of the *S. cerevisiae* grown in either Glucose or Galactose, since the system is assumed to be nonlinear and the description above only is valid locally near the steady-state of the dynamical mode ( $\underline{x}_{iC}$ ); see *e.g.* Nordling *et al.* (2007b) for a discussion on dynamical modes. This model formalism is known as S-systems (Savageau, 1969, 1976, 1987; Crampin *et al.*, 2004) and was recently used in (Jörnsten *et al.*, 2011). Here  $t$  in parenthesis marks that the system is assumed to be dynamical, *i.e.* the state depends on time, while the lower index  $t \in \{1, 2, \dots, 5\}$  marks the gene that is perturbed in the experiment. In this model, the self-loops  $a_{ii}$  represent the degradation of the mRNA of gene  $i$ . All other interactions represent the influence from gene  $j$  on gene  $i$  mediated by translation of mRNA into protein that acts as a transcription factor activating or repressing transcription of gene  $i$ , except  $a_{25}$  and  $a_{52}$  that represent the formation of a protein complex between the Gal4 protein and Gal80 protein, as described in Cantone *et al.* (2009). For further detail see Section 2.1 and the original article by Cantone *et al.*

We next extend this model by adding 7 state variables that represent proteins, protein complexes, or free and bound mRNA. We first present the resulting interaction matrix. Then we explain what the additional state variables represent and motivate why we added them; all in Section B.1.1. We discuss published evidence for the interactions that we introduce in Section B.1.2. While we in Section B.1.3 derive the transfer functions between the five original states that the additional latent states give rise to. We, finally in Section B.1.4, identify the coefficients of the full interaction matrix of the extended IRMA model based on the inferred Glucose and Galactose model in Section 4.6.

### B.1.1 Extension of the IRMA example

We here first present the interaction matrix that results from our extension of the IRMA network by 7 additional state variables. Then we motivate the additional state variables and why we added them. The interaction matrix of the extended IRMA model is shown in Table B.1. Note that we have removed the interactions:  $a_{14}$ ,  $a_{32}$ ,  $a_{25}$ , and  $a_{52}$ , since we add states representing the free and bound *ASH1* mRNA, a state representing the *Gal4* protein mediating the second interaction, and the *Gal4-Gal80* protein complex that is represented by the last two interactions. We also change the sign of  $a_{21}$  from positive to negative, because we add the *Cbf1* protein mediating the activation that this interaction represents in (B.1.1).

The first three of our additional states:  $x_6$ =Gal4p,  $x_7$ =Gal80p, and  $x_8$ = Gal4p-Gal80p represent the *Gal4* protein, *Gal80* protein, and *Gal4-Gal80* protein complex. Addition of these allow us to describe the mechanism behind the protein complex more accurately and gives us four strictly proper transfer functions of order 3 with relative order 1 or 3. For a brief introduction to transfer functions and their order see Section B.2. Two of the four interactions introduced by these transfer functions are required to explain the Glucose and Galactose steady-state data recorded by Cantone *et al.*

**Table B.1: The full interaction matrix ( $A$ ) of the extended IRMA network.**

State \ State	<i>CBF1</i>	<i>GAL4</i>	<i>SWI5</i>	<i>ASH1</i>	<i>GAL80</i>	<i>Gal4p</i>	<i>Gal80p</i>	<i>Gal4p-Gal80p</i>	<i>Cbf1p</i>	<i>Swi5p-Unk</i>	<i>ASH1f</i>	<i>ASH1b</i>
<i>CBF1</i>	$a_{11}$	0	$a_{13}$	0	0	$a_{16}$	0	0	0	0	$a_{1,11}$	0
<i>GAL4</i>	$a_{21}$	$a_{22}$	0	0	0	0	0	0	$a_{29}$	0	0	0
<i>SWI5</i>	0	0	$a_{33}$	0	0	$a_{36}$	0	0	$a_{39}$	0	0	0
<i>ASH1</i>	0	0	$a_{43}$	$a_{44}$	0	0	0	0	$a_{49}$	$a_{4,10}$	0	$a_{4,12}$
<i>GAL80</i>	0	0	$a_{53}$	0	$a_{55}$	0	0	0	0	$a_{5,10}$	0	0
<i>Gal4p</i>	0	$a_{62}$	0	0	0	$a_{66}$	0	$a_{68}$	0	0	0	0
<i>Gal80p</i>	0	0	0	0	$a_{75}$	0	$a_{77}$	$a_{78}$	0	0	0	0
<i>Gal4p-Gal80p</i>	0	0	0	0	0	$a_{86}$	$a_{87}$	$a_{88}$	0	0	0	0
<i>Cbf1p</i>	$a_{91}$	0	0	0	0	0	0	0	$a_{99}$	0	0	0
<i>Swi5p-Unk</i>	0	0	$a_{10,3}$	0	0	0	0	0	0	$a_{10,10}$	0	0
<i>ASH1f</i>	0	0	$a_{11,3}$	0	0	0	0	0	$a_{11,9}$	$a_{11,10}$	$a_{11,11}$	$a_{11,12}$
<i>ASH1b</i>	0	0	0	0	0	0	0	0	0	0	$a_{12,11}$	$a_{12,12}$

The fourth additional state  $x_9=\text{Cbf1p}$  represents the Cbf1 protein and it gives us two strictly proper transfer functions of order 1 with relative order 1, that we use to explain the Glucose and Galactose steady-state data recorded by Cantone *et al.* We also use it to create a feedforward loop that counteracts the engineered activation of *GAL4*, since it provides us an opportunity to illustrate how mRNA and protein interactions can cancel each other and this interaction is not needed to explain the steady-state data. The resulting transfer function is semi-proper of order 1 with relative order 0. Note that the relative order of a semi-proper transfer function always is zero.

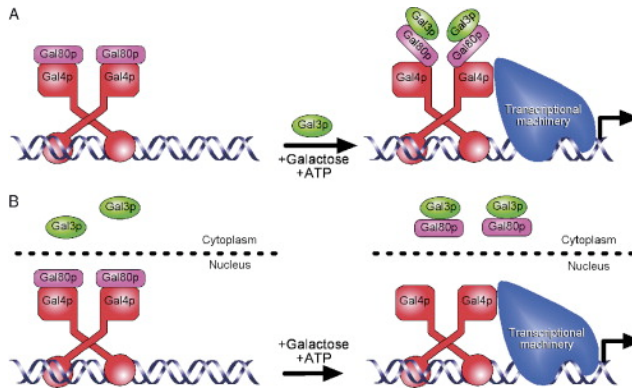
The fifth additional state  $x_{10}=\text{Swi5p-Unk}$  represents an unknown state variable that depends on the Swi5 protein and we use it to illustrate a feedforward loop that cancels the  $a_{43}$  interaction at steady-state in Glucose, while it instead cancels the  $a_{53}$  interaction in Galactose. No influence from *SWI5* to *ASH1* is needed to explain the Glucose steady-state data, while this interaction is present in the model chosen to explain the Galactose steady-state data. No influence from *SWI5* to *GAL80* is, on the other hand, needed to explain the Galactose steady-state data, while this interaction is present in the model chosen to explain the Glucose steady-state data. The resulting transfer function is in both cases semi-proper of order 1.

The sixth and seventh additional states  $x_{11}=\text{ASH1f}$ , and  $x_{12}=\text{ASH1b}$  represent free *ASH1* mRNA that can be translated and *ASH1* mRNA that is bound to some protein and therefore cannot be translated. We use them to illustrate a feedback loop with an integrator that maintains the state of  $x_{11}$  unchanged such that the repression of *CBF1* by *ASH1* is cancelled. No influence from *ASH1* to *CBF1* is needed to explain the Glucose nor the Galactose steady-state data. This gives rise to two semi-proper transfer functions of order 3: one from *SWI5* to *CBF1* and a self-loop, both mediated by *ASH1f*. Assuming that Cantone *et al.* have measured the total abundance of *ASH1* mRNA,  $x_4 = x_{11} + x_{12}$  and the full interaction matrix in Table B.1 is rank deficient.

### B.1.2 Published evidence of the introduced interactions

We considered alternative mechanisms for the desired illustration that could explain the Glucose and Galactose steady-state data recorded by Cantone *et al.* before adding any state variable or interaction and tried to add known mechanisms as far as possible. We here discuss published evidence of the added states and interactions, that we found in either the Saccharomyces Genome Database (SGD) (Weng *et al.*, 2003; Nash *et al.*, 2007), Biological General Repository for Interaction Datasets (BioGRID) (Stark *et al.*, 2011), Information Hyperlinked over Proteins (iHOP) (Hoffmann and Valencia, 2004), or through PubMed ([www.pubmed.org](http://www.pubmed.org)).

The Gal4 protein is a DNA binding transcriptional activator that binds to upstream activating sequences in the GAL promoter region (Bhat and Murthy, 2001; Sellick *et al.*, 2008). In the absence of Galactose, however, the Gal80 protein forms a complex with the Gal4 protein, which prevents transcriptional activation. The activation domain of Gal4p is masked by the binding of Gal80p. In the presence of



**Figure B.1: Alternative models for *GAL* gene activation.** (A) Activation occurs by formation of a Gal3p-Gal4p-Gal80p complex that leads to a conformational change in Gal80p. (B) Binding of Gal3p to Gal80p decreases the amount of Gal80p available in the nucleus and formation of the Gal4p-Gal80p complex, which leads to transcriptional activation. Reprinted from Sellick et al. (2008) with permission from Elsevier.

Galactose the Gal3 protein binds to Gal80p and Gal4p can activate transcription. Note that *GAL3* is transcribed and Gal3p is present also in the absence of Galactose, but the binding is only stable in the presence of Galactose (Suzuki-Fujimoto et al., 1996b; Sellick et al., 2008). Gal3p either forms a Gal3p-Gal4p-Gal80p complex that leads to a conformational change in Gal80p (Figure B.1A) or decreases the amount of Gal80p available in the nucleus and thereby the formation of the Gal4p-Gal80p complex (Figure B.1B) (Sellick et al., 2008). Despite the mechanism being debated, it is well documented that it leads to activation of transcription of the *GAL* genes. Introduction of  $x_6 = \text{Gal4p}$ ,  $x_7 = \text{Gal80p}$ , and  $x_8 = \text{Gal4p-Gal80p}$ , with  $a_{62}$  and  $a_{75}$  positive to represent translation of the *GAL4* and *GAL80* mRNA to protein, enables us to represent the effect of the mentioned mechanisms on the observed genes. Complex formation decreases the amount of free Gal4 and Gal80 protein so we assume that  $a_{68}$  and  $a_{78}$  are negative to implement the prevention by Gal80 of transcriptional activation by the Gal4 protein. An increase in the abundance of Gal4 and Gal80 protein should increase formation of the Gal4p-Gal80p complex, so  $a_{86}$  and  $a_{87}$  should be positive. The expression of *GAL4* is however reduced 4-5 fold in Glucose (Johnston et al., 1994). We therefore expect that Gal4p is rare compared to Gal80p and an increase of Gal80p has no or little effect on the formation of the Gal4p-Gal80p complex, while an increase of Gal4p has a significant effect. To capture this we assume that  $a_{87} = 0$  in the absence of Galactose. Note that the same effect would result from an increase in expression of *GAL80* in Glucose or if less Gal80p was bound to Gal3 in the absence of Galactose, but no increase in Gal80p nor Gal3p-Gal80p immunoprecipitation in the presence of Galactose has been detected (Suzuki-Fujimoto et al., 1996b). Insensitivity of formation of the

Gal4p-Gal80p complex to changes in Gal80p abundance and sensitivity to Gal4p is crucial to ensure that all Gal4p is bound and obtain a robust switch that prevents transcriptional activation. In the absence of Galactose, Gal4p occupancy of the *GAL1/10* and *GAL3* promoters has been reported to have a half-life of 5 min or less (Nalley et al., 2006), which explains how nuclear Gal80p can bind essentially all Gal4p. In IRMA the *GAL10* promoter is fused to *SWI5*, so we assume that  $a_{36}$  is positive to represent transcriptional activation by Gal4p, implying that over-expression of *GAL4* should lead to increased transcription of *SWI5*. This is observed in both the Glucose and Galactose steady-state data sets recorded by Cantone et al. Both of them also contains fairly strong support for an activation of *SWI5* by *GAL4*, which is consistent with the previously reported mechanisms that we have described above. If we did not assume that  $a_{87} = 0$ , then over-expression of *GAL80* would lead to increased formation of the Gal4p-Gal80p complex, preventing transcriptional activation by Gal4p and repression of *SWI5*. While this is not seen in the Glucose steady-state data by Cantone et al., it is evident in the Galactose data and we therefore assume  $a_{87}$  to be positive in Galactose. Moreover, the Galactose data cannot be explained without a repression from *GAL80* to *SWI5*. In order for this to be consistent with formation of a Gal3p-Gal4p-Gal80p complex in Galactose (Figure B.1A) the amount of Gal3p would have to be so low that the additional Gal80p binds to Gal4p alone and decreases the rate of transcription of *SWI5*. In order for over-expression of *GAL4* to increase transcription of *SWI5* essentially all Gal80p need to be bound to Gal3p, so a delicate balance between the amount of Gal80p and Gal3p is required, which would make the switch sensitive. The second mechanism (Figure B.1B), on the other hand, is consistent as long as binding of Gal80p to Gal3p does not completely prevent binding of Gal80p to Gal4p, which makes it more robust. No definitive choice between the two alternative mechanisms can be made, since we e.g. lack data on how changes in expression of *GAL3* affects the system, but the data by Cantone et al. favours the second mechanism based on our robustness argument. The inability to explain the Galactose data without a repression from *GAL80* to *SWI5* hence adds yet an argument to this ongoing scientific debate. The Glucose steady-state data by Cantone et al. contains fairly strong support for an activation of *CBF1* by *GAL4*, while the Galactose data does not, so we assume  $a_{16}$  positive in Glucose and zero in Galactose. If we do not include this interaction in the Glucose model, then we need to include some other interactions to compensate for it in order to obtain a model that can explain the Glucose data, while the Galactose data can be explained without it. We have not found published evidence of Gal4p directly binding to the *HO* promoter, which Cantone et al. fused to *CBF1*. But transcription of *HO* requires chromatin remodelling by the Swi-Snf complex and Gal4p is an activator of the chromatin remodelling complex (Kishore and Kundu, 2003; Cosma et al., 1999). More precisely, Swi5p binds to the *HO* promoter and recruits the Swi-Snf complex, which in turn recruits the SAGA complex and SBF to start transcription (Peterson and Workman, 2000; Cosma et al., 1999). Both Gal4p and Swi5p also interacts directly with Gal11p in the Mediator complex that recruits RNA Polymerase II to the promoter, which catalyses the synthesis of mRNA (Bhoite

et al., 2001; Kishore and Kundu, 2003). We therefore suggest that Gal4p in the absence of Galactose helps to recruit the Swi-Snf and/or Mediator complexes to *HO*, while it in Galactose is stably bound to the *GAL1/10* promoter. The Gal4p promoter complex has been observed to have a half-life of approximately 1h when *S. cerevisiae* is grown in Galactose (Nalley et al., 2006).

The *CBF1* gene encodes a protein belonging to the bHLH family called Cbf1p (Robinson and Lopes, 2000), which we have introduced as our fourth additional state,  $x_9$ . Translation of *CBF1* is represented by  $a_{91}$ , which we therefore assume to be positive. Cbf1p is involved in chromosome segregation and transcriptional control of the *MET* genes (Robinson and Lopes, 2000). It binds to a region called CDE1 in the centromeres and promoters of e.g. *MET16*, *MET25*, *TRP1*, and *GAL2*. Cantone et al. state that they selected the *MET16* promoter for *GAL4* so that it would be controlled by *CBF1*, since Cbf1p is essential for transcription of *MET16*, whereas it is not for the other *MET* genes (Ferreiro et al., 2004). We therefore assume that  $a_{29}$  is positive. Nonetheless, no influence from *CBF1* to *GAL4* is needed to explain the steady-state data recorded by Cantone et al.. In particular, the Galactose steady-state data can be explained without this interaction even if the uncertainty was one hundred of the estimated uncertainty. Activation of *MET16* by Cbf1p has been shown to depend on the location of the CDE1 site relative to the Gcn4p binding site and on recruitment of other transcriptional activators, such as Met4p and Met28p (Robinson and Lopes, 2000). The lack of an influence from *CBF1* to *GAL4* could therefore be explained by changes caused by the fusion of the *MET16* promoter to *GAL4*. *MET16* transcription is, however, regulated by different pathways: one dependent on binding of a complex containing Cbf1p, Met4p, and Met28p to the CDE1 site, and another through the general control of amino acids, which depends on Cbf1p and binding of Gcn4p to the AP-1 site (Ferreiro et al., 2004). These two pathways form a feedforward loop, which may cancel out the effect of Cbf1p. We have here chosen to illustrate this type of cancellation in the simplest possible manner by assuming that both the mRNA of *CBF1* and Cbf1p influences the transcription of *GAL4*, which requires  $a_{21}$  to be negative. The Glucose steady-state data of Cantone et al. cannot be explained without an influence from *CBF1* to *SWI5* and the Galactose data also contains strong support for this interaction, so we assume that Cbf1p activates transcription of *SWI5*, i.e.  $a_{39} > 0$ , which has the *GAL10* promoter. The promoter of the *GAL2* gene contains the CDE1 site (Mellor et al., 1990), but we have not found evidence that *GAL10* does. At least it does not contain the TCACGTG sequence over which the Cbf1p-Met4p-Met28p complex is formed (Kuras et al., 1997), based on a BLAST similarity search of NCBI reference sequence: NM\_001178367.1. If it did, then it could explain this interaction. We have not found any publication nor data base containing this interaction, so it appears to be a novel interaction. The Glucose data also contains fairly strong evidence for an influence from *CBF1* to *ASH1*, so we assume that Cbf1p also activates transcription of *ASH1*. More precisely, transcription of *ASH1f*, since we distinguish between free and bound *ASH1* mRNA, and we therefore assume  $a_{11,9} > 0$ . An interaction between *CBF1* and *ASH1* has previously been reported in a study of cooperative binding among transcription

factors deduced from ChIP-chip data (Datta and Zhao, 2008). This interaction also has some support in the Galactose data and we therefore also included it in our Galactose model. Note that Cantone *et al.* fused the *ASH1* promoter to both *ASH1* and *GAL80*, but the evidence for an influence from *CBF1* to *GAL80* is considerably weaker and we have therefore not included any. The steady-state data is however not informative enough to say that one is necessary while the other is not needed.

*SWI5* encodes a zinc finger protein (Bobola *et al.*, 1996). In wild type *S. cerevisiae* Swi5p accumulates in the cytoplasm during G2 and M phase due to phosphorylation of the Serine 522, 646, and 664 sites by Cdc28p and only enters the nucleus after dephosphorylation by Cdc14p during anaphase (Visintin *et al.*, 1998). Cantone *et al.* substituted the Serine in these three phosphorylation sites by Alanin to avoid cell cycle control of Swi5p and to obtain constant entry of Swi5p to the nucleus (Cantone *et al.*, 2009). Constant entry is required because Swi5p is highly unstable once it enters the nucleus (Visintin *et al.*, 1998). Swi5p is a transcriptional activator that binds to both the *ASH1* and *HO* promoters (Cosma, 2004; Bobola *et al.*, 1996). Transcription of *HO* is however delayed until the late G1 phase of the cell cycle due to sequential recruitment of the chromatin remodelling complex, while transcription of *ASH1* starts directly (Cosma *et al.*, 1999; Cantone *et al.*, 2009). We should therefore assume that a time-delay exists between over-expression of *SWI5* and the activation of transcription of *HO*, but we here only deal with steady-state data, which is unaffected by any time-delay. Cantone *et al.* designed the IRMA network so that transcription of both *ASH1* and *GAL80* is controlled by the *ASH1* promoter, while *CBF1* is controlled by the *HO* promoter. Over-expression of *SWI5* should therefore activate all three, but both the Glucose and Galactose steady-state data can, strictly speaking, be explained without these interactions. The support for an activation of *CBF1* and *GAL80* is however fairly strong based on the Glucose data, while the Galactose data supports activation of *CBF1* and *ASH1*, but not *GAL80*. This suggest that activation of these genes by Swi5p depends on the growth medium. We have searched for publications addressing differences in the activation of *HO* and *ASH1* by Swi5p in different growth media but not found any. We therefore introduce our fifth additional state  $x_{10}$ =Swi5p-Unk as an unknown state variable depending on Swi5p, which is affected by Galactose, in order to explain Galactose induced differences. We assume that Swi5p-Unk forms a feedforward loop to both *ASH1* and *GAL80*, with  $a_{10,3}$  positive and  $a_{5,10}$  and  $a_{11,10}$  negative. Swi5p is known to have many interactions, forms, and different functions, of which we next mention a few, so many potential candidates for this state variable exist. More than 200 unique interactors of Swi5p is listed in the BioGRID database. Swi5p is also a transcription factor of at least *SIC1* and *PCL9* (Visintin *et al.*, 1998). Swi5p is phosphorylated by Pho85p at a region distinct from the three Cdc28p phosphorylation sites (Measday *et al.*, 2000), where Cantone *et al.* substituted Serine by Alanin to avoid cell cycle control of Swi5p. Swi5p interacts directly with Gal11p in the Mediator complex that recruits RNA Polymerase II to the promoter, which catalyzes the synthesis of mRNA (Bhoite *et al.*, 2001; Kishore and Kundu, 2003). Swi5p is also involved in DNA repair (Haruta *et al.*, 2008). Note that the function of Swi5p in the strain engineered by Cantone *et al.*

differs from wild type due to the modifications mentioned earlier and due to deletion of the *ACE2* gene, which cooperates with *Swi5p* in regulating transcription of the *ASH1* promoter. Expression of *ASH1* is reduced when *ACE2* is deleted (Measday et al., 2000).

*Ash1p* is a repressor that inhibits *HO* transcription after accumulation in daughter cells in late anaphase, which enables mating type switching in *S. cerevisiae* (Cosma, 2004; Bobola et al., 1996). *Swi5p* binds to both the *ASH1* and *HO* promoters and *ASH1* is immediately transcribed while the transcription of *HO* is delayed until the late G1 phase of the cell cycle. Cantone et al. fused the *HO* promoter to *CBF1*, so over-expression of *ASH1* should down-regulate *CBF1*, but their steady-state data can be explained without this interaction. We next discuss the mechanism of *ASH1* translation, which can explain this. The *ASH1* mRNA contains four minimal *cis*-acting localization sequences that are recognized by *She2p* already in the nucleus (Cosma, 2004; Shen et al., 2009). Disruption of the four localization elements results in complete de-localization of *ASH1* mRNA and symmetrical distribution of *Ash1p* between mother and daughter (Chartrand et al., 2002). Several other proteins, such as *Npl3p*, *Rrp5p*, *Khd1p*, *Scp160p*, *Ydl124wp*, *Gcy1p*, *Pcs60p*, *Mdh3p*, *Sec1p*, *Sec16p*, etc., has also been shown to bind to the mRNA of *ASH1* (Tsvetanova et al., 2010). But *She2p* is crucial, since the *Myo4p-She3p* complex, which facilitates the transport to the distal tip along actin filaments, does not bind to the mRNA independent of *She2p* (Gonsalvez et al., 2003; Aronov et al., 2007; Chung and Takizawa, 2010). Exclusion of *She2p* from the nucleus also disrupts the binding of *Loc1p* and *Puf6p*, which both are involved in translational repression, to the *ASH1* mRNA (Shen et al., 2009). Once the *ASH1* mRNA is transported to the distal tip of the bud it is anchored by factors such as *Khd1p*, *Bud6p-Aip3p*, and/or *She5* (Cosma, 2004). Phosphorylation of *Puf6p* then releases the translational repression and *Ash1p* enters the nucleus of the daughter cell and represses *HO* transcription (Shen et al., 2009). Deletion of *FUN12* reduces the level of *Ash1p* by more than 80%, while over-expression increases the level of *Ash1p* and *Puf6p* suppresses *ASH1* mRNA translation via *Fun12p* (Deng et al., 2008). Cantone et al. deleted the *SHE2* gene to obtain a homogeneous population in which transcription of the *HO* promoter is not developmentally controlled (Gonsalvez et al., 2003; Cantone et al., 2009). Based on the description above, it should be clear that translational control is at least as important for production of *Ash1p* and repression of the *HO* promoter as transcriptional control. We therefore decided to introduce  $x_{11} = \text{ASH1f}$  and  $x_{12} = \text{ASH1b}$  and use them to illustrate how a feedback loop at the translational level can cancel the repression of *CBF1* mediated by *Ash1p*. Here *ASH1f* represents free *ASH1* mRNA that can be translated, while *ASH1b* represents *ASH1* mRNA that is bound to some protein and therefore cannot be translated. We assume that they form a negative feedback loop in which *ASH1b* acts as an integrator, i.e.  $a_{12,12} = 0$ , such that the amount of *ASH1f* is kept constant at steady-state. A mechanism like this is plausible, since deletion of *SHE2* both affects the localization and translational control, but we have not found any direct evidence for it in the literature. We assume that the total amount of *ASH1* mRNA is measured by Cantone et al., i.e.  $x_4 = x_{11} + x_{12}$ ,

since they reported that the quality of the purified RNA was good and purified RNA of good quality should be free of protein (Fleige and Pfaffl, 2006). This implies that *ASH1f* takes the role that Cantone *et al.* assumed for *ASH1* and we therefore assume  $a_{11,3}$  and  $a_{12,11}$  positive and  $a_{1,11}$  and  $a_{11,12}$  negative, while the coefficients of *ASH1* is given by the linear dependence above.

### B.1.3 Transfer functions in the extended IRMA example

We would like to know the influence of the additional states and their interactions on the interactions between the original observed states and therefore here derive transfer functions representing these. For a brief introduction to transfer functions, see Section B.2. The transfer functions are obtained by taking the Laplace transform of the extended IRMA system. First, the state variables are partitioned into a set containing the original observed states,  $x_1$  to  $x_5$ , and one containing the latent states,  $x_6$  to  $x_{12}$ . Then the full interaction matrix in Table B.1 is partitioned in the same way

$$\mathbf{A} = \left[ \begin{array}{c|c} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \hline \mathbf{A}_{21} & \mathbf{A}_{22} \end{array} \right], \quad (\text{B.1.2})$$

so that the  $\mathbf{A}_{11}$  block contains the direct links among the observed states, the  $\mathbf{A}_{12}$  block the links from the latent states to the observed ones, *etc.*. The desired interaction matrix of the observed states alone is now given by

$$\mathbf{A}(s) = \mathbf{A}_{11} + \mathbf{A}_{12} (s\mathbf{I} - \mathbf{A}_{22})^{-1} \mathbf{A}_{21}, \quad (\text{B.1.3})$$

which is a matrix of transfer functions as marked by the  $s$  in parenthesis. We have thus replaced the additional latent protein states by transfer functions from gene  $j$  to gene  $i$ , which does not depend on any state representing a gene, *i.e.* any of the original observed states. We next give the resulting transfer functions, followed by the resulting interaction matrix. The Gal4-Gal80 protein complex mechanism involves the first three additional states:  $x_6$ =Gal4p,  $x_7$ =Gal80p, and  $x_8$ =Gal4p-Gal80p. We assume that Gal4p activates *SWI5* and *CBF1*. The first part of the assumption gives us a transfer function from *GAL4* to *SWI5*

$$a_{32}(s) = \frac{a_{36}a_{62}((s - a_{77})(s - a_{88}) - a_{78}a_{87})}{\star}, \quad (\text{B.1.4})$$

$$\star \triangleq (s - a_{66})(s - a_{77})(s - a_{88}) - a_{78}a_{87}(s - a_{66}) - a_{68}a_{86}(s - a_{77}),$$

replacing the interaction  $a_{32}$  in the original IRMA model, and one from *GAL80* to *SWI5*

$$a_{35}(s) = \frac{a_{36}a_{68}a_{87}a_{75}}{\star}. \quad (\text{B.1.5})$$

The second part gives us a transfer function from *GAL4* to *CBF1*

$$a_{12}(s) = \frac{a_{16}a_{62}((s - a_{77})(s - a_{88}) - a_{78}a_{87})}{\star}, \quad (\text{B.1.6})$$

and from *GAL80* to *CBF1*

$$a_{15}(s) = \frac{a_{16}a_{68}a_{87}a_{75}}{\star}. \quad (\text{B.1.7})$$

In these transfer functions we assume that all non-diagonal coefficients  $a_{ij}$  are positive, except  $a_{68}$  and  $a_{78}$  that we assume to be negative, since an increase in Gal4p-Gal80p requires both Gal4p and Gal80p. In Glucose  $a_{87}$  is assumed to be zero and the amount of Gal4p-Gal80p is, in practice, determined by the abundance of Gal4p alone. This makes both  $a_{35}(s)$  and  $a_{15}(s)$  zero, which is consistent with the fact that the Glucose steady-state data can be explained without these interactions. In Galactose, on the other hand,  $a_{87}$  is assumed positive to make  $a_{35}(s)$  negative, which is consistent with the Galactose steady-state data. Actually the Galactose data cannot be explained without a nonzero interaction  $a_{35}(s)$ . In Galactose  $a_{16}$  is for simplicity assumed to be zero, since neither  $a_{12}(s)$  nor  $a_{15}(s)$  is necessary to explain the steady-state data. Note that all diagonal coefficients, corresponding to a self-loop  $a_{ii}$ , are assumed to be negative, in order to represent degradation of mRNA or proteins.

We assume that the Cbf1 protein  $x_9$ =Cbf1p activates *SWI5* and *ASH1f*, which gives us the following two transfer functions:

$$a_{31}(s) = \frac{a_{39}a_{91}}{s - a_{99}}, \quad (\text{B.1.8})$$

$$a_{41}(s) = \frac{a_{11,9}a_{91}}{s - a_{99}}. \quad (\text{B.1.9})$$

The later transfer function stems from our assumption that Cantone *et al.* have measured the total amount of *ASH1* mRNA, *i.e.*  $x_4 = x_{11} + x_{12}$ , since the time derivative of *ASH1* then is

$$\begin{aligned} \dot{x}_4 = \dot{x}_{11} + \dot{x}_{12} = & a_{11,3}x_3 + a_{11,9}x_9 + a_{11,10}x_{10} \\ & + (a_{11,11} + a_{12,11})x_{11} + (a_{11,12} + a_{12,12})x_{12}, \end{aligned} \quad (\text{B.1.10})$$

and the Laplace transform is

$$\begin{aligned} sx_4(s) = & \underbrace{a_{11,9} \frac{a_{91}}{s - a_{99}}}_{=a_{41}(s)} x_1(s) + \left( \underbrace{a_{11,3} + a_{11,10} \frac{a_{10,3}}{s - a_{10,10}}}_{=a_{43}(s)} \right) x_3(s) \\ & + (a_{11,11} + a_{12,11})x_{11}(s) + (a_{11,12} + a_{12,12})x_{12}(s). \end{aligned} \quad (\text{B.1.11})$$

Here all non-diagonal coefficients  $a_{ij}$  are assumed positive, except  $a_{11,10}$  and  $a_{11,12}$ , which enables us to explain the Glucose and Galactose steady-state data. The

Glucose data cannot be explained without  $a_{31}(s)$  being positive. The data can be explained without  $a_{41}(s)$ , but the support for this interaction is quite strong so we would have to replace it with some other interaction. Both interactions also have fairly strong support in the Galactose data. On the other hand, no influence from *CBF1* to *GAL4* is needed to explain neither of the steady-state data sets and we cancel this interaction by a feedforward loop. To implement a feedforward loop that cancels the activation of *GAL4* by *CBF1* in the simplest possible way, we assume that both the mRNA of *CBF1* and the Cbf1 protein  $x_9$  influence transcription of *GAL4*. This gives us the transfer function

$$a_{21}(s) = \frac{a_{21}(s - a_{99}) + a_{29}a_{91}}{s - a_{99}}, \quad (\text{B.1.12})$$

which is zero at steady-state if we assume  $a_{29}a_{91} = a_{21}a_{99}$ , with  $a_{29}$  and  $a_{91}$  positive and  $a_{21}$  and  $a_{99}$  negative. This transfer function is not strictly proper, which at first sight may seem physically questionable, since the gain should approach zero at high frequencies, but it is a mere artefact of how the system is represented. To obtain the effect of  $x_1$  on the state  $x_2$  this transfer function needs to be multiplied by  $1/(s - a_{22})$ , which makes it strictly proper. Note that all the coefficients  $a_{ij}$  that do not depend on any latent state also are semi-proper transfer functions.

We assume that  $x_{10}$ —the unknown state variable that depends on Swi5p—represses *ASH1f* and *GAL80*. This gives us two feedforward loops:

$$a_{43}(s) = \frac{a_{11,3}(s - a_{10,10}) + a_{11,10}a_{10,3}}{s - a_{10,10}}, \quad (\text{B.1.13})$$

$$a_{53}(s) = \frac{a_{53}(s - a_{10,10}) + a_{5,10}a_{10,3}}{s - a_{10,10}}. \quad (\text{B.1.14})$$

The former transfer function involves the free and bounded *ASH1* and is derived in (B.1.11). Here all non-diagonal coefficients  $a_{ij}$  are assumed positive, except  $a_{11,10}$  and  $a_{5,10}$ . We assume that  $a_{11,10}a_{10,3} = a_{11,3}a_{10,10}$  and  $a_{5,10}a_{10,3} > a_{53}a_{10,10}$  in Glucose, because this makes the first transfer function zero and the second positive at steady-state. This assumption is consistent with the fact that the first interaction is not needed to explain the Glucose steady-state data. In Galactose, we instead assume  $a_{11,10}a_{10,3} > a_{11,3}a_{10,10}$  and  $a_{5,10}a_{10,3} = a_{53}a_{10,10}$ , so that the first transfer function is positive and the second is zero at steady-state, since the second interaction is not needed to explain the Galactose steady-state data. Actually neither is needed to explain the data, but each data set contains fairly strong support for one of the interactions, meaning that if we did not include it then we would have to include some alternative interactions.

The last two states,  $x_{11} = \text{ASH1f}$ , and  $x_{12} = \text{ASH1b}$ , represent free and bound mRNA. We assume that the change in *ASH1* mRNA abundance measured by Cantone *et al.* equals the total change of both free and bound mRNA. This makes  $x_4$  a linear combination of  $x_{11}$  and  $x_{12}$  and the full interaction matrix  $\mathbf{A}$  in Table B.1 rank

deficient. To make the degradation term of *ASH1*,  $a_{44}$ , visible we rewrite (B.1.11)

$$\begin{aligned}
 sx_4(s) &= \underbrace{a_{11,9} \frac{a_{91}}{s - a_{99}}}_{=a_{41}(s)} x_1(s) + \underbrace{\left( a_{11,3} + a_{11,10} \frac{a_{10,3}}{s - a_{10,10}} \right)}_{=a_{43}(s)} x_3(s) \\
 &+ \underbrace{(a_{11,11} + a_{12,11})}_{a_{44}} \underbrace{(x_{11}(s) + x_{12}(s))}_{x_4(s)} \\
 &+ \underbrace{(a_{11,12} + a_{12,12} - a_{11,11} - a_{12,11})}_{a_{4,12}} x_{12}(s). \tag{B.1.15}
 \end{aligned}$$

To simplify this expression we assume that  $a_{11,12} + a_{12,12} = a_{11,11} + a_{12,11}$ , since the last term then is zero and the degradation of *ASH1* mRNA is  $a_{11,11} + a_{12,11}$ . We assume that only *ASH1f* can be translated and have therefore replaced  $a_{14}$  by  $a_{1,11}$ . The feedback loop formed by *ASH1f* and *ASH1b* then gives us two transfer functions. One from *SWI5* to *CBF1*

$$\begin{aligned}
 a_{13}(s) &= \frac{1}{(s - a_{10,10}) \bullet} \left( a_{13}(s - a_{10,10}) \bullet \right. \\
 &+ a_{1,11} a_{11,3} (s - a_{10,10}) (s - a_{12,12}) \\
 &+ a_{1,11} a_{11,10} a_{10,3} (s - a_{12,12}) \left. \right), \tag{B.1.16} \\
 \bullet &\triangleq (s - a_{11,11})(s - a_{12,12}) - a_{11,12} a_{12,11},
 \end{aligned}$$

and the self-loop

$$\begin{aligned}
 a_{11}(s) &= \frac{1}{(s - a_{99}) \bullet} \left( a_{11}(s - a_{99}) \bullet \right. \\
 &+ a_{1,11} a_{11,9} a_{91} (s - a_{12,12}) \left. \right), \tag{B.1.17}
 \end{aligned}$$

in which all non-diagonal coefficients  $a_{ij}$  are assumed positive, except  $a_{1,11}$ ,  $a_{11,10}$ , and  $a_{11,12}$ . These both contain a part of the transfer function

$$\begin{aligned}
 x_{11}(s) &= \frac{(s - a_{12,12})}{\bullet} \left( a_{11,3} x_3(s) + a_{11,9} x_9(s) \right. \\
 &+ a_{11,10} x_{10}(s) \left. \right), \tag{B.1.18}
 \end{aligned}$$

that describes how a change in *SWI5*, *Cbf1p*, and *Swi5p-Unk* affects *ASH1f* while accounting for the feedback loop through *ASH1b*. If we assume  $a_{12,12} = 0$ , then we have a pure integrator in the feedback loop and the steady-state of  $x_{11}$  is unaffected by the other states. Insertion of this in (B.1.16) and (B.1.17) gives

$$a_{13}(s) = a_{13} + a_{1,11} \frac{a_{11,3}(s - a_{10,10}) + a_{11,10} a_{10,3}}{(s - a_{10,10}) ((s - a_{11,11})s - a_{11,12} a_{12,11})} s \tag{B.1.19}$$

and

$$a_{11}(s) = a_{11} + a_{1,11} \frac{a_{1,11}a_{11,9}a_{91}}{(s - a_{99})((s - a_{11,11})s - a_{11,12}a_{12,11})} s. \quad (\text{B.1.20})$$

So *Cbf1p*, *Swi5p-Unk*, and *ASH1* have no influence on the steady-state value of *CBF1*, since  $a_{13}(0) = a_{13}$  and  $a_{11}(0) = a_{11}$ , which agrees with the steady-state data. Note that our previous assumption in Glucose,  $a_{11,10}a_{10,3} = a_{11,3}a_{10,10}$ , alone implies that neither of these states have any influence on the steady-state value of *CBF1*, so we do not need to assume that *ASH1b* is a pure integrator in Glucose. The interaction matrix of the observed states of the extended IRMA system is

$$\mathbf{A}(s) = \begin{bmatrix} a_{11}(s) & a_{12}(s) & a_{13}(s) & 0 & a_{15}(s) \\ a_{21}(s) & a_{22} & 0 & 0 & 0 \\ a_{31}(s) & a_{32}(s) & a_{33} & 0 & a_{35}(s) \\ a_{41}(s) & 0 & a_{43}(s) & a_{44} & 0 \\ 0 & 0 & a_{53}(s) & 0 & a_{55} \end{bmatrix}, \quad (\text{B.1.21})$$

with  $s$  in parenthesis marking the interactions that are transfer functions and given above, while the other interactions are constant coefficients.

### B.1.4 Identification of interaction coefficients in the extended IRMA example

To simulate perturbation experiments on our extended IRMA example and illustrate some properties of them, numerical values need to be assigned to the coefficients. We therefore identify the coefficients of the full interaction matrix in Table B.1 based on the inferred Glucose and Galactose models in Section 4.6. The full interaction matrix contains 35 parameters that we assume to be nonzero. While the steady-state data sets by Cantone *et al.* only contain 25 data points each so some parameters are structurally unidentifiable, implying that our parameter identification problem has many different solutions. Actually, only six parameters are practically selectable in Glucose and two in Galactose, implying that the number of practically identifiable parameters at best is equally low. For our illustrations it is, however, sufficient to select one of these solutions for Glucose and one for Galactose. We here select the ones that minimize the sum of the absolute values of the parameters while having identical steady-state response to the Glucose or Galactose models that we inferred in Section 4.6. The parameter values that we obtain are probably not biologically correct, but we lack necessary information for identification of biological values and this selection at least emphasizes low frequency behaviour. By emphasizing low frequency behaviour we avoid large overshoots and introduction of high frequency behaviour that the steady-state data cannot contain any information about. Both the resulting Glucose and Galactose models are stable systems, characterised by the real part of all eigenvalues of the interaction matrix being negative. The degree of interampatteness, Definition 3.2.2, of the Glucose model is 442 and Galactose model 854. For convenience the identified value and assumption on each coefficient of the full interaction matrix in Table B.1 is summarized in Table B.2 both for Glucose and Galactose. We assume all parameters to be zero that are not included in the table. The structure of our extended IRMA network is shown in Figure 6.1 for Glucose and Figure 6.2 for Galactose.

We now state how we assigned values to all parameters, including the three optimization problems that we solved to minimize the sum of the absolute value of the parameters. To make the steady-state response of the five states present in the original IRMA network identical to that of the Glucose or Galactose model in (A.6.3) and (A.6.6) respectively, we for (B.1.3) require that

$$\lim_{s \rightarrow 0} A(s) = \begin{cases} \hat{A}_{\text{Glu}} & \text{in Glucose,} \\ \hat{A}_{\text{Gal}} & \text{in Galactose.} \end{cases} \quad (\text{B.1.22})$$

**Table B.2: Coefficients (part 1/2).** Summary of the assumptions on and values assigned to all coefficients of the full interaction matrix  $\mathbf{A}$  in Table B.1. The assumptions are denoted by + positive, 0 zero, and - negative. If the same assumption is made in the Galactose model (Gal) as in the Glucose model (Glu), then the element is left empty. Part 2 is shown in Table B.3.

Coeff.	Assumption		Value	
	Glu	Gal	Glu	Gal
$a_{11}$	-		-0.143	-0.349
$a_{13}$	+		0.073	0.237
$a_{16}$	+	0	0.0223	0
$a_{1,11}$	-		-0.001	-0.001
$a_{21}$	-		-0.016	-0.016
$a_{22}$	-		-0.183	-0.227
$a_{29}$	+		0.0016	0.0016
$a_{33}$	-		-0.219	-0.407
$a_{36}$	+		0.0471	0.0471
$a_{39}$	+		0.0103	0.0196
$a_{43}$	+		0.0567	0.0567
$a_{44}$	-		-0.177	-0.205
$a_{49}$	+		0.0036	0.0073
$a_{4,10}$	-		-0.0126	-0.001
$a_{4,12}$	0		0	0
$a_{53}$	+		0.0135	0.0135
$a_{55}$	-		-0.152	-0.141
$a_{5,10}$	-		-0.001	-0.003
$a_{62}$	+		0.0365	0.0365
$a_{66}$	-		-0.0061	-0.0061
$a_{68}$	-		-0.0228	-0.0228
$a_{75}$	+		0.0329	0.0329
$a_{77}$	-		-0.0112	-0.0112
$a_{78}$	-		-0.0195	-0.0195
$a_{86}$	+		0.0115	0.0115
$a_{87}$	0	+	0	0.0314
$a_{88}$	-		-0.0206	-0.0206
$a_{91}$	+		0.01	0.01
$a_{99}$	-		-0.001	-0.001

**Table B.3: Coefficients (part 2/2).** Summary of the assumptions on and values assigned to all coefficients of the full interaction matrix  $\mathbf{A}$  in Table B.1. The assumptions are denoted by + positive, 0 zero, and - negative. If the same assumption is made in the Galactose model (Gal) as in the Glucose model (Glu), then the element is left empty. Part 1 is shown in Table B.2.

Coeff.	Assumption		Value	
	Glu	Gal	Glu	Gal
$a_{10,3}$	+		0.0045	0.0045
$a_{10,10}$	-		-0.001	-0.001
$a_{11,3}$	+		0.0567	0.0567
$a_{11,9}$	+		0.0036	0.0073
$a_{11,10}$	-		-0.0126	-0.001
$a_{11,11}$	-		-0.2166	-0.2166
$a_{11,12}$	-		-0.177	-0.205
$a_{12,11}$	+		0.0396	0.0116
$a_{12,12}$	0		0	0

This provides us with conditions on each coefficient and transfer function in (B.1.21), which we next use to assign values to the parameters. The following parameters are thus assigned the value of the corresponding parameter in  $\hat{\mathbf{A}}_{\text{Glu}}$  or  $\hat{\mathbf{A}}_{\text{Gal}}$ :  $a_{11}$ ,  $a_{22}$ ,  $a_{33}$ ,  $a_{44}$ ,  $a_{55}$ , and  $a_{13}$ . We determine all parameters related to the Gal4p-Gal80p complex,  $\boldsymbol{\theta}_{\text{g}} \triangleq [a_{16}, a_{36}, a_{62}, a_{66}, a_{68}, a_{75}, a_{77}, a_{78}, a_{86}, a_{87}, a_{88}, a_{16}^*, a_{87}^*]^T$ , by solving the following optimization problem:

$$\min_{\boldsymbol{\theta}_{\text{g}}} \sum_{i=1}^{13} |\theta_i^{\text{g}}| \quad (\text{B.1.23a})$$

$$\text{s.t. } \boldsymbol{\theta}_{\text{g}} \in \mathcal{T}_{\text{g}}, \quad (\text{B.1.23b})$$

$$\frac{a_{36}a_{62}(a_{77}a_{88} - a_{78}a_{87})}{\blacktriangleleft} = a_{32}^{\text{Glu}}, \quad (\text{B.1.23c})$$

$$\frac{a_{36}a_{68}a_{87}a_{75}}{\blacktriangleleft} = a_{35}^{\text{Glu}}, \quad (\text{B.1.23d})$$

$$\frac{a_{16}a_{62}(a_{77}a_{88} - a_{78}a_{87})}{\blacktriangleleft} = a_{12}^{\text{Glu}}, \quad (\text{B.1.23e})$$

$$\frac{a_{16}a_{68}a_{87}a_{75}}{\blacktriangleleft} = a_{15}^{\text{Glu}}, \quad (\text{B.1.23f})$$

$$\frac{a_{36}a_{62}(a_{77}a_{88} - a_{78}a_{87}^*)}{\blacktriangleright} = a_{32}^{\text{Gal}}, \quad (\text{B.1.23g})$$

$$\frac{a_{36}a_{68}a_{87}^*a_{75}}{\blacktriangleright} = a_{35}^{\text{Gal}}, \quad (\text{B.1.23h})$$

$$\frac{a_{16}^*a_{62}(a_{77}a_{88} - a_{78}a_{87}^*)}{\blacktriangleright} = a_{12}^{\text{Gal}}, \quad (\text{B.1.23i})$$

$$\frac{a_{16}^*a_{68}a_{87}^*a_{75}}{\blacktriangleright} = a_{15}^{\text{Gal}}, \quad (\text{B.1.23j})$$

$$\blacktriangleleft \triangleq -a_{66}a_{77}a_{88} + a_{78}a_{87}a_{66} + a_{68}a_{86}a_{77},$$

$$\blacktriangleright \triangleq -a_{66}a_{77}a_{88} + a_{78}a_{87}^*a_{66} + a_{68}a_{86}a_{77}.$$

Here  $\mathcal{T}_{\text{g}}$  denotes the set of parameter values fulfilling the assumptions made in the previous section and listed in Table B.2. Parameter  $a_{16}$  and  $a_{87}$  take different values in Glucose and Galactose, and we therefore denote the Galactose values by  $a_{16}^*$  and  $a_{87}^*$ , respectively. This nomenclature is also used in the following two optimization problems. We determine all parameters related to Cbf1p,  $\boldsymbol{\theta}_{\text{c}} \triangleq [a_{21}, a_{29}, a_{39}, a_{91}, a_{99}, a_{11,9}, a_{39}^*, a_{11,9}^*]^T$ , by solving the following optimization

problem:

$$\min_{\theta_c} \sum_{i=1}^8 |\theta_i^c| \quad (\text{B.1.24a})$$

$$\text{s.t. } \theta_c \in \mathcal{T}_c, \quad (\text{B.1.24b})$$

$$- \frac{a_{39}a_{91}}{a_{99}} = a_{31}^{\text{Glu}}, \quad (\text{B.1.24c})$$

$$- \frac{a_{11,9}a_{91}}{a_{99}} = a_{41}^{\text{Glu}}, \quad (\text{B.1.24d})$$

$$\frac{a_{21}a_{99} - a_{29}a_{91}}{a_{99}} = a_{21}^{\text{Glu}}, \quad (\text{B.1.24e})$$

$$- \frac{a_{39}^*a_{91}}{a_{99}} = a_{31}^{\text{Gal}}, \quad (\text{B.1.24f})$$

$$- \frac{a_{11,9}^*a_{91}}{a_{99}} = a_{41}^{\text{Gal}}. \quad (\text{B.1.24g})$$

Similarly, we determine all parameters related to *Swi5p-Unk* and *ASH1*,  $\theta_s \triangleq [a_{53}, a_{5,10}, a_{10,3}, a_{10,10}, a_{11,3}, a_{11,10}, a_{11,11}, a_{12,11}, a_{5,10}^*, a_{11,10}^*, a_{12,11}^*]^T$ , by solving the following optimization problem:

$$\min_{\theta_s} \sum_{i=1}^{11} |\theta_i^s| \quad (\text{B.1.25a})$$

$$\text{s.t. } \theta_s \in \mathcal{T}_s, \quad (\text{B.1.25b})$$

$$\frac{a_{11,3}a_{10,10} - a_{11,10}a_{10,3}}{a_{10,10}} = a_{43}^{\text{Glu}}, \quad (\text{B.1.25c})$$

$$\frac{a_{53}a_{10,10} - a_{5,10}a_{10,3}}{a_{10,10}} = a_{53}^{\text{Glu}}, \quad (\text{B.1.25d})$$

$$a_{11,11}a_{12,11} = a_{44}^{\text{Glu}}, \quad (\text{B.1.25e})$$

$$\frac{a_{11,3}a_{10,10} - a_{11,10}^*a_{10,3}}{a_{10,10}} = a_{43}^{\text{Gal}}, \quad (\text{B.1.25f})$$

$$\frac{a_{53}a_{10,10} - a_{5,10}^*a_{10,3}}{a_{10,10}} = a_{53}^{\text{Gal}}, \quad (\text{B.1.25g})$$

$$a_{11,11}a_{12,11}^* = a_{44}^{\text{Gal}}. \quad (\text{B.1.25h})$$

Based on our assumptions in the previous section we set  $a_{11,12} = a_{44}$ ,  $a_{43} = a_{11,3}$ ,  $a_{49} = a_{11,9}$ ,  $a_{4,10} = a_{11,10}$ ,  $a_{4,12} = 0$ , and  $a_{12,12} = 0$ . We set  $a_{1,11} = -0.001$  to give this parameter a small negative value. The steady-state solution is not affected by  $a_{1,11}$ , so we could have selected any negative value since we previously merely assumed it to be negative. Finally, we make minor adjustment of the following coefficients to maintain exact cancellation of the influence from *CBF1* to *GAL4* and from *SWI5* to *ASH1* in Glucose and *GAL80* in Galactose when rounding the values to 4 digits:  $a_{21}$ ,  $a_{29}$ ,  $a_{91}$ ,  $a_{99}$ ,  $a_{53}$ ,  $a_{5,10}$ ,  $a_{10,3}$ ,  $a_{10,10}$ ,  $a_{11,3}$ , and  $a_{11,10}$ .

## B.2 Transfer function basics

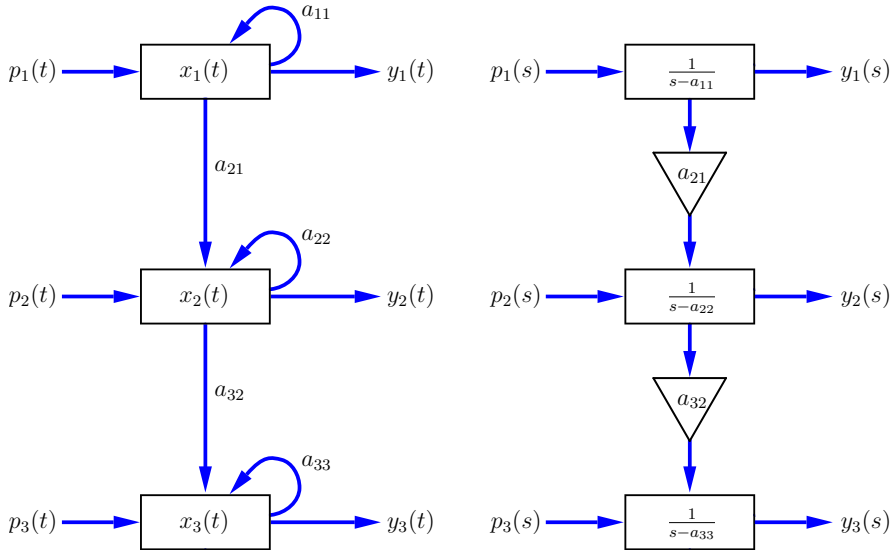
This section gives a brief introduction to transfer functions and properties of them, which we use in this work. We have also included references to basic textbooks in signals and systems theory where additional information can be found in order to provide a good starting point for readers unfamiliar with transfer functions. The relation between ODEs and transfer functions is also illustrated graphically, in order to stress that transfer functions merely are a different representation of the system.

Any linear state-space model can through Laplace transformation be written as a transfer function. The transfer function concept is widely used in signal and systems theory, since it simplifies the division of the system into blocks, *i.e.* subsystems, and we here use it in particular to represent the effect of latent states. Let us therefore briefly introduce the transfer function concept and show the relation between the network and block diagram views of a system. A good introduction to transfer functions and block diagrams can be found in *e.g.* Lindner (1999, ch. 10). The ODEs in (B.1.1) can be transformed one-by-one from the time domain, identified by dependence on  $t$ , to the Laplace domain, denoted by dependence on  $s$ ,

$$\dot{x}_i(t) = a_{i1}x_1(t) + \dots + a_{in}x_n(t) + p_i(t) \xrightarrow{\mathcal{L}^-} \quad (\text{B.2.1})$$

$$\begin{aligned} s x_i(s) - x_i(0^-) &= a_{i1}x_1(s) + \dots + a_{ii}x_i(s) + \dots + a_{in}x_n(s) + p_i(s) \Rightarrow \\ x_i(s) &= \frac{a_{i1}}{s - a_{ii}}x_1(s) + \dots + \frac{a_{i,i-1}}{s - a_{ii}}x_{i-1}(s) + \frac{a_{i,i+1}}{s - a_{ii}}x_{i+1}(s) + \dots \\ &\quad + \frac{a_{in}}{s - a_{ii}}x_n(s) + \frac{1}{s - a_{ii}}(p_i(s) + x_i(0^-)). \end{aligned} \quad (\text{B.2.2})$$

Here  $p_i(s)$  denotes the Laplace transform of the exogenous input, and  $x_i(0^-)$  denotes the initial value. We typically assume that the initial value is zero, because it simplifies the calculations, it is zero at the steady-state, and any nonzero initial value acts as a perturbation and can be compensated for in the perturbation, as evident from (B.2.2). A transfer function is defined as the ratio between the Laplace transform of an input and output signal (Lindner, 1999, p. 297), so if we consider  $x_i$  as our output then  $G_i(s) = \frac{1}{s - a_{ii}}$  is the transfer function from the exogenous input  $p_i$  to our output. If we instead studied the direct effect of  $x_j$ ,  $j \neq i$  on  $x_i$  then  $G_{ij}(s) = \frac{a_{ij}}{s - a_{ii}}$  would be the transfer function of interest. One can hence easily introduce different transfer functions depending on the signals of interest. The effect of a subsystem containing latent states can typically be represented in a concise manner using transfer functions. The relation between different signals is typically visualized in a block diagram, with each block representing a transfer function. The main difference between the block diagram and the network graph of a system is that the signals are explicit in the first one, while the state variables are explicit in the later one. For illustration both the network graph and the block diagram of a cascade system with three states is shown in Figure B.2. Another advantage of transfer functions, except the explicit representation of signals, is that it is easy to calculate the final value and steady-state response. The final value is given by the



**Figure B.2: Graphical representation of systems.** Network graph (left) and block diagram (right) of a cascade system with three states. Here each state is perturbed and measured. The state variables are explicit in the network graph, while the signals are explicit in the block diagram.

final value theorem (Lindner, 1999, p. 266)

$$\lim_{t \rightarrow \infty} x(t) = \lim_{s \rightarrow 0} sx(s). \quad (\text{B.2.3})$$

The steady-state response, *i.e.* the response to a constant input, is consequently the strength of the input times the static gain, which is obtained by taking the limit of the transfer function as the frequency  $\omega \rightarrow 0$ , *e.g.*  $G_i(0) = \lim_{\omega \rightarrow 0} G_i(i\omega) = \frac{1}{-a_{ii}}$ . Note that  $s$  is replaced by  $i\omega$ , *i.e.* we evaluate the limit along the imaginary axis of the complex plane. We here use this way of calculating the steady-state response, since it normally is easier than setting all derivatives equal to zero and solving the system of ODEs.

For convenience of the reader we now define several properties of transfer functions that we use frequently in this work. Consider a rational transfer function of the form

$$G(s) = \frac{b_{n_z} s^{n_z} + \dots + b_1 s + b_0}{s^n + a_{n-1} s^{n-1} + \dots + a_1 s + a_0}. \quad (\text{B.2.4})$$

The order of the transfer function is equal to the order of the denominator or pole polynomial  $n$ , which in the single input-single output case also is the order of the system that it represents, *i.e.* the number of states in the system (Lindner, 1999, p. 299). The roots of the denominator polynomial is typically called poles in control

theory, while the roots of the numerator polynomial is called zeros. The order of the numerator or zero polynomial is equal to the number of zeros  $n_z$ . The relative order is equal to the pole excess  $n - n_z$  (Skogestad and Postlethwaite, 1996, p. 4). A transfer function  $G(s)$  is strictly proper if  $n > n_z$  or equivalently if  $G(i\omega) \rightarrow 0$  as  $\omega \rightarrow \infty$ , *i.e.* the gain goes to zero at high frequencies  $\omega$ . And a transfer function  $G(s)$  is proper if  $n \geq n_z$  or equivalently if  $G(i\omega) \rightarrow C$  as  $\omega \rightarrow \infty$ , with  $C$  being some finite constant. The transfer function is said to be semi-proper if  $n = n_z$ . Note that all practical systems have zero gain at sufficiently high frequency and they are therefore strictly proper. Some transfer functions are though proper due to our division of the system into subsystems.



---

## Design of perturbation experiments in depth

---

We here present additional information on the Snf1 example and *in vivo* steady-state data that we use in Chapters 4 and 7. The alternative network models that we have inferred are described in Sections C.1 and C.2 together with the residual sum of squares and model rejection criteria that we used. The *in silico* experiments used to demonstrate the SVD design in Section 7.6 are described in Section C.3.

## C.1 Additional information on the Snf1 example

The Snf1 example and *in vivo* steady-state data that we use in Chapter 7 was published in Lorenz et al. (2009). Here we only present information on the alternative network models that we have inferred. Readers interested in further details should see the original publication and its supplemental.

Two models are indistinguishable based on input-output data if the response of both models to all possible perturbations is identical, implying that the parameters of the models are structurally unidentifiable (Faller et al., 2003; Ljung, 1999). In practice models are compared based on how well they explain observed input-output data. The residuals or prediction errors is a measure of the difference between the observed values of the dependent variables and the values predicted by the model (Jaqaman and Danuser, 2006; Ljung, 1999). All measurements are corrupted by noise and the expected norm of the residuals is therefore nonzero. In order for two models to be practically distinguishable based on input-output data the difference between the norm of the residuals must be sufficiently large compared to the expected value, *i.e.* the difference must be statistically significant. Several statistical hypothesis tests exist for comparing alternative models, *e.g.* the F-test (Cedersund and Roll, 2009). Instead of comparing the relative fit of models, one can evaluate the goodness of fit of a model directly based on the assumed noise model. We take the common engineering view that a model can not be validated, but if the norm of the residuals is sufficiently large compared to the expected value, then it can and should be rejected at a specified confidence or significance level, since it then is unlikely to be useful. One such test which we use is the  $\chi^2$  goodness of fit test. It is based on the expected minimised weighted sum of squared residuals being equal to the number of degrees of freedom (Jaqaman and Danuser, 2006; Cedersund and Roll, 2009; Pearson, 1900). The  $\chi^2$  test statistic of the errors  $v_{ij}$ , which are distributed according to a multivariate normal distribution with zero mean and covariance matrix  $\Lambda_{\mathbf{Y}}$ , is

$$\begin{aligned} \hat{\chi}^2(\mathbf{A}) &\triangleq \vec{\mathbf{Y}}^T \Lambda_{\mathbf{Y}}^{-1} \vec{\mathbf{Y}} \\ &= \begin{cases} \sum_{i=1}^m \sum_{j=1}^n \frac{v_{ij}^2}{\lambda_{ij}} & \text{if } \Lambda_{\mathbf{Y}} = \text{diag}(\lambda_{11}, \dots, \lambda_{mn}) \\ \|\mathbf{Y}\|_{Fro}^2 / \lambda & \text{if } \Lambda_{\mathbf{Y}} = \lambda \mathbf{I} \end{cases} . \end{aligned} \quad (\text{C.1.1})$$

If this test statistic is larger than  $\chi^{-2}(\alpha, nm)$ , which denotes the inverse of the chi-square cumulative distribution with  $nm$  degrees of freedom such that  $\text{P}[\chi^2(nm) > \chi^{-2}(\alpha, nm)] = \alpha$ , then the corresponding model  $\mathbf{A}$  is rejected at the significance

level  $\alpha$  (Chew, 1966). We here use the significance level  $\alpha = 0.05$  and reject all models with p-values smaller than 0.05, corresponding to a 95% confidence level. By selecting the confidence level we select how conservative the test is. This because we use model rejection to do robust network inference, as explained in detail in Chapter 5. In short, if we can reject all models with a specific parameter set to zero, then the model structure needed to adequately represent the input-output data must contain this parameter and the data is informative enough for rejecting all models lacking this parameter. In network inference each parameter of the linear models that we restrict ourselves to corresponds to a link, so if all models in which a specific parameter is zero are rejected, then the link is statistically significant at the selected significance level.

Lorenz et al. (2009) assumed the data model

$$p_{ij} = - \sum_{k=1}^n a_{ik} y_{kj} \quad (\text{C.1.2})$$

where both the perturbations  $p_{ij}$  and responses  $y_{kj}$  are measured in each experiment  $j$  with errors assumed to be independent and zero mean normally distributed. Errors in both the input and output of the system implies that we from a parameter estimation perspective are dealing with an errors in variables regression problem (see e.g. Griliches and Hausman, 1986; Söderström, 2007). Lorenz et al. provide estimates of the standard error of each measured response both for the perturbations and responses and we therefore use weighted residual sums of squares to measure the goodness of fit to observed data and calculate the  $\chi^2$  test statistic of both the perturbations and responses. Like all estimates these standard errors are associated with a degree of uncertainty and we therefore also use unweighted residual sums of squares. To obtain alternative network models that cannot be rejected, we first generated a large set of alternative models with structure picked at random and fitted them to the data recorded by Lorenz et al. (2009). We minimized a mixture of unweighted and weighted residual sums of squares for the perturbations and responses to obtain candidates. We then selected a subset of these that had desired properties by testing if they could be rejected, as described next.

Since we have an errors in variables regression problem, we define the weighted residual sum of squares (WRSS) as

$$\text{WRSS}(\mathbf{Y}) \triangleq \sum_{j=1}^m (\hat{\mathbf{y}}_j - \mathbf{y}_j)^T \mathbf{C}_{\mathbf{y}_j}^{-1} (\hat{\mathbf{y}}_j - \mathbf{y}_j) \quad \text{with } \hat{\mathbf{Y}} \text{ given by (C.1.7)} \quad (\text{C.1.3})$$

and

$$\text{WRSS}(\mathbf{P}) \triangleq \sum_{j=1}^m (\tilde{\mathbf{p}}_j - \mathbf{p}_j)^T \mathbf{C}_{\mathbf{p}_j}^{-1} (\tilde{\mathbf{p}}_j - \mathbf{p}_j) \quad \text{with } \hat{\mathbf{P}} \text{ given by (C.1.7)}. \quad (\text{C.1.4})$$

And the unweighted residual sum of squares (RSS) as

$$\text{RSS}(\mathbf{Y}) \triangleq \left\| \hat{\mathbf{Y}} - \mathbf{Y} \right\|_{\text{Fro}}^2 \quad \text{with } \hat{\mathbf{Y}} \text{ given by (C.1.7)} \quad (\text{C.1.5})$$

and

$$\text{RSS}(\mathbf{P}) \triangleq \left\| \hat{\mathbf{P}} - \mathbf{P} \right\|_{\text{Fro}}^2 \quad \text{with } \hat{\mathbf{P}} \text{ given by (C.1.7)}. \quad (\text{C.1.6})$$

For any given interaction matrix  $\mathbf{A}$  we obtain the optimal response and perturbation matrices by solving the optimisation problem

$$\{\hat{\mathbf{Y}}, \hat{\mathbf{P}}\} = \arg \min_{\mathbf{Y}, \mathbf{P}} \sum_{j=1}^m (\tilde{\mathbf{y}}_j - \mathbf{y}_j)^T \mathbf{C}_{\mathbf{y}_j}^{-1} (\tilde{\mathbf{y}}_j - \mathbf{y}_j) + (\tilde{\mathbf{p}}_j - \mathbf{p}_j)^T \mathbf{C}_{\mathbf{p}_j}^{-1} (\tilde{\mathbf{p}}_j - \mathbf{p}_j) \quad (\text{C.1.7a})$$

$$\text{s.t. } \mathbf{A}\tilde{\mathbf{Y}} = -\tilde{\mathbf{P}} \quad (\text{C.1.7b})$$

$$\left\| \tilde{\mathbf{Y}} - \mathbf{Y} \right\|_{\text{Fro}} \leq 8.7 \quad (\text{C.1.7c})$$

$$\left\| \tilde{\mathbf{P}} - \mathbf{P} \right\|_{\text{Fro}} \leq 1.8 \quad (\text{C.1.7d})$$

with the observed response matrix  $\mathbf{Y} \triangleq [\mathbf{y}_1, \dots, \mathbf{y}_j, \dots, \mathbf{y}_m]$ , observed perturbation matrix  $\mathbf{P} \triangleq [\mathbf{p}_1, \dots, \mathbf{p}_j, \dots, \mathbf{p}_m]$ , covariance matrix  $\mathbf{C}_{\mathbf{y}_j}$  of the responses in experiment  $j$ , and covariance matrix  $\mathbf{C}_{\mathbf{p}_j}$  of the perturbations in experiment  $j$ . Similar to [Lorenz et al. \(2009\)](#) we assumed that each measurement is independent and used the square of the standard errors of the fold changes reported in their supplementary Table S3 as estimates of the variance of each observed response, yielding diagonal covariance matrices  $\mathbf{C}_{\mathbf{y}_j}$ . We also assumed that each perturbation is independent and used the square of the standard errors of each measured perturbation reported in their supplementary Table S4, except for unperturbed genes that we assumed to have a standard error of 0.1, as estimates of the variance of each perturbation, yielding diagonal covariance matrices  $\mathbf{C}_{\mathbf{p}_j}$ . The standard error 0.1 is approximately one quarter to one third of the reported standard errors of the perturbed genes and chosen in lack of better information to represent biological variation in the cell cultures. The constraints of the unweighted residual sums of squares 8.7 and 1.8 are based on the expected Frobenius norm, *i.e.* square root of the residual sum of squares, at significance level 0.05 being 8.74 for the responses and 1.86 for the perturbations using the covariance matrices above in a Monte Carlo simulation with 10000 random noise realizations. They are included to ensure that the models cannot be rejected based on the RSS. We solve this problem in Matlab ([www.mathworks.com](http://www.mathworks.com)) using CVX ([cvxr.com/cvx](http://cvxr.com/cvx)). The residual sums of squares of the three models in Table C.2, Table C.4, and Table C.6 for the corresponding optimal perturbation matrices in Table C.3, Table C.5, and Table C.7 are given in Table C.1. Note that we here only report the interaction matrices and optimal perturbation matrices, since the optimal response matrices are obtained from the former as  $\hat{\mathbf{Y}} = -\mathbf{A}^\dagger \hat{\mathbf{P}}$ , with  $\mathbf{A}^\dagger$  denoting the Moore-Penrose generalized inverse ([Horn and Johnson, 1990](#), p. 421).

**Table C.1: Residual sum of squares for the Snf1 interaction matrices in Table C.2, Table C.4, and Table C.6.** These are calculated using the corresponding optimal perturbation matrices in Table C.3, Table C.5, and Table C.7 and responses given by  $\hat{\mathbf{Y}} = -\mathbf{A}^{-1}\hat{\mathbf{P}}$ . The weighted residual sum of squares is defined in (C.1.3) and (C.1.4), while the unweighted residual sum of squares is defined in (C.1.5) and (C.1.6).

Measure \ model	$\mathbf{A}_{\text{fin}}$ TC.2	$\mathbf{A}_1$ TC.4	$\mathbf{A}_2$ TC.6
WRSS( $\mathbf{Y}$ )	92.1	48.6	39.5
WRSS( $\mathbf{P}$ )	98.2	66.6	44.8
RSS( $\mathbf{Y}$ )	1.73	0.828	1.28
RSS( $\mathbf{P}$ )	3.24	0.694	0.513

**Table C.2: The interaction matrix inferred by Lorenz *et al.* for the Snf1 network. Reproduced from supplementary Table S9 in Lorenz *et al.* (2009).**

Gene $i \setminus j$	CAT8	FBP1	HXK2	MED8	MIG1	SIP2	SIP4	SNF1	SNF4	SUC2
CAT8	-0.485	-0.0446	0.109	-0.198	0.112	0	0	0.431	-0.024	0.016
FBP1	-0.027	-0.832	0.549	0.17	0.028	-0.025	0	0.271	0	0
HXK2	0	0	-4.258	1.521	0	0	0	-1.307	0	0
MED8	0	0	2.093	0	0	0.133	0	1.049	0	0
MIG1	-0.012	0	0.169	0.719	-0.843	-0.063	0	0	0.016	0.025
SIP2	-0.023	0	0.699	0.089	0.114	-0.783	0	0.585	0	0
SIP4	0.003	0	0	0	0.005	-0.009	-0.085	0	0	0
SNF1	-0.013	0	0.66	1.52	0.053	0	0	-1.599	0.036	0.051
SNF4	0.053	0	0	0.68	-0.08	0.032	0	0.042	-0.564	0.049
SUC2	-0.058	0	0.666	0.675	0.078	0.02	0	0.143	0.014	-0.811

**Table C-3: The optimal perturbation matrix corresponding to the Snf1 interaction matrix inferred by Lorenz et al.** This perturbation matrix minimizes the weighted residual sum of squares of the responses and perturbations while fulfilling constraints on the unweighted residual sum of squares in (C.1.7) for the interaction matrix in Table C.2.

Experiment	1	2	3	4	5	6	7	8	9	10
	Perturbed gene									
<i>CAT8</i>	2.123	-0.001	0.014	-0.001	-0.013	-0.04	-0.003	-0.002	0.002	-0.004
<i>FBP1</i>	0.022	2.137	0.038	0.063	-0.013	-0.029	-0.082	0.048	0.007	0.01
<i>HXK2</i>	0.099	-0.017	2.382	-0.237	0.036	-0.077	0.147	-0.054	0.024	0.022
<i>MED8</i>	0.158	-0.025	-0.63	0.8	0.075	-0.092	0.346	-0.097	0.058	0.027
<i>MIG1</i>	-0.021	0.021	0.031	0.066	2.134	0.033	-0.029	0.011	0.012	-0.02
<i>SIP2</i>	0.034	-0.021	-0.053	0.041	-0.008	2.095	-0.029	-0.003	-0.022	0.021
<i>SIP4</i>	-0.001	0.019	0.004	0.001	0.001	-0.002	2.139	0.002	0.003	0.007
<i>SNF1</i>	0.008	-0.025	-0.024	-0.017	-0.002	0.014	0.02	1.423	-0.013	0.014
<i>SNF4</i>	-0.028	0.034	0.176	0.156	-0.022	0.006	-0.07	-0.002	3.181	-0.016
<i>SUC2</i>	0.003	0.02	0.127	0.116	0.007	-0.005	-0.064	0.014	-0.013	2.453



**Table C.5: The optimal perturbation matrix corresponding to alternative interaction matrix 1.** This perturbation matrix minimizes the weighted residual sum of squares of the responses and perturbations while fulfilling constraints on the unweighted residual sum of squares in (C.1.7) for the interaction matrix in Table C.4.

Experiment	1	2	3	4	5	6	7	8	9	10
Perturbed gene										
<i>CAT8</i>	2.113	-0.069	0.008	-0.125	0.127	-0.106	-0.079	0.157	-0.048	0.05
<i>FBP1</i>	-0.106	2.138	0.069	-0.027	-0.018	-0.159	-0.195	0.041	0	-0.003
<i>HXK2</i>	0.008	-0.002	3.117	-0.003	0.002	0.005	0	-0.005	0	-0.002
<i>MED8</i>	-0.014	0.008	-0.018	1.986	0.001	-0.023	-0.006	-0.007	-0.008	-0.022
<i>MIG1</i>	-0.143	0.036	-0.04	-0.081	2.13	-0.21	-0.231	-0.119	0.028	-0.01
<i>SIP2</i>	-0.015	-0.014	-0.005	0.007	0.013	2.156	0.003	-0.014	-0.01	0.016
<i>SIP4</i>	0.013	0.019	0.003	0.004	0.012	-0.026	2.164	0.001	0.004	0.007
<i>SNF1</i>	0.009	-0.014	0	-0.024	-0.029	-0.107	0.005	1.485	-0.023	0.024
<i>SNF4</i>	0.052	0.048	0.03	-0.013	-0.193	-0.027	-0.277	-0.041	3.183	0.063
<i>SUC2</i>	-0.305	0.038	0.09	-0.103	0.026	-0.095	-0.27	-0.12	-0.013	2.457

Table C.6: Alternative interaction matrix 2 that explains the steady-state data recorded by [Lorenz et al. \(2009\)](#) for the Snfl network.

Gene $i \setminus j$	CAT8	FBP1	HXK2	MED8	MIG1	SIP2	SIP4	SNF1	SNF4	SUC2
CAT8	-0.494	0	0	0	0	0	0	0	0	0
FBP1	0	-0.839	0	0	0	0	0	0	0	0
HXK2	0.713	0	-7.369	8.312	0.849	0	0.169	0	0	0
MED8	0.746	0	0	9.426	0	0.758	0.212	3.445	0.142	0
MIG1	0	0	0	0.931	-0.936	0	0	0	0	0
SIP2	0	0	0	0	0	-0.832	0	0	0	0
SIP4	0	0	0	0	0	0	-0.096	0	0	0
SNF1	0	0	0	2.051	0	0	0	-1.919	0	0
SNF4	0	0	0	0	0	0	0	0	-0.59	0
SUC2	0	0	0	1.252	0	0	0	0	0	-0.835

**Table C.7: The optimal perturbation matrix corresponding to alternative interaction matrix 2.** This perturbation matrix minimizes the weighted residual sum of squares of the responses and perturbations while fulfilling constraints on the unweighted residual sum of squares in (C.1.7) for the interaction matrix in Table C.6.

Experiment	1	2	3	4	5	6	7	8	9	10
	Perturbed gene									
<i>CAT8</i>	2.032	-0.073	0.002	-0.129	0.134	-0.106	-0.077	0.143	-0.053	0.05
<i>FBP1</i>	-0.106	2.138	0.069	-0.027	-0.018	-0.159	-0.195	0.041	0	-0.003
<i>HXK2</i>	0	-0.001	3.165	-0.003	-0.001	0.021	0	-0.016	0.002	-0.001
<i>MED8</i>	-0.006	-0.005	0.004	2.193	0.015	-0.008	0.006	-0.004	0.001	0.007
<i>MIG1</i>	0.013	0.037	-0.046	-0.032	2.144	-0.119	0.04	-0.005	0.036	-0.008
<i>SIP2</i>	-0.142	-0.017	0.057	-0.125	0.11	2.052	-0.215	0.062	-0.04	0.017
<i>SIP4</i>	0.014	0.021	0.003	0.004	0.014	-0.03	2.354	0.001	0.004	0.008
<i>SNF1</i>	0.022	-0.025	-0.005	-0.01	-0.001	-0.017	-0.012	1.414	-0.013	0.036
<i>SNF4</i>	0.052	0.047	0.03	-0.013	-0.192	-0.027	-0.279	-0.042	3.194	0.063
<i>SUC2</i>	-0.069	0.035	0.093	-0.018	0.036	0.017	0.043	0.022	0.011	2.467

## C.2 Non-rejectable network models

We here prove that the steady-state data recorded by [Lorenz et al. \(2009\)](#) for the Snf1 network is not informative enough for robust inference of anyone of the possible interactions, by for each possible interaction finding at least one model lacking it that explains their data. If a model lacking interaction  $a_{ij}$ , corresponding to an interaction matrix with  $a_{ij} = 0$ , that can explain the data exists, *i.e.* it cannot be rejected at the desired significance level, then the data set is not informative enough for robust inference of this interaction, since it is not possible to prove that this interaction is necessary to explain the data.

Of the 100 possible interactions 84 is lacking in either of the two alternative models in [Table C.4](#) and [Table C.6](#). We here present 16 additional models that each lack one of the 16 interactions present in both of the previous two models. Together these 18 models prove that the steady-state data recorded by [Lorenz et al. \(2009\)](#) for the Snf1 network is not informative enough for robust inference of anyone of the possible interactions at any significance level below 0.05. The expected WRSS, given by an inverse  $\chi^2$  distribution with 100 degrees of freedom, is 124.3, the expected  $\text{RSS}(\mathbf{Y})$  is 76.4, and the expected  $\text{RSS}(\mathbf{P})$  is 3.46, at significance level 0.05, using the noise assumptions stated in [Section C.1](#). Clearly all residual sums of squares are for all models below these values, as seen in [Table C.1](#) and [Table C.8](#), so the existence of these models proves that the data is not informative enough. As a curiosity, we note that the final model inferred by [Lorenz et al.](#)  $\mathbf{A}_{\text{fin}}$  has higher residual sums of squares than all of our alternative models with the exception of  $\text{RSS}(\mathbf{Y})$  of  $\mathbf{A}_{11}$  and  $\mathbf{A}_{17}$ , so it will in general be rejected before any of our alternative models are.

Our alternative interaction matrices together with the corresponding optimal perturbation matrices are here reported in [Table C.9](#) to [Table C.40](#). Note that we here only report the interaction matrices and optimal perturbation matrices, since the optimal response matrices are obtained from the former as  $\hat{\mathbf{Y}} = -\mathbf{A}^\dagger \hat{\mathbf{P}}$ , with  $\mathbf{A}^\dagger$  denoting the Moore-Penrose generalized inverse ([Horn and Johnson, 1990](#), p. 421).

**Table C.8: Residual sum of squares for the Snf1 interaction matrices in Table C.9-C.39 used to prove that a non-rejectable model lacking any possible interaction exists.** These are calculated using the corresponding optimal perturbation matrices in Table C.10-C.40 and responses given by  $\hat{Y} = -A^{-1}\hat{P}$ . The weighted residual sum of squares is defined in (C.1.3) and (C.1.4), while the unweighted residual sum of squares is defined in (C.1.5) and (C.1.6).

Model \ measure	WRSS( $Y$ )	WRSS( $P$ )	RSS( $Y$ )	RSS( $P$ )
$A_3$ TC.9	7.76	21.9	0.296	1.07
$A_4$ TC.11	13.5	38.5	1.37	2.18
$A_5$ TC.13	10.6	4.75	0.248	0.179
$A_6$ TC.15	6.73	5.9	0.138	0.459
$A_7$ TC.17	11.3	4.6	0.207	0.205
$A_8$ TC.19	8.6	3.62	0.229	0.137
$A_9$ TC.21	9.78	8.65	0.289	0.468
$A_{10}$ TC.23	10.9	10.5	0.279	0.803
$A_{11}$ TC.25	12.4	15	2.65	1.18
$A_{12}$ TC.27	10.7	19.6	0.38	1.21
$A_{13}$ TC.29	9.71	22.4	0.345	1.62
$A_{14}$ TC.31	0.66	15.1	0.078	0.345
$A_{15}$ TC.33	2.34	0.997	0.036	0.0157
$A_{16}$ TC.35	4.91	5.38	0.0867	0.303
$A_{17}$ TC.37	15.5	52.9	3.66	3.24
$A_{18}$ TC.39	13.2	66.2	0.744	3.24

Table C.9: Alternative interaction matrix 3 that explains the steady-state data recorded by Lorenz et al. (2009) for the Snf1 network and lacks the  $a_{11}$  interaction.

Gene $i \setminus j$	CAT8	FBP1	HXK2	MED8	MIG1	SIP2	SIP4	SNF1	SNF4	SUC2
CAT8	0	-0.269	0.369	6.629	0.521	0.463	0.155	1.226	0.18	0.207
FBP1	-0.117	-0.767	0.51	-1.413	-0.118	-0.131	-0.037	0.197	-0.035	-0.042
HXK2	0.988	-0.75	-3.824	15.86	1.353	0.792	0.342	0.501	0.421	0.449
MED8	0.791	-0.478	2.985	13.18	0.791	1.07	0.321	3.079	0.415	0.352
MIG1	0.096	-0.049	0.254	2.564	-0.735	0.067	0.044	0.231	0.089	0.067
SIP2	-0.068	0.027	0.691	-0.634	0.014	-0.839	-0.02	0.493	-0.023	-0.013
SIP4	0	0.009	-0.002	-0.051	0.002	-0.012	-0.086	-0.005	-0.001	0.001
SNF1	0.094	-0.193	0.788	3.412	0.113	0.125	0.041	-1.318	0.056	0.129
SNF4	0.033	0.04	0.132	0.431	-0.107	0.006	-0.007	-0.029	-0.569	0.041
SUC2	-0.043	0.015	0.674	0.931	0.076	0.037	0.004	0.17	0.017	-0.802

**Table C.10: The optimal perturbation matrix corresponding to alternative interaction matrix 3.** This perturbation matrix minimizes the weighted residual sum of squares of the responses and perturbations while fulfilling the constraints on the unweighted residual sum of squares in (C.1.7) for the interaction matrix in Table C.9.

Experiment	1	2	3	4	5	6	7	8	9	10
	Perturbed gene									
<i>CAT8</i>	1.905	-0.055	0.238	0.258	0.039	-0.044	0.005	0.068	0.01	-0.02
<i>FBP1</i>	0.009	2.142	0.006	-0.015	-0.002	0.001	-0.007	-0.002	0.001	0
<i>HXK2</i>	0.003	0.009	2.679	-0.053	-0.007	0.007	-0.001	-0.01	-0.002	0.003
<i>MED8</i>	-0.002	0.018	-0.08	1.355	-0.012	0.014	-0.001	-0.023	-0.003	0.006
<i>MIG1</i>	0.009	-0.001	-0.001	-0.034	2.155	0.001	-0.006	0	0	0
<i>SIP2</i>	0.006	0.001	0.004	-0.042	-0.003	2.139	0.004	-0.004	-0.001	0.001
<i>SIP4</i>	0.001	0.001	0.002	0.002	0	-0.001	2.145	0.001	0	0
<i>SNF1</i>	0.007	-0.003	0.001	0.004	0.002	-0.001	-0.001	1.473	0.001	-0.001
<i>SNF4</i>	-0.005	0.001	0.005	-0.006	0.002	0.001	-0.001	0	3.181	0.001
<i>SUC2</i>	0.019	0	-0.001	-0.043	-0.002	0.001	0.005	-0.001	-0.001	2.453

Table C.11: Alternative interaction matrix 4 that explains the steady-state data recorded by Lorenz et al. (2009) for the Snf1 network and lacks the  $G_{22}$  interaction.

Gene $i \setminus j$	CAT8	FBP1	HXK2	MED8	MIG1	SIP2	SIP4	SNF1	SNF4	SUC2
CAT8	-0.469	-0.041	0.14	0.098	0.123	0.006	0.007	0.48	-0.008	0.019
FBP1	-0.547	0	-0.051	-9.46	-0.623	-0.633	-0.211	-0.634	-0.268	-0.277
HXK2	0.988	-0.75	-3.824	15.86	1.353	0.792	0.342	0.5	0.421	0.449
MED8	0.791	-0.478	2.985	13.18	0.791	1.07	0.321	3.079	0.415	0.352
MIG1	0.096	-0.049	0.254	2.564	-0.735	0.067	0.044	0.231	0.089	0.067
SIP2	-0.068	0.027	0.692	-0.634	0.014	-0.839	-0.02	0.493	-0.023	-0.013
SIP4	0	0.009	-0.002	-0.051	0.002	-0.012	-0.086	-0.005	-0.001	0.001
SNF1	0.094	-0.193	0.788	3.412	0.113	0.125	0.041	-1.319	0.056	0.129
SNF4	0.033	0.04	0.132	0.431	-0.107	0.006	-0.007	-0.029	-0.569	0.041
SUC2	-0.043	0.015	0.674	0.931	0.076	0.037	0.004	0.17	0.017	-0.802

**Table C.12: The optimal perturbation matrix corresponding to alternative interaction matrix 4.** This perturbation matrix minimizes the weighted residual sum of squares of the responses and perturbations while fulfilling the constraints on the unweighted residual sum of squares in (C.1.7) for the interaction matrix in Table C.11.

Experiment	1	2	3	4	5	6	7	8	9	10
Perturbed gene										
<i>CAT8</i>	2.142	0.028	0.019	-0.02	0.002	-0.002	0.005	0.036	-0.001	0
<i>FBP1</i>	-0.045	1.47	-0.234	-0.29	-0.051	0.03	0.016	-0.228	0.023	0.013
<i>HXK2</i>	-0.012	-0.013	2.294	-0.091	-0.014	0.009	0.005	-0.072	0.006	0.004
<i>MED8</i>	-0.013	-0.017	-0.079	1.351	-0.015	0.009	0.005	-0.057	0.007	0.004
<i>MIG1</i>	-0.004	-0.012	-0.025	-0.057	2.128	0.001	-0.005	-0.033	0.001	0
<i>SIP2</i>	0.002	0.008	0.029	-0.017	0.002	2.132	0.002	0.03	-0.001	0
<i>SIP4</i>	0.002	-0.005	-0.003	0.001	0	-0.001	2.139	0	0	0
<i>SNF1</i>	-0.008	0	-0.055	-0.086	-0.014	0.006	0.003	1.166	0.006	0.003
<i>SNF4</i>	-0.002	-0.02	-0.01	-0.011	-0.002	0.001	-0.001	-0.014	3.182	0.001
<i>SUC2</i>	-0.002	-0.013	0.006	-0.048	-0.001	0.001	0.005	-0.022	0.001	2.456

Table C.13: Alternative interaction matrix 5 that explains the steady-state data recorded by Lorenz et al. (2009) for the Snf1 network and lacks the  $a_{31}$  interaction.

Gene $i \setminus j$	CAT8	FBP1	HXK2	MED8	MIG1	SIP2	SIP4	SNF1	SNF4	SUC2
CAT8	-0.469	-0.041	0.14	0.098	0.123	0.006	0.007	0.48	-0.008	0.019
FBP1	-0.117	-0.767	0.51	-1.413	-0.118	-0.131	-0.037	0.197	-0.035	-0.042
HXK2	0	-0.269	-4.305	2.102	0.515	-0.17	0.03	-1.07	0.024	0.053
MED8	0.791	-0.478	2.985	13.18	0.791	1.07	0.321	3.079	0.415	0.352
MIG1	0.096	-0.049	0.254	2.564	-0.735	0.067	0.044	0.231	0.089	0.067
SIP2	-0.068	0.027	0.692	-0.634	0.014	-0.839	-0.02	0.493	-0.023	-0.013
SIP4	0	0.009	-0.002	-0.051	0.002	-0.012	-0.086	-0.005	-0.001	0.001
SNF1	0.094	-0.193	0.788	3.412	0.113	0.125	0.041	-1.319	0.056	0.129
SNF4	0.033	0.04	0.132	0.431	-0.107	0.006	-0.007	-0.029	-0.569	0.041
SUC2	-0.043	0.015	0.674	0.931	0.076	0.037	0.004	0.17	0.017	-0.802

**Table C.14: The optimal perturbation matrix corresponding to alternative interaction matrix 5.** This perturbation matrix minimizes the weighted residual sum of squares of the responses and perturbations while fulfilling the constraints on the unweighted residual sum of squares in (C.1.7) for the interaction matrix in Table C.13.

Experiment	1	2	3	4	5	6	7	8	9	10
	Perturbed gene									
<i>CAT8</i>	2.021	0	0.002	-0.001	0.003	0.001	0.005	-0.003	0	0.001
<i>FBP1</i>	0.07	2.137	-0.02	-0.07	-0.015	0.013	-0.007	-0.015	-0.001	0.004
<i>HXK2</i>	0.06	0.005	2.831	-0.062	-0.006	0.011	0.001	-0.02	-0.001	0.004
<i>MED8</i>	0.016	0	-0.003	2.102	0	0.002	0	-0.002	0	0.001
<i>MIG1</i>	-0.034	-0.001	0.008	0.017	2.123	-0.006	-0.006	0.008	0	-0.002
<i>SIP2</i>	0.071	0.008	-0.014	-0.084	-0.012	2.142	0.005	-0.014	-0.003	0.005
<i>SIP4</i>	0.004	0.001	-0.001	-0.002	0	0	2.145	0	0	0
<i>SNF1</i>	0.001	-0.001	0.006	0.001	-0.002	-0.002	-0.001	1.479	0	0
<i>SNF4</i>	0.009	0.002	-0.008	-0.01	-0.004	0.001	-0.001	-0.001	3.185	0.002
<i>SUC2</i>	0.031	0.004	-0.03	-0.051	-0.005	0.006	0.006	-0.008	-0.001	2.46

Table C.15: Alternative interaction matrix 6 that explains the steady-state data recorded by Lorenz et al. (2009) for the Snf1 network and lacks the  $a_{33}$  interaction.

Gene $i \setminus j$	CAT8	FBP1	HXK2	MED8	MIG1	SIP2	SIP4	SNF1	SNF4	SUC2
CAT8	-0.469	-0.041	0.14	0.098	0.123	0.006	0.007	0.48	-0.008	0.019
FBP1	-0.117	-0.767	0.51	-1.413	-0.118	-0.131	-0.037	0.197	-0.035	-0.042
HXK2	1.126	-0.93	0	19.65	1.36	1.096	0.431	1.925	0.542	0.466
MED8	0.791	-0.478	2.985	13.18	0.791	1.07	0.321	3.079	0.415	0.352
MIG1	0.096	-0.049	0.254	2.564	-0.735	0.067	0.044	0.231	0.089	0.067
SIP2	-0.068	0.027	0.692	-0.634	0.014	-0.839	-0.02	0.493	-0.023	-0.013
SIP4	0	0.009	-0.002	-0.051	0.002	-0.012	-0.086	-0.005	-0.001	0.001
SNF1	0.094	-0.193	0.788	3.413	0.113	0.125	0.041	-1.318	0.056	0.129
SNF4	0.033	0.04	0.132	0.431	-0.107	0.006	-0.007	-0.029	-0.569	0.041
SUC2	-0.043	0.015	0.674	0.931	0.076	0.037	0.004	0.17	0.017	-0.802

**Table C.16: The optimal perturbation matrix corresponding to alternative interaction matrix 6.** This perturbation matrix minimizes the weighted residual sum of squares of the responses and perturbations while fulfilling the constraints on the unweighted residual sum of squares in (C.1.7) for the interaction matrix in Table C.15.

Experiment	1	2	3	4	5	6	7	8	9	10
	Perturbed gene									
<i>CAT8</i>	2.137	0	-0.002	-0.017	-0.001	0.004	0.005	0.019	-0.001	0.003
<i>FBP1</i>	0.008	2.136	-0.022	0.055	0.012	0.034	-0.008	0.034	0.003	0.022
<i>HXK2</i>	0.006	0.006	2.776	0.044	0.005	0.031	0	0.038	0.002	0.02
<i>MED8</i>	-0.006	-0.007	0.027	1.645	-0.005	-0.032	0	-0.04	-0.002	-0.021
<i>MIG1</i>	-0.004	-0.002	0.013	-0.049	2.146	-0.022	-0.006	-0.028	-0.001	-0.009
<i>SIP2</i>	0.007	0.01	-0.014	0.05	0.009	2.143	0.004	0.031	0.003	0.022
<i>SIP4</i>	0.002	0.002	-0.001	0.003	0	0	2.145	0.001	0	0
<i>SNF1</i>	-0.001	-0.005	0.018	-0.033	0	-0.02	-0.001	1.301	0	-0.009
<i>SNF4</i>	-0.001	0.002	-0.008	0.005	0.002	0.001	-0.001	-0.002	3.183	0.003
<i>SUC2</i>	0.002	0.005	-0.03	0.014	0.003	0.012	0.005	0.007	0.001	2.462

Table C.17: Alternative interaction matrix 7 that explains the steady-state data recorded by Lorenz et al. (2009) for the Snf1 network and lacks the  $g_{34}$  interaction.

Gene $i \setminus j$	CAT8	FBP1	HXK2	MED8	MIG1	SIP2	SIP4	SNF1	SNF4	SUC2
CAT8	-0.469	-0.041	0.14	0.098	0.123	0.006	0.007	0.48	-0.008	0.019
FBP1	-0.117	-0.767	0.51	-1.413	-0.118	-0.131	-0.037	0.197	-0.035	-0.042
HXK2	0.074	-0.152	-4.703	0	0.405	-0.271	-0.011	-1.199	-0.027	0.031
MED8	0.791	-0.478	2.985	13.18	0.791	1.07	0.321	3.079	0.415	0.352
MIG1	0.096	-0.049	0.254	2.564	-0.735	0.067	0.044	0.231	0.089	0.067
SIP2	-0.068	0.027	0.691	-0.634	0.014	-0.839	-0.02	0.493	-0.023	-0.013
SIP4	0	0.009	-0.002	-0.051	0.002	-0.012	-0.086	-0.005	-0.001	0.001
SNF1	0.094	-0.193	0.788	3.412	0.113	0.125	0.041	-1.319	0.056	0.129
SNF4	0.033	0.04	0.132	0.431	-0.107	0.006	-0.007	-0.029	-0.569	0.041
SUC2	-0.043	0.015	0.674	0.931	0.076	0.037	0.004	0.17	0.017	-0.802

**Table C.18: The optimal perturbation matrix corresponding to alternative interaction matrix 7.** This perturbation matrix minimizes the weighted residual sum of squares of the responses and perturbations while fulfilling the constraints on the unweighted residual sum of squares in (C.1.7) for the interaction matrix in Table C.17.

Experiment	1	2	3	4	5	6	7	8	9	10
<i>Perturbed gene</i>										
<i>CAT8</i>	2.138	0	-0.002	-0.01	0.003	0.001	0.006	-0.009	0	0
<i>FBP1</i>	0	2.139	-0.023	-0.078	-0.028	0.008	-0.006	-0.037	-0.001	-0.008
<i>HXK2</i>	-0.001	0.004	2.792	-0.072	-0.01	0.007	0.001	-0.053	-0.001	-0.007
<i>MED8</i>	0	0	-0.005	2.02	-0.002	0.002	0.001	-0.01	0	-0.002
<i>MIG1</i>	0	-0.001	0.008	0.013	2.108	-0.004	-0.006	0.015	0	0.001
<i>SIP2</i>	-0.001	0.006	-0.016	-0.096	-0.021	2.141	0.005	-0.032	-0.002	-0.007
<i>SIP4</i>	0.002	0.001	-0.001	-0.001	0	0	2.144	0	0	0
<i>SNF1</i>	0.001	0	0.003	-0.012	-0.006	0	0	1.481	0	-0.001
<i>SNF4</i>	-0.001	0.001	-0.009	-0.012	-0.007	0.001	-0.001	-0.004	3.185	0
<i>SUC2</i>	-0.002	0.004	-0.033	-0.063	-0.009	0.004	0.006	-0.024	-0.001	2.448

Table C.19: Alternative interaction matrix 8 that explains the steady-state data recorded by Lorenz et al. (2009) for the Snf1 network and lacks the  $g_{37}$  interaction.

Gene $i \setminus j$	CAT8	FBP1	HXK2	MED8	MIG1	SIP2	SIP4	SNF1	SNF4	SUC2
CAT8	-0.469	-0.041	0.14	0.098	0.123	0.006	0.007	0.48	-0.008	0.019
FBP1	-0.117	-0.767	0.51	-1.413	-0.118	-0.131	-0.037	0.197	-0.035	-0.042
HXK2	0.145	-0.226	-4.666	1.494	0.459	-0.215	0	-1.397	-0.002	0.056
MED8	0.791	-0.478	2.985	13.18	0.791	1.07	0.321	3.079	0.415	0.352
MIG1	0.096	-0.049	0.254	2.564	-0.735	0.067	0.044	0.231	0.089	0.067
SIP2	-0.068	0.027	0.691	-0.634	0.014	-0.839	-0.02	0.493	-0.023	-0.013
SIP4	0	0.009	-0.002	-0.051	0.002	-0.012	-0.086	-0.005	-0.001	0.001
SNF1	0.094	-0.193	0.788	3.412	0.113	0.125	0.041	-1.319	0.056	0.129
SNF4	0.033	0.04	0.132	0.431	-0.107	0.006	-0.007	-0.029	-0.569	0.041
SUC2	-0.043	0.015	0.674	0.931	0.076	0.037	0.004	0.17	0.017	-0.802

**Table C.20: The optimal perturbation matrix corresponding to alternative interaction matrix 8.** This perturbation matrix minimizes the weighted residual sum of squares of the responses and perturbations while fulfilling the constraints on the unweighted residual sum of squares in (C.1.7) for the interaction matrix in Table C.19.

Experiment	1	2	3	4	5	6	7	8	9	10
	Perturbed gene									
<i>CAT8</i>	2.137	0	-0.002	-0.01	0.002	0.001	0.008	-0.007	-0.001	0.001
<i>FBP1</i>	-0.003	2.139	-0.019	-0.075	-0.019	0.01	0.033	-0.018	0.002	-0.002
<i>HXK2</i>	-0.003	0.004	2.868	-0.064	-0.007	0.008	0.018	-0.022	0.001	-0.001
<i>MED8</i>	0	0	-0.003	2.062	-0.001	0.002	0.007	-0.004	0	0
<i>MIG1</i>	0	-0.001	0.007	0.017	2.12	-0.005	-0.025	0.009	-0.001	0
<i>SIP2</i>	-0.003	0.007	-0.014	-0.091	-0.014	2.142	0.033	-0.016	0.001	-0.001
<i>SIP4</i>	0.002	0.001	-0.001	-0.001	0	0	2.122	0	0	0
<i>SNF1</i>	0.001	-0.001	0.006	0.001	-0.003	-0.002	0.006	1.485	0	0
<i>SNF4</i>	-0.002	0.001	-0.007	-0.01	-0.005	0.001	-0.001	-0.001	3.184	0.001
<i>SUC2</i>	-0.003	0.004	-0.028	-0.057	-0.006	0.005	0.024	-0.009	0.001	2.455

Table C.21: Alternative interaction matrix 9 that explains the steady-state data recorded by Lorenz et al. (2009) for the Snf1 network and lacks the  $a_{41}$  interaction.

Gene $i \setminus j$	CAT8	FBP1	HXK2	MED8	MIG1	SIP2	SIP4	SNF1	SNF4	SUC2
CAT8	-0.469	-0.041	0.14	0.098	0.123	0.006	0.007	0.48	-0.008	0.019
FBP1	-0.117	-0.767	0.51	-1.413	-0.118	-0.131	-0.037	0.197	-0.035	-0.042
HXK2	0.988	-0.75	-3.824	15.86	1.353	0.792	0.342	0.501	0.421	0.449
MED8	0	-0.093	2.6	2.177	0.12	0.3	0.072	1.822	0.097	0.035
MIG1	0.096	-0.049	0.254	2.564	-0.735	0.067	0.044	0.231	0.089	0.067
SIP2	-0.068	0.027	0.692	-0.634	0.014	-0.839	-0.02	0.493	-0.023	-0.013
SIP4	0	0.009	-0.002	-0.051	0.002	-0.012	-0.086	-0.005	-0.001	0.001
SNF1	0.094	-0.193	0.788	3.412	0.113	0.125	0.041	-1.319	0.056	0.129
SNF4	0.033	0.04	0.132	0.431	-0.107	0.006	-0.007	-0.029	-0.569	0.041
SUC2	-0.043	0.015	0.674	0.931	0.076	0.037	0.004	0.17	0.017	-0.802

**Table C.22: The optimal perturbation matrix corresponding to alternative interaction matrix 9.** This perturbation matrix minimizes the weighted residual sum of squares of the responses and perturbations while fulfilling the constraints on the unweighted residual sum of squares in (C.1.7) for the interaction matrix in Table C.21.

Experiment	1	2	3	4	5	6	7	8	9	10
Perturbed gene										
<i>CAT8</i>	2.235	-0.012	0.048	0.039	0.007	-0.009	0.006	0.02	0.001	-0.004
<i>FBP1</i>	-0.023	2.131	0.016	0.013	0.005	-0.003	-0.007	0.008	0.002	-0.003
<i>HXK2</i>	0.004	0	3.08	-0.001	0	0.001	0	0.002	0	0
<i>MED8</i>	0.105	0.011	-0.078	1.514	-0.009	0.017	-0.001	-0.019	-0.002	0.007
<i>MIG1</i>	-0.072	-0.007	0.041	0.035	2.152	-0.011	-0.004	0.016	0.001	-0.005
<i>SIP2</i>	-0.068	-0.008	0.008	0.034	0.013	2.143	0.005	0.014	0.002	-0.005
<i>SIP4</i>	0.003	0.002	-0.004	-0.002	0	0	2.141	0	0	0
<i>SNF1</i>	-0.032	-0.002	0.032	0.007	0.005	-0.003	0	1.394	0.001	-0.002
<i>SNF4</i>	-0.018	-0.001	0.007	0.003	0.002	-0.001	-0.001	0	3.182	0
<i>SUC2</i>	-0.074	-0.008	0.079	0.038	0.008	-0.01	0.006	0.012	0.002	2.448

Table C.23: Alternative interaction matrix 10 that explains the steady-state data recorded by Lorenz et al. (2009) for the Snf1 network and lacks the  $a_{44}$  interaction.

Gene $i \setminus j$	CAT8	FBP1	HXK2	MED8	MIG1	SIP2	SIP4	SNF1	SNF4	SUC2
CAT8	-0.469	-0.041	0.14	0.098	0.123	0.006	0.007	0.48	-0.008	0.019
FBP1	-0.117	-0.767	0.51	-1.413	-0.118	-0.131	-0.037	0.197	-0.035	-0.042
HXK2	0.988	-0.75	-3.824	15.85	1.353	0.792	0.342	0.501	0.421	0.449
MED8	0.031	0.019	2.254	0	0.003	0.185	0.028	1.666	0.043	0.005
MIG1	0.096	-0.049	0.254	2.564	-0.735	0.067	0.044	0.231	0.089	0.067
SIP2	-0.068	0.027	0.692	-0.634	0.014	-0.839	-0.02	0.493	-0.023	-0.013
SIP4	0	0.009	-0.002	-0.051	0.002	-0.012	-0.086	-0.005	-0.001	0.001
SNF1	0.094	-0.193	0.788	3.412	0.113	0.125	0.041	-1.319	0.056	0.129
SNF4	0.033	0.04	0.132	0.431	-0.107	0.006	-0.007	-0.029	-0.569	0.041
SUC2	-0.043	0.015	0.674	0.931	0.076	0.037	0.004	0.17	0.017	-0.802

**Table C.24: The optimal perturbation matrix corresponding to alternative interaction matrix 10.** This perturbation matrix minimizes the weighted residual sum of squares of the responses and perturbations while fulfilling the constraints on the unweighted residual sum of squares in (C.1.7) for the interaction matrix in Table C.23.

Experiment	1	2	3	4	5	6	7	8	9	10
Perturbed gene										
<i>CAT8</i>	2.134	-0.01	0.047	0.039	0.013	-0.006	0.005	0.053	0.001	0.007
<i>FBP1</i>	0.001	2.133	0.021	0.014	0.012	-0.002	-0.008	0.021	0.002	0.003
<i>HXK2</i>	0	0.001	2.94	-0.005	-0.001	0.002	0	0	0	-0.001
<i>MED8</i>	-0.002	0.013	-0.115	1.325	-0.023	0.015	0	-0.058	-0.003	-0.014
<i>MIG1</i>	0	-0.007	0.043	0.036	2.161	-0.007	-0.006	0.037	0.001	0.007
<i>SIP2</i>	0.001	-0.008	0.01	0.037	0.028	2.143	0.004	0.04	0.002	0.008
<i>SIP4</i>	0.002	0.002	-0.005	-0.002	-0.001	0	2.145	-0.001	0	0
<i>SNF1</i>	0.001	0	0.02	-0.002	0.007	0	-0.001	1.263	0	0.001
<i>SNF4</i>	-0.001	0	0.008	0.003	0.004	0	-0.001	-0.001	3.181	0.002
<i>SUC2</i>	0	-0.007	0.094	0.039	0.016	-0.007	0.005	0.028	0.002	2.466

Table C.25: Alternative interaction matrix 11 that explains the steady-state data recorded by Lorenz et al. (2009) for the Snf1 network and lacks the  $a_{4T}$  interaction.

Gene $i \setminus j$	CAT8	FBP1	HXK2	MED8	MIG1	SIP2	SIP4	SNF1	SNF4	SUC2
CAT8	-0.469	-0.041	0.14	0.098	0.123	0.006	0.007	0.48	-0.008	0.019
FBP1	-0.117	-0.767	0.51	-1.413	-0.118	-0.131	-0.037	0.197	-0.035	-0.042
HXK2	0.988	-0.75	-3.824	15.85	1.353	0.792	0.342	0.5	0.421	0.449
MED8	-0.001	0.014	2.194	-0.304	-0.049	0.124	0	1.297	0.019	-0.016
MIG1	0.096	-0.049	0.254	2.564	-0.735	0.067	0.044	0.231	0.089	0.067
SIP2	-0.068	0.027	0.692	-0.634	0.014	-0.839	-0.02	0.493	-0.023	-0.013
SIP4	0	0.009	-0.002	-0.051	0.002	-0.012	-0.086	-0.005	-0.001	0.001
SNF1	0.094	-0.193	0.788	3.412	0.113	0.125	0.041	-1.319	0.056	0.129
SNF4	0.033	0.04	0.132	0.431	-0.107	0.006	-0.007	-0.029	-0.569	0.041
SUC2	-0.043	0.015	0.674	0.931	0.076	0.037	0.004	0.17	0.017	-0.802

**Table C.26: The optimal perturbation matrix corresponding to alternative interaction matrix 11.** This perturbation matrix minimizes the weighted residual sum of squares of the responses and perturbations while fulfilling the constraints on the unweighted residual sum of squares in (C.1.7) for the interaction matrix in Table C.25.

Experiment	1	2	3	4	5	6	7	8	9	10
Perturbed gene										
<i>CAT8</i>	2.121	-0.014	0.049	0.04	0.011	-0.009	-0.03	0.032	-0.002	0.002
<i>FBP1</i>	0.002	2.127	0.026	0.015	0.01	-0.003	-0.021	0.014	0	0
<i>HXK2</i>	-0.001	0.003	2.847	-0.009	-0.002	0.003	0.012	-0.001	0	0
<i>MED8</i>	-0.007	0.02	-0.143	1.153	-0.023	0.026	0.066	-0.045	0.003	-0.004
<i>MIG1</i>	0.003	-0.009	0.047	0.038	2.16	-0.011	-0.059	0.025	-0.001	0.002
<i>SIP2</i>	0.004	-0.011	0.014	0.039	0.024	2.144	-0.03	0.025	-0.002	0.002
<i>SIP4</i>	0.002	0.002	-0.004	-0.002	0	0	2.274	0	0	0
<i>SNF1</i>	0.002	-0.002	0.031	0.004	0.008	-0.002	-0.019	1.346	0	0.001
<i>SNF4</i>	0	-0.001	0.011	0.004	0.003	-0.001	0.005	0	3.186	0.001
<i>SUC2</i>	0.002	-0.01	0.104	0.042	0.015	-0.011	-0.038	0.019	-0.002	2.459

Table C.27: Alternative interaction matrix 12 that explains the steady-state data recorded by Lorenz et al. (2009) for the Snf1 network and lacks the  $q_{55}$  interaction.

Gene $i \setminus j$	CAT8	FBP1	HXK2	MED8	MIG1	SIP2	SIP4	SNF1	SNF4	SUC2
CAT8	-0.469	-0.041	0.14	0.098	0.123	0.006	0.007	0.48	-0.008	0.019
FBP1	-0.117	-0.767	0.51	-1.413	-0.118	-0.131	-0.037	0.197	-0.035	-0.042
HXK2	0.988	-0.75	-3.824	15.86	1.353	0.792	0.342	0.501	0.421	0.449
MED8	0.791	-0.478	2.985	13.18	0.791	1.07	0.321	3.079	0.415	0.352
MIG1	0.5	-0.321	0.266	9.442	0	0.537	0.203	1.052	0.29	0.243
SIP2	-0.068	0.027	0.691	-0.634	0.014	-0.839	-0.02	0.493	-0.023	-0.013
SIP4	0	0.009	-0.002	-0.051	0.002	-0.012	-0.086	-0.005	-0.001	0.001
SNF1	0.094	-0.193	0.788	3.412	0.113	0.125	0.041	-1.318	0.056	0.129
SNF4	0.033	0.04	0.132	0.431	-0.107	0.006	-0.007	-0.029	-0.569	0.041
SUC2	-0.043	0.015	0.674	0.931	0.076	0.037	0.004	0.17	0.017	-0.802

**Table C.28: The optimal perturbation matrix corresponding to alternative interaction matrix 12.** This perturbation matrix minimizes the weighted residual sum of squares of the responses and perturbations while fulfilling the constraints on the unweighted residual sum of squares in (C.1.7) for the interaction matrix in Table C.27.

Experiment	1	2	3	4	5	6	7	8	9	10
Perturbed gene										
<i>CAT8</i>	2.1	-0.004	0.014	-0.016	-0.016	-0.001	0.006	0.005	-0.002	0.002
<i>FBP1</i>	0.001	2.116	0.027	0.003	0.001	0.001	-0.008	0.005	0	0
<i>HXK2</i>	-0.016	0.007	2.469	-0.058	0.003	-0.002	0	-0.021	0.004	-0.008
<i>MED8</i>	-0.022	0.012	-0.085	1.457	0.011	-0.003	0	-0.027	0.007	-0.012
<i>MIG1</i>	0.065	-0.031	0.218	0.191	1.809	0.008	0	0.078	-0.017	0.031
<i>SIP2</i>	-0.001	0	-0.001	-0.029	-0.015	2.136	0.004	0	-0.001	0
<i>SIP4</i>	0.002	0.001	0.002	0.002	-0.002	-0.001	2.143	0.001	0	0
<i>SNF1</i>	-0.01	0.004	-0.044	-0.045	0.001	-0.001	-0.001	1.392	0.003	-0.005
<i>SNF4</i>	-0.002	0.004	-0.019	-0.016	0.027	0	-0.001	-0.004	3.178	-0.001
<i>SUC2</i>	-0.008	0	0.018	-0.035	-0.005	0	0.005	-0.005	0	2.451

Table C.29: Alternative interaction matrix 13 that explains the steady-state data recorded by Lorenz et al. (2009) for the Snf1 network and lacks the  $a_{66}$  interaction.

Gene $i \setminus j$	CAT8	FBP1	HXK2	MED8	MIG1	SIP2	SIP4	SNF1	SNF4	SUC2
CAT8	-0.469	-0.041	0.14	0.098	0.123	0.006	0.007	0.48	-0.008	0.019
FBP1	-0.117	-0.767	0.51	-1.413	-0.118	-0.131	-0.037	0.197	-0.035	-0.042
HXK2	0.988	-0.75	-3.824	15.86	1.353	0.792	0.342	0.501	0.421	0.449
MED8	0.791	-0.478	2.985	13.18	0.791	1.07	0.321	3.079	0.415	0.352
MIG1	0.096	-0.049	0.254	2.564	-0.735	0.067	0.044	0.231	0.089	0.067
SIP2	0.43	-0.264	1.241	7.652	0.519	0	0.173	1.526	0.222	0.206
SIP4	0	0.009	-0.002	-0.051	0.002	-0.012	-0.086	-0.005	-0.001	0.001
SNF1	0.094	-0.193	0.788	3.412	0.113	0.125	0.041	-1.318	0.056	0.129
SNF4	0.033	0.04	0.132	0.431	-0.107	0.006	-0.007	-0.029	-0.569	0.041
SUC2	-0.043	0.015	0.674	0.931	0.076	0.037	0.004	0.17	0.017	-0.802

**Table C.30: The optimal perturbation matrix corresponding to alternative interaction matrix 13.** This perturbation matrix minimizes the weighted residual sum of squares of the responses and perturbations while fulfilling the constraints on the unweighted residual sum of squares in (C.1.7) for the interaction matrix in Table C.29.

Experiment	1	2	3	4	5	6	7	8	9	10
Perturbed gene										
<i>CAT8</i>	2.137	0.004	-0.01	-0.065	-0.002	0.004	0.006	-0.006	-0.001	0
<i>FBP1</i>	0.001	2.126	-0.003	-0.02	-0.001	0.009	-0.007	-0.003	0.001	-0.001
<i>HXK2</i>	0	0.003	3.042	-0.031	-0.003	0.006	0	-0.007	0	-0.001
<i>MED8</i>	-0.002	0.022	-0.053	1.076	-0.016	0.016	0.004	-0.038	0	-0.008
<i>MIG1</i>	-0.001	0.002	-0.014	-0.063	2.147	0.012	-0.004	-0.007	0	-0.001
<i>SIP2</i>	0.004	-0.047	0.124	0.244	0.034	1.532	-0.009	0.084	0.001	0.017
<i>SIP4</i>	0.002	0.002	-0.004	0	0	0.001	2.147	0	0	0
<i>SNF1</i>	0.001	0.002	-0.014	-0.024	-0.001	0.006	0	1.453	0	-0.001
<i>SNF4</i>	-0.001	0.001	-0.003	-0.01	0	0.003	-0.001	-0.002	3.184	0.001
<i>SUC2</i>	-0.001	0.002	-0.012	-0.065	-0.002	0.007	0.006	-0.004	0	2.455

Table C.31: Alternative interaction matrix 14 that explains the steady-state data recorded by Lorenz et al. (2009) for the Snf1 network and lacks the  $q_{TT}$  interaction.

Gene $i \setminus j$	CAT8	FBP1	HXK2	MED8	MIG1	SIP2	SIP4	SNF1	SNF4	SUC2
CAT8	-0.469	-0.041	0.14	0.098	0.123	0.006	0.007	0.48	-0.008	0.019
FBP1	-0.117	-0.767	0.51	-1.413	-0.118	-0.131	-0.037	0.197	-0.035	-0.042
HXK2	0.988	-0.75	-3.824	15.86	1.353	0.792	0.342	0.501	0.421	0.449
MED8	0.791	-0.478	2.985	13.18	0.791	1.07	0.321	3.079	0.415	0.352
MIG1	0.096	-0.049	0.254	2.564	-0.735	0.067	0.044	0.231	0.089	0.067
SIP2	-0.068	0.027	0.692	-0.634	0.014	-0.839	-0.02	0.493	-0.023	-0.013
SIP4	0.212	-0.122	0.209	3.55	0.226	0.24	0	0.471	0.105	0.099
SNF1	0.094	-0.193	0.788	3.413	0.113	0.125	0.041	-1.318	0.056	0.129
SNF4	0.033	0.04	0.132	0.431	-0.107	0.006	-0.007	-0.029	-0.569	0.041
SUC2	-0.043	0.015	0.674	0.931	0.076	0.037	0.004	0.17	0.017	-0.802

**Table C.32: The optimal perturbation matrix corresponding to alternative interaction matrix 14.** This perturbation matrix minimizes the weighted residual sum of squares of the responses and perturbations while fulfilling the constraints on the unweighted residual sum of squares in (C.1.7) for the interaction matrix in Table C.31.

Experiment	1	2	3	4	5	6	7	8	9	10
	Perturbed gene									
<i>CAT8</i>	2.138	0	0.001	-0.017	0.001	-0.001	0.006	0.001	-0.001	0.001
<i>FBP1</i>	0.001	2.129	0.009	0	0	0.001	-0.006	0.001	0.001	-0.001
<i>HXK2</i>	0	0.003	2.952	-0.028	-0.004	0.002	0.001	-0.005	0	0
<i>MED8</i>	-0.001	0.006	-0.036	1.751	-0.007	0.003	-0.001	-0.009	0.001	-0.001
<i>MIG1</i>	-0.001	0.001	-0.006	-0.025	2.133	0	-0.004	-0.002	0	0
<i>SIP2</i>	0	0	0.003	-0.017	0	2.138	0.005	0.001	0	0
<i>SIP4</i>	0.007	-0.037	0.219	0.275	0.045	-0.02	2.064	0.061	-0.005	0.005
<i>SNF1</i>	0.001	0.001	-0.011	-0.018	-0.002	0.001	0.001	1.438	0	0
<i>SNF4</i>	-0.001	0	0.001	-0.003	-0.001	0.001	0.001	-0.001	3.184	0.001
<i>SUC2</i>	-0.001	0	0.004	-0.021	-0.001	0	0.007	-0.001	0	2.456

Table C.33: Alternative interaction matrix 15 that explains the steady-state data recorded by Lorenz et al. (2009) for the Snf1 network and lacks the  $q_{84}$  interaction.

Gene $i \setminus j$	CAT8	FBP1	HXK2	MED8	MIG1	SIP2	SIP4	SNF1	SNF4	SUC2
CAT8	-0.469	-0.041	0.14	0.098	0.123	0.006	0.007	0.48	-0.008	0.019
FBP1	-0.117	-0.767	0.51	-1.413	-0.118	-0.131	-0.037	0.197	-0.035	-0.042
HXK2	0.988	-0.75	-3.824	15.86	1.353	0.792	0.342	0.501	0.421	0.449
MED8	0.791	-0.478	2.985	13.18	0.791	1.07	0.321	3.079	0.415	0.352
MIG1	0.096	-0.049	0.254	2.564	-0.735	0.067	0.044	0.231	0.089	0.067
SIP2	-0.068	0.027	0.691	-0.634	0.014	-0.839	-0.02	0.493	-0.023	-0.013
SIP4	0	0.009	-0.002	-0.051	0.002	-0.012	-0.086	-0.005	-0.001	0.001
SNF1	-0.102	-0.065	0.599	0	-0.091	-0.104	-0.035	-1.684	-0.04	0.039
SNF4	0.033	0.04	0.132	0.431	-0.107	0.006	-0.007	-0.029	-0.569	0.041
SUC2	-0.043	0.015	0.674	0.931	0.076	0.037	0.004	0.17	0.017	-0.802

**Table C.34: The optimal perturbation matrix corresponding to alternative interaction matrix 15.** This perturbation matrix minimizes the weighted residual sum of squares of the responses and perturbations while fulfilling the constraints on the unweighted residual sum of squares in (C.1.7) for the interaction matrix in Table C.33.

Experiment	1	2	3	4	5	6	7	8	9	10
Perturbed gene										
<i>CAT8</i>	2.139	0.003	-0.014	-0.033	-0.004	0.001	0.006	-0.006	-0.001	-0.001
<i>FBP1</i>	0.001	2.132	0.004	-0.002	-0.002	0.001	-0.007	-0.001	0.001	-0.001
<i>HXK2</i>	0	0	3.154	-0.003	-0.001	0	0	0.001	0	0
<i>MED8</i>	0	0	-0.001	2.12	0.001	0	0	-0.003	0	0
<i>MIG1</i>	0	-0.001	0.008	0.011	2.134	-0.001	-0.006	0.003	0	0.001
<i>SIP2</i>	0	0	0.001	-0.014	-0.003	2.138	0.004	-0.003	0	-0.001
<i>SIP4</i>	0.002	0.001	0	0	0	-0.001	2.144	0	0	0
<i>SNF1</i>	-0.001	0.004	-0.058	-0.058	-0.008	0.007	0	1.378	-0.001	-0.005
<i>SNF4</i>	-0.001	-0.001	0.008	0.009	0.001	0	-0.001	0.002	3.185	0.002
<i>SUC2</i>	0	-0.001	0.013	0.013	0.002	-0.001	0.005	0.003	0	2.455

Table C.35: Alternative interaction matrix 16 that explains the steady-state data recorded by Lorenz et al. (2009) for the Snf1 network and lacks the  $a_{88}$  interaction.

Gene $i \setminus j$	CAT8	FBP1	HXK2	MED8	MIG1	SIP2	SIP4	SNF1	SNF4	SUC2
CAT8	-0.469	-0.041	0.14	0.098	0.123	0.006	0.007	0.48	-0.008	0.019
FBP1	-0.117	-0.767	0.51	-1.413	-0.118	-0.131	-0.037	0.197	-0.035	-0.042
HXK2	0.988	-0.75	-3.824	15.86	1.353	0.792	0.342	0.501	0.421	0.449
MED8	0.791	-0.478	2.985	13.18	0.791	1.07	0.321	3.079	0.415	0.352
MIG1	0.096	-0.049	0.254	2.564	-0.735	0.067	0.044	0.231	0.089	0.067
SIP2	-0.068	0.027	0.692	-0.634	0.014	-0.839	-0.02	0.493	-0.023	-0.013
SIP4	0	0.009	-0.002	-0.051	0.002	-0.012	-0.086	-0.005	-0.001	0.001
SNF1	0.295	-0.312	1.422	6.673	0.33	0.38	0.13	0	0.171	0.197
SNF4	0.033	0.04	0.132	0.431	-0.107	0.006	-0.007	-0.029	-0.569	0.041
SUC2	-0.043	0.015	0.674	0.931	0.076	0.037	0.004	0.17	0.017	-0.802

**Table C.36: The optimal perturbation matrix corresponding to alternative interaction matrix 16.** This perturbation matrix minimizes the weighted residual sum of squares of the responses and perturbations while fulfilling the constraints on the unweighted residual sum of squares in (C.1.7) for the interaction matrix in Table C.35.

Experiment	1	2	3	4	5	6	7	8	9	10
Perturbed gene										
<i>CAT8</i>	2.11	0.004	0.004	0.037	0.003	0.007	0.006	-0.016	-0.001	0.003
<i>FBP1</i>	0.001	2.129	-0.001	-0.003	0.002	0.002	-0.007	-0.003	0.001	0
<i>HXK2</i>	-0.001	-0.001	3.168	-0.014	0	-0.004	0	0.006	0	-0.001
<i>MED8</i>	-0.015	-0.003	-0.009	1.704	-0.004	-0.017	-0.001	0.003	0	-0.004
<i>MIG1</i>	-0.011	-0.001	-0.004	-0.05	2.139	-0.01	-0.006	0.012	0	-0.002
<i>SIP2</i>	0.002	0.001	-0.001	0.002	0.002	2.129	0.004	-0.009	0	0.001
<i>SIP4</i>	0.002	0.001	0	0.001	0	-0.001	2.144	0	0	0
<i>SNF1</i>	0.043	0.008	0.022	0.128	0.01	0.05	0.001	1.162	0	0.011
<i>SNF4</i>	-0.004	-0.001	-0.002	-0.021	-0.002	-0.004	-0.001	0.005	3.185	0
<i>SUC2</i>	-0.013	-0.001	-0.004	-0.052	-0.003	-0.007	0.004	0.01	0	2.456

Table C.37: Alternative interaction matrix 17 that explains the steady-state data recorded by Lorenz et al. (2009) for the Snf1 network and lacks the  $g_{99}$  interaction.

Gene $i \setminus j$	CAT8	FBP1	HXK2	MED8	MIG1	SIP2	SIP4	SNF1	SNF4	SUC2
CAT8	-0.469	-0.041	0.14	0.098	0.123	0.006	0.007	0.48	-0.008	0.019
FBP1	-0.117	-0.767	0.51	-1.413	-0.118	-0.131	-0.037	0.197	-0.035	-0.042
HXK2	0.988	-0.75	-3.824	15.86	1.353	0.792	0.342	0.501	0.421	0.449
MED8	0.791	-0.478	2.985	13.18	0.791	1.07	0.321	3.079	0.415	0.352
MIG1	0.096	-0.049	0.254	2.564	-0.735	0.067	0.044	0.231	0.089	0.067
SIP2	-0.068	0.027	0.692	-0.634	0.014	-0.839	-0.02	0.493	-0.023	-0.013
SIP4	0	0.009	-0.002	-0.051	0.002	-0.012	-0.086	-0.005	-0.001	0.001
SNF1	0.094	-0.193	0.788	3.412	0.113	0.125	0.041	-1.319	0.056	0.129
SNF4	0.768	-0.444	0.918	12.9	0.667	0.882	0.282	1.643	0	0.362
SUC2	-0.043	0.015	0.674	0.931	0.076	0.037	0.004	0.17	0.017	-0.802

**Table C.38: The optimal perturbation matrix corresponding to alternative interaction matrix 17.** This perturbation matrix minimizes the weighted residual sum of squares of the responses and perturbations while fulfilling the constraints on the unweighted residual sum of squares in (C.1.7) for the interaction matrix in Table C.37.

Experiment	1	2	3	4	5	6	7	8	9	10
Perturbed gene										
<i>CAT8</i>	2.134	-0.001	0.006	-0.057	0.014	-0.001	0.005	0.011	0.014	0.001
<i>FBP1</i>	0	2.104	0.041	0.001	0.004	-0.001	-0.008	0.009	-0.001	0
<i>HXK2</i>	0.004	0.011	2.213	-0.093	-0.032	0.012	0.001	-0.038	0.009	-0.007
<i>MED8</i>	0.009	0.025	-0.19	0.819	-0.077	0.025	0.002	-0.069	0.013	-0.015
<i>MIG1</i>	0.001	0.009	-0.062	-0.107	1.928	0.004	-0.005	-0.023	-0.002	-0.004
<i>SIP2</i>	0	-0.001	0.019	-0.053	0.009	2.136	0.004	0.007	0.011	0
<i>SIP4</i>	0.002	0.001	0.001	0.003	0.002	-0.001	2.145	0.001	-0.001	0
<i>SNF1</i>	0.003	0.009	-0.07	-0.08	-0.026	0.008	0	1.34	0.016	-0.005
<i>SNF4</i>	-0.015	-0.044	0.319	0.279	0.131	-0.044	-0.004	0.128	2.761	0.027
<i>SUC2</i>	-0.001	0.001	0.014	-0.072	0.001	0.002	0.005	-0.01	0.007	2.456

Table C.39: Alternative interaction matrix 18 that explains the steady-state data recorded by Lorenz et al. (2009) for the Snf1 network and lacks the  $a_{10,10}$  interaction.

Gene $i \setminus j$	CAT8	FBP1	HXK2	MED8	MIG1	SIP2	SIP4	SNF1	SNF4	SUC2
CAT8	-0.469	-0.041	0.14	0.098	0.123	0.006	0.007	0.48	-0.008	0.019
FBP1	-0.117	-0.767	0.51	-1.413	-0.118	-0.131	-0.037	0.197	-0.035	-0.042
HXK2	0.988	-0.75	-3.824	15.86	1.353	0.792	0.342	0.501	0.421	0.449
MED8	0.791	-0.478	2.985	13.18	0.791	1.07	0.321	3.079	0.415	0.352
MIG1	0.096	-0.049	0.254	2.564	-0.735	0.067	0.044	0.231	0.089	0.067
SIP2	-0.068	0.027	0.692	-0.634	0.014	-0.839	-0.02	0.493	-0.023	-0.013
SIP4	0	0.009	-0.002	-0.051	0.002	-0.012	-0.086	-0.005	-0.001	0.001
SNF1	0.094	-0.193	0.788	3.412	0.113	0.125	0.041	-1.319	0.056	0.129
SNF4	0.033	0.04	0.132	0.431	-0.107	0.006	-0.007	-0.029	-0.569	0.041
SUC2	0.532	-0.366	0.76	10.06	0.607	0.651	0.215	0.946	0.268	0

**Table C.40: The optimal perturbation matrix corresponding to alternative interaction matrix 18.** This perturbation matrix minimizes the weighted residual sum of squares of the responses and perturbations while fulfilling the constraints on the unweighted residual sum of squares in (C.1.7) for the interaction matrix in Table C.39.

Experiment	1	2	3	4	5	6	7	8	9	10
	Perturbed gene									
<i>CAT8</i>	2.126	0	-0.009	-0.031	-0.001	-0.001	0.005	0.021	-0.001	-0.011
<i>FBP1</i>	0.002	2.119	0.03	0.009	0.003	0	-0.007	0.016	0.003	-0.007
<i>HXK2</i>	-0.006	0.011	2.66	-0.081	-0.019	0.005	0	-0.059	-0.008	0.038
<i>MED8</i>	-0.013	0.024	-0.176	1.333	-0.041	0.009	0.001	-0.097	-0.018	0.083
<i>MIG1</i>	-0.004	0.004	-0.032	-0.059	2.114	0.001	-0.005	-0.035	-0.003	0.016
<i>SIP2</i>	0	0	0.004	-0.029	0.001	2.139	0.004	0.015	0	-0.004
<i>SIP4</i>	0.002	0.001	0.001	0.002	0.001	-0.001	2.146	0.001	0	-0.002
<i>SNF1</i>	-0.008	0.017	-0.124	-0.124	-0.03	0.006	0	1.167	-0.013	0.046
<i>SNF4</i>	-0.002	0.001	-0.005	-0.01	-0.003	0.001	-0.001	-0.012	3.177	-0.01
<i>SUC2</i>	0.032	-0.057	0.411	0.357	0.095	-0.023	-0.002	0.254	0.042	1.132

### C.3 Simulation of perturbation experiments on the Snf1 example

We here describe how the additional perturbation experiments used to demonstrate the SVD design in Section 7.6 were simulated *in silico*, *i.e.* on a standard computer. We used the alternative network model in Table C.4, which is depicted in Figure 7.1 (left), as the “true” network to generate data, since many different models can explain the data recorded by Lorenz *et al.* (2009), the correct network is unknown, and this network is sparse with only 25 links, interampatte with degree 127, and generates a stable linear system. It is in other words at least as good as any other model that we could have used. Actually, we also tested the SVD design on a few alternative models but the results were so similar that it is pointless to present them.

We simulated the *in silico* perturbation experiments in Matlab ([www.mathworks.com](http://www.mathworks.com)) using the data model

$$\mathbf{Y} = -\hat{\mathbf{A}}_1^{-1}(\mathbf{P} - \mathbf{F}) + \mathbf{E}, \quad (\text{C.3.1})$$

which essentially is the transpose of (7.3.1). Here  $\mathbf{P} \in \mathbb{R}^{n \times m}$  is the matrix of designed perturbations,  $\mathbf{Y} \in \mathbb{R}^{n \times m}$  the matrix of observed responses in the  $n = 10$  genes in  $m$  experiments. We generated each perturbation error in  $\mathbf{F}$  using `randn` in Matlab, which generates an independent and identical normally distributed sequence of numbers with zero mean and standard deviation one, and scaled it such that the standard deviation equals the one reported by Lorenz *et al.* for each perturbation (supplementary Table S4). Similarly `randn` was used to generate the response errors in  $\mathbf{E}$  by scaling such that the standard deviation equals the mean value obtained by Lorenz *et al.* for each gene (supplementary Table S3). For the ten experiments performed *in vivo* by Lorenz *et al.* we, however, used  $\mathbf{F} = \mathbf{P} - \check{\mathbf{P}}$  and  $\mathbf{E} = \mathbf{Y} - \check{\mathbf{Y}}$ , with  $\mathbf{P}$  and  $\mathbf{Y}$  being the perturbations and responses reported by Lorenz *et al.* (supplementary Table S2 and S4),  $\check{\mathbf{P}}$  the optimal perturbation matrix in Table C.5, and  $\check{\mathbf{Y}} = -\hat{\mathbf{A}}_1^{-1}\check{\mathbf{P}}$ . This ensures that the simulated data has the same error level in both the input and output as observed in real *in vivo* data.

---

## Abbreviations

---

<i>E. coli</i>	<i>Escherichia coli</i>
<i>S. cerevisiae</i>	<i>Saccharomyces cerevisiae</i>
AIC	Akaike Information Criterion
AIDS	Acquired ImmunoDeficiency Syndrome
BIC	Bayesian Information Criterion
BioGRID	Biological General Repository for Interaction Datasets
cDNA	complementary DNA
ChIP	Chromatin ImmunoPrecipitation
DNA	DeoxyriboNucleic Acid
DoE	Design of Experiments
DREAM	Dialogue for Reverse Engineering Assessments and Methods
EGF	Epidermal Growth Factor
GFP	Green Fluorescence Protein
GRN	Gene Regulatory Network
HIV	Human Immunodeficiency Virus
iHOP	information Hyperlinked Over Proteins
INTERAMPATTE	INTERactions enabling simultaneous AMPlification and AT-Tenuation of different signals
IRMA	<i>In vivo</i> Reverse-engineering and Modelling Assessment
LASSO	Least Absolute Shrinkage and Selection Operator
LTI	Linear Time Invariant
MAPK	Mitogen-Activated Protein Kinase

---

mRNA	messenger RNA
ODE	Ordinary Differential Equations
OED	Optimal Experimental Design
P4	Predictive, Personalized, Preventive, and Participatory medicine
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
qPCR	quantitative real-time Polymerase Chain Reaction
qRT-PCR	quantitative real-time Reverse Transcription Polymerase Chain Reaction
RNA	RiboNucleic Acid
RSS	Residual Sum of Squares
RT	Reverse Transcription
RT-PCR	Reverse Transcription Polymerase Chain Reaction
SGD	Saccharomyces Genome Database
SIC	Strong Irrepresentability Condition
siRNA	silencing RNA
SNR	Signal to Noise Ratio
SSV	Structured Singular Value
SVD	Singular Value Decomposition
TF	Transcription Factor
TGF- $\beta$	Transforming Growth Factor beta
TOEL	Two gene Example
WRSS	Weighted Residual Sum of Squares

---

# Index

---

- activating interaction, 29, 108, 129
- adjacency matrix, 29, 40, 60
- Akaike information criterion, 69, 88
- alternative interaction, 66, 67, 128, 129, 131, 132, 181
- amplification efficiency, 24, 26
- attribute selection, 69, 86
- automated discovery, 87, 138, 203
  
- Bayesian information criterion, 69, 88
- biological function, 3, 13, 34
- block diagram, 265, 266
- bow-tie, 57
  
- cascade, 35, 49, 187, 265
- causality, 170
- characteristic mode, 13, 34, 46, 55–57, 187, 199
- chi-square distribution, 101, 135, 185
- chi-square test, 101
- class of functions, 95
- co-regulated genes, 38, 45, 58, 199
- collinearity affects selectability, 118
- column uncertainty, 99
- comparative  $C_T$  method, 22–24
- condition number, 42, 45
- conditioning problem, 89, 113, 138
- confidence level, 98
- confidence score, 79, 80, 133, 135, 190, 224
- connected set, 121
- connectivity matrix, 40
- consistent, 97, 99, 102, 103, 183
- constrained optimization, 227
- controllability, 160
- controllability matrix, 160
  
- controlled variables, 94
- convex cone, 120
- convex set, 120
- correlated perturbations, 46, 57, 187
- correlated response, 38, 46, 48
- correlation coefficient, 220
  
- data model, 31, 79, 94, 182
- data property, 43
- data space, 129
- degree of interampattiness, 44, 45, 50, 51, 53, 54, 56, 187
- degrees of freedom, 70, 185
- $\Delta\Delta C_T$  method, 22, 24, 79
- dependent variables, 95
- deterministic error, 79, 97, 99, 100, 182
- direct interaction, 169
- distance to rank deficiency, 76
- dyadic expansion, 209
- dynamical mode, 28
- dynamical structure function, 151, 165
  
- E. coli*, 54, 57
- effective rank, 115, 208, 209
- eigengene, 13, 34, 46, 57, 187, 199
- eigenvalue, 42
- element-wise error, 99
- error assumption, 98
- error model, 25, 91, 94, 97, 100, 215
- errors-in-variables, 31, 89, 94
- existing interaction, 66, 107, 129, 131, 132, 181
- experiment design, 9, 12, 54, 57, 177
- explained variables, 95
- explanatory variables, 94

- exposure variables, 94  
 feasible, 104  
 feasible model, 75  
 feasible solution, 104  
 feature, 95  
 feature reduction, 69, 86  
 feature selection, 69, 86  
 feedback, 35, 49, 50, 57, 187, 204  
 feedforward, 49, 50  
 fundamental interconnections, 49  
  
 gain, 43  
 gene regulatory network, 1, 6, 20, 21, 28, 36, 37, 39–41, 44, 60, 81, 176, 179, 180, 186, 187, 192, 197, 201, 204  
 gene space, 27, 190  
 goodness of fit test, 72, 73, 179, 180  
 Granger causality, 171  
 growth medium, 22  
  
 halfspace, 120  
 hidden state, 161  
 housekeeping gene, 25, 26  
 hyperplane, 119  
  
 identifiability, 70, 181  
 ill-conditioned, 13, 14, 34, 35, 38, 44, 51, 54, 55, 58, 89, 113, 186  
 impulse response, 159  
 independent variables, 94  
 indirect interaction, 169  
 influence, 29, 82, 170  
 informative enough, 183, 184, 189  
 input space, 39, 44  
 input variables, 95  
 input-output relationship, 157, 158  
 interaction matrix, 28, 30, 40, 63, 64, 81, 182  
 interampatte, 34, 36, 37, 44–46, 54, 57, 72, 186, 187, 199, 204  
 IRMA, 20, 21, 25, 63, 64, 79–81, 154–156, 201, 214, 215, 246, 248  
  
 iterative design, 186–188, 193, 204  
  
 Jacobian, 40  
  
 Laplace transform, 161, 163, 265  
 LASSO, 7, 68, 71  
 latent state, 29, 150, 160, 161, 168, 203  
 likelihood ratio test, 88  
 Lilliefors test, 25  
 linear mapping, 30  
 linear regression, 63  
 linearly independent, 190  
 loop-gain, 50  
  
 manipulated variables, 94  
 McMillan degree, 168  
 measure of confidence, 62, 71  
 measured variables, 95  
 microarray, 40  
 minimal realization, 162  
 model invalidation, 138  
 model selection, 69, 86, 138, 177  
 model set, 95  
 mRNA abundance, 27  
 multi-gene perturbation, 192–194  
 multicollinearity, 89, 113  
  
 near collinearity, 89, 113, 114  
 network graph, 265, 266  
 network inference, 2, 4, 5, 27, 28, 51, 54, 60, 63, 69, 93, 177, 179, 181, 197  
 network matrix, 40  
 network model, 94, 182  
 network motifs, 49  
 network reconstruction, 4, 60, 93  
 network structure, 41  
 non-evidential interaction, 66, 67, 126, 129, 181  
 non-existing interaction, 66, 123, 129, 181  
 normal distribution, 25, 52, 79, 101, 135, 182, 185, 216  
 normalization, 25

- NP-hard problem, 8, 87
- observability, 160
- observability matrix, 160
- observed variables, 28, 95, 182
- Occam's razor, 122
- optimal experimental design, 10, 177
- order of the system, 266
- output space, 39, 44
- output variables, 95
- parameter estimation, 69, 138, 209
- path length, 169
- PCR, 23
- perturbation experiments, 37, 64, 179, 186–188, 190
- poles, 266
- poor data, 176, 186, 194
- practical, 106
- practical alternative, 127, 129, 131, 181
- practical assignability, 108, 114, 119–121, 129
- practical collinearity, 113, 130, 131, 136, 184, 188
- practical excludability, 123, 124, 129–131, 136, 181, 183
- practical existence, 105, 110, 132
- practical identifiability, 106
- practical independence, 112, 115, 130–133, 135, 136, 184, 188, 224
- practical negligibility, 125–127, 129–131, 181
- practical omissibility, 122
- practical rank, 114, 115, 208
- practical selectability, 107, 114, 116–118, 129–131, 136, 181, 183, 185, 225
- practical span, 110, 184, 185, 188, 190
- practical uniqueness, 105, 111, 112
- practical zero vector, 125, 130
- practically indistinguishable, 104
- practically negative, 107, 120, 129
- practically nonzero, 107, 129
- practically positive, 107, 120, 129
- practically unique, 129
- practically zero, 107, 129
- precision, 4, 179
- prediction error, 53, 54, 72, 73, 180
- predictor variables, 94
- principle of parsimony, 66, 122, 162
- process error, 27
- projection matrix, 67, 119
- proper, 267
- pseudo direct, 170
- qPCR, 40
- qRT-PCR, 22, 23, 26, 215
- rank, 114
- regressand, 75, 94, 182
- regression problem, 31
- regressor, 75, 94, 182
- regularization coefficient, 68
- relative degree, 169
- relative order, 267
- repressing interaction, 29, 108, 129
- residual sum of squares, 72, 73, 180, 271, 272
- responding variables, 95
- response variables, 95
- reverse engineering, 4, 35, 60, 93
- RNA interference, 40
- robust, 87, 97
- robust network inference, 82, 94, 103, 137, 181, 182, 184, 189, 193
- robust variable selection, 77, 91, 103, 137, 181, 183, 188, 189
- S-system, 28
- S. cerevisiae*, 20, 26, 54, 57, 62, 72, 79, 82, 179
- semi-proper, 267
- sensitivity, 4, 179
- set of feasible solutions, 104
- set of indistinguishable models, 74, 104
- set of nonrejectable models, 104

- signal amplification, 37, 45, 57, 187
- signal attenuation, 37, 45, 57, 187, 195, 204
- signal to noise ratio, 130, 185, 208
- signed adjacency matrix, 29
- significance level, 97, 98, 133, 185
- similarity transformation, 157, 161
- singular value, 42, 43, 51
- singular value decomposition, 30, 38, 42, 43, 56, 134, 178, 190, 208
- sloppy parameters, 194
- span, 66, 110
- sparse, 6, 7, 54, 67, 72, 106, 180, 197
- state space, 27, 188
- state vector, 27, 40
- state-space model, 27, 40, 153
- static gain, 30, 266
- steady-state data, 29, 41, 80, 82
- steady-state experiments, 29, 37, 169, 179, 190
- stochastic error, 101
- stochastic noise, 97, 100, 182
- strictly proper, 267
- strong gain, 56
- strong irrepresentability condition, 71
- structured singular value, 80, 132, 134, 145, 224
- subnetwork, 151
- subset selection, 69, 86
- system, 39, 43, 187
- system of linear ordinary differential equations, 27
- system property, 13, 43, 46, 57
  
- t-test, 72, 73
- Taylor approximation, 28
- time-constant, 48, 51
- time-series data, 29
- time-series experiments, 48, 169
- TOEL, 37, 99, 100, 129, 131, 229
- transcription factor, 6
- transfer function, 255, 265, 266
- transfer matrix, 163, 164
  
- uncertainty cone, 129, 131
- uncertainty set, 75, 77, 91, 94, 96, 98–101, 129, 182, 186, 233
- uncorrelated response, 46
- uniform distribution, 101
  
- variable selection, 5, 8, 69, 86, 200, 209
  
- weak gain, 57, 72, 187, 188
- weak mode, 46, 48, 57, 199
  
- zeros, 267

---

## Bibliography

---

- S. Aelst, J. A. Khan, and R. H. Zamar. Fast robust variable selection. In P. Brito, editor, *Compstat 2008*, 359–370. Physica-Verlag HD (2008).
- H. Akaike. Information Theory and an Extension of the Maximum Likelihood Principle. In B. N. Petrov and F. Csaki, editors, *Second International Symposium on Information Theory*, 267 – 281. Akademiai Kiado, Budapest (1973).
- B. Aksasse, Y. Stitou, Y. Berthoumieu, and M. Najim. 3-D AR model order selection via rank test procedure. *IEEE Transactions on Signal Processing*, 54(7): 2672–2677 (2006).
- J. Alander, A. Autere, O. Kannianen, J. Koljonen, T. E. M. Nordling, and P. Välisuo. Near infrared wavelength relevance detection of ultraviolet radiation-induced erythema. *Journal of Near Infrared Spectroscopy*, 16(3): 233 (2008).
- A. Alin. Multicollinearity. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(3): 370–374 (2010).
- U. Alon, M. G. Surette, N. Barkai, and S. Leibler. Robustness in bacterial chemotaxis. *Nature*, 397(6715): 168–171 (1999).
- O. Alter. Genomic signal processing: from matrix algebra to genetic networks. *Methods Mol Biol*, 377: 17–60 (2007).
- O. Alter, P. O. Brown, and D. Botstein. Singular value decomposition for genome-wide expression data processing and modeling. *Proc Natl Acad Sci U S A*, 97(18): 10101–10106 (2000).
- O. Alter and G. H. Golub. Singular value decomposition of genome-scale mRNA lengths distribution reveals asymmetry in RNA gel electrophoresis band broadening. *Proc Natl Acad Sci U S A*, 103(32): 11828–11833 (2006).
- H. Anton and C. Rorres. *Elementary linear algebra: applications version*. John Wiley & Sons, Inc., New York, NY, USA, 8th edition (2000).
- H. Anton and C. Rorres. *Elementary linear algebra with applications*. John Wiley & Sons, Inc., New York, NY, USA, 9th edition (2005).

- K. Aoki, Y. Ogata, and D. Shibata. Approaches for extracting practical information from gene co-expression networks in plant biology. *Plant Cell Physiol*, 48(3): 381–390 (2007).
- J. F. Apgar, J. E. Toettcher, D. Endy, F. M. White, and B. Tidor. Stimulus Design for Model Selection and Validation in Cell Signaling. *PLoS Comput Biol*, 4(2): e30 (2008).
- M. I. Arnone and E. H. Davidson. The hardwiring of development: Organization and function of genomic regulatory systems. *Development*, 124(10): 1851–1864 (1997).
- S. Aronov, R. Gelin-Licht, G. Zipor, L. Haim, E. Safran, and J. E. Gerst. mRNAs encoding polarity and exocytosis factors are cotransported with the cortical endoplasmic reticulum to the incipient bud in *Saccharomyces cerevisiae*. *Molecular and cellular biology*, 27(9): 3441–55 (2007).
- M. Ashyraliyev, Y. Fomekong-Nanfack, J. A. Kaandorp, and J. G. Blom. Systems biology: parameter estimation for biochemical models. *The FEBS journal*, 276(4): 886–902 (2009).
- R. C. Aster, B. Borchers, and C. H. Thurber. *Parameter estimation and inverse problems*, volume 90. Elsevier Academic Press, Amsterdam (2005).
- J. R. Banga and E. Balsa-Canto. Parameter estimation and optimal experimental design. *Essays in biochemistry: Systems biology*, 45: 195–209 (2008).
- M. Bansal, V. Belcastro, A. Ambesi-Impiombato, and D. di Bernardo. How to infer gene networks from expression profiles. *Molecular systems biology*, 3: 78 (2007).
- M. Bansal, G. D. Gatta, and D. di Bernardo. Inference of gene regulatory networks and compound mode of action from time course gene expression profiles. *Bioinformatics*, 22(7): 815–822 (2006).
- A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439): 509–512 (1999).
- A.-L. Barabási, N. Gulbahce, and J. Loscalzo. Network medicine: a network-based approach to human disease. *Nature reviews. Genetics*, 12(1): 56–68 (2011).
- A. Baralla, W. I. Mentzen, and A. de la Fuente. Inferring gene networks: dream or nightmare? *Annals of the New York Academy of Sciences*, 1158: 246–56 (2009).
- N. Barkai and S. Leibler. Robustness in simple biochemical networks. *Nature*, 387(6636): 913–7 (1997).
- Y. Barlas. Formal aspects of model validity and validation in system dynamics. *System Dynamics Review*, 12(3): 183–210 (1996).

- J. M. S. Bartlett and D. Stirling. A short history of the polymerase chain reaction. *Methods in molecular biology (Clifton, N.J.)*, 226: 3–6 (2003).
- K. Basso, A. a. Margolin, G. Stolovitzky, U. Klein, R. Dalla-Favera, and A. Califano. Reverse engineering of regulatory networks in human B cells. *Nature genetics*, 37(4): 382–90 (2005).
- R. Bellman and K. J. Åström. On structural identifiability. *Mathematical Biosciences*, 7(3-4): 329–339 (1970).
- D. A. Belsley. *Conditioning diagnostics : collinearity and weak data in regression*. John Wiley and Sons, Inc., New York (1991).
- P. N. Benfey and T. Mitchell-Olds. From genotype to phenotype: systems biology meets natural variation. *Science (New York, N.Y.)*, 320(5875): 495–7 (2008).
- M. Bengtsson, M. Hemberg, P. Rorsman, and A. Ståhlberg. Quantification of mRNA in single cells and modelling of RT-qPCR induced noise. *BMC molecular biology*, 9: 63 (2008).
- P. J. Bhat and T. V. Murthy. Transcriptional control of the GAL/MEL regulation of yeast *Saccharomyces cerevisiae*: mechanism of galactose-mediated signal transduction. *Molecular microbiology*, 40(5): 1059–66 (2001).
- L. T. Bhoite, Y. Yu, and D. J. Stillman. The Swi5 activator recruits the Mediator complex to the HO promoter without RNA polymerase II. *Genes & development*, 15(18): 2457–69 (2001).
- N. H. Bingham and J. M. Fry. *Regression, Linear models in statistics*. Springer London (2010).
- N. Bobola, R. P. Jansen, T. H. Shin, and K. Nasmyth. Asymmetric accumulation of Ash1p in postanaphase nuclei depends on a myosin and restricts yeast mating-type switching to mother cells. *Cell*, 84(5): 699–709 (1996).
- S. Boccaletti. Complex networks: structure and dynamics. *Physics Reports*, 424(4-5): 175–308 (2006).
- R. Bonneau. Learning biological networks: from modules to dynamics. *Nature chemical biology*, 4(11): 658–64 (2008).
- R. Bonneau, D. J. Reiss, P. Shannon, M. Facciotti, L. Hood, N. S. Baliga, and V. Thorsson. The Inferelator: an algorithm for learning parsimonious regulatory networks from systems-biology data sets de novo. *Genome biology*, 7(5): R36 (2006).
- G. E. P. Box and N. R. Draper. *Empirical model-building and response surfaces*. John Wiley and Sons, Inc. (1987).

- S. P. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, Cambridge (2004).
- P. Brazhnik. Inferring gene networks from steady-state response to single-gene perturbations. *J Theor Biol*, 237(4): 427–440 (2005).
- P. Brazhnik, A. de la Fuente, and P. Mendes. Gene networks: how to put the function in genomics. *Trends Biotechnol*, 20(11): 467–472 (2002).
- K. S. Brown, C. C. Hill, G. A. Calero, C. R. Myers, K. H. Lee, J. P. Sethna, and R. A. Cerione. The statistical mechanics of complex signaling networks: nerve growth factor signaling. *Physical biology*, 1(3-4): 184–95 (2004).
- M.-J. Bruwer and J. F. MacGregor. Robust multi-variable identification: Optimal experimental design with constraints. *Journal of Process Control*, 16(6): 581–600 (2006).
- K. P. Burnham and D. R. Anderson. *Model selection and multimodel inference: a practical information-theoretic approach*. Springer New York, New York, NY, USA, 2nd edition (2002).
- D. Butler. Human genome at ten: Science after the sequence. *Nature*, 465(7301): 1000–1 (2010).
- G. Buzzi Ferraris, P. Forzatti, G. Emig, and H. Hofmann. Sequential experimental design for model discrimination in the case of multiple responses. *Chemical Engineering Science*, 39(1): 81–85 (1984).
- M. Cacciottolo, V. Belcastro, S. Laval, K. Bushby, D. di Bernardo, and V. Nigro. Reverse engineering gene network identifies new dysferlin-interacting proteins. *The Journal of biological chemistry*, 286(7): 5404–13 (2011).
- E. Candes and M. Wakin. An Introduction To Compressive Sampling. *IEEE Signal Processing Magazine*, 25(2): 21–30 (2008).
- E. J. Candès and Y. Plan. Near-ideal model selection by  $\ell_1$  minimization. *The Annals of Statistics*, 37(5A): 2145–2177 (2009).
- I. Cantone, L. Marucci, F. Iorio, M. A. Ricci, V. Belcastro, M. Bansal, S. Santini, M. di Bernardo, D. di Bernardo, and M. P. Cosma. A yeast synthetic network for in vivo assessment of reverse-engineering and modeling approaches. *Cell*, 137(1): 172–81 (2009).
- G. Casella and R. L. Berger. *Statistical Inference*. Duxbury Press, 2nd edition (2001).

- M. Cassman, A. P. Arkin, F. Doyle, F. Katagiri, D. Lauffenburger, and C. Stokes. *WTEC panel report on International Research and Development in Systems Biology*. World Technology Evaluation Center, Inc., 2809 Boston Street, Suite 441, Baltimore, Maryland 21224, USA (2005).
- G. Cedersund and J. Roll. Systems biology: model based evaluation and comparison of potential explanations for given biological data. *The FEBS journal*, 276(4): 903–22 (2009).
- K. Chaloner and I. Verdinelli. Bayesian experimental design: A review. *Statistical Science*, 10(3): 273–304 (1995).
- M. Chapman and K. Godfrey. Nonlinear compartmental model indistinguishability. *Automatica*, 32(3): 419–422 (1996).
- P. Chartrand, X. H. Meng, S. Huttelmaier, D. Donato, and R. H. Singer. Asymmetric sorting of ash1p in yeast results from inhibition of translation by localization elements in the mRNA. *Molecular cell*, 10(6): 1319–30 (2002).
- B. H. Chen and S. P. Asprey. On the Design of Optimally Informative Dynamic Experiments for Model Discrimination in Multiresponse Nonlinear Situations. *Industrial & Engineering Chemistry Research*, 42(7): 1379–1390 (2003).
- V. Chew. Confidence, Prediction, and Tolerance Regions for the Multivariate Normal Distribution. *Journal of the American Statistical Association*, 61(315): 605 – 617 (1966).
- K. H. Cho, S. M. Choo, S. H. Jung, J. R. Kim, H. S. Choi, and J. Kim. Reverse engineering of gene regulatory networks. *IET Syst Biol*, 1(3): 149–163 (2007).
- C. Christensen, J. Thakar, and R. Albert. Systems-level insights into cellular regulation: inferring, analysing, and modelling intracellular networks. *IET Syst Biol*, 1(2): 61–77 (2007).
- J. S. Chuang, O. Rivoire, and S. Leibler. Cooperation and Hamilton’s rule in a simple synthetic microbial system. *Molecular systems biology*, 6(398): 398 (2010).
- S. Chung and P. A. Takizawa. Multiple Myo4 motors enhance ASH1 mRNA transport in *Saccharomyces cerevisiae*. *The Journal of cell biology*, 189(4): 755–67 (2010).
- S. Cikos, A. Bukovská, and J. Koppel. Relative quantification of mRNA: comparison of methods currently used for real-time PCR data analysis. *BMC molecular biology*, 8: 113 (2007).
- S. Cikos and J. Koppel. Transformation of real-time PCR fluorescence data to target gene quantity. *Analytical biochemistry*, 384(1): 1–10 (2009).

- C. Cosentino, W. Curatola, F. Montefusco, M. Bansal, D. di Bernardo, and F. Amato. Linear matrix inequalities approach to reconstruction of biological networks. *IET Syst Biol*, 1(3): 164–173 (2007).
- E. J. Cosgrove, T. S. Gardner, and E. D. Kolaczyk. On the choice and number of microarrays for transcriptional regulatory network inference. *BMC bioinformatics*, 11(1): 454 (2010).
- E. J. Cosgrove, Y. Zhou, T. S. Gardner, and E. D. Kolaczyk. Predicting gene targets of perturbations via network-based filtering of mRNA expression compendia. *Bioinformatics (Oxford, England)*, 24(21): 2482–90 (2008).
- M. P. Cosma. Daughter-specific repression of *Saccharomyces cerevisiae* HO: Ash1 is the commander. *EMBO reports*, 5(10): 953–7 (2004).
- M. P. Cosma, T. Tanaka, and K. Nasmyth. Ordered recruitment of transcription and chromatin remodeling factors to a cell cycle- and developmentally regulated promoter. *Cell*, 97(3): 299–311 (1999).
- C. Cotta, C. Sloper, and P. Moscato. Evolutionary Search of Thresholds for Robust Feature Set Selection: Application to the Analysis of Microarray Data. In G. R. Raidl, S. Cagnoni, J. Branke, D. W. Corne, R. Drechsler, Y. Jin, C. G. Johnson, P. Machado, E. Marchiori, F. Rothlauf, G. D. Smith, and G. Squillero, editors, *Applications of Evolutionary Computing, Lecture notes in Computer Science*, volume 3005 of *Lecture Notes in Computer Science*, 21–30. Springer Berlin Heidelberg, Berlin, Heidelberg (2004).
- T. M. Cover and J. M. Van Campenhout. On the Possible Orderings in the Measurement Selection Problem. *IEEE Transactions on Systems, Man, and Cybernetics*, 7(9): 657–661 (1977).
- E. J. Crampin. System Identification Challenges from Systems Biology. In B. Ninness and H. Hjalmarson, editors, *Proc. 14th IFAC Symposium on System Identification*, volume 14, 81–93. The International Federation of Automatic Control, Australia (2006).
- E. J. Crampin, S. Schnell, and P. E. McSharry. Mathematical and computational techniques to deduce complex biochemical reaction mechanisms. *Prog Biophys Mol Biol*, 86(1): 77–112 (2004).
- F. Crick. Central dogma of molecular biology. *Nature*, 227(5258): 561–563 (1970).
- M. Csete and J. Doyle. Bow ties, metabolism and disease. *Trends Biotechnol*, 22(9): 446–450 (2004).
- M. E. Csete and J. C. Doyle. Reverse engineering of biological complexity. *Science (New York, N.Y.)*, 295(5560): 1664–9 (2002).

- M. Dash and H. Liu. Feature selection for classification. *Intelligent data analysis*, 1: 131–156 (1997).
- D. Datta and H. Zhao. Statistical methods to infer cooperative binding among transcription factors in *Saccharomyces cerevisiae*. *Bioinformatics (Oxford, England)*, 24(4): 545–52 (2008).
- P. P. N. de Groen. An Introduction to Total Least Squares. *Nieuw archief voor wiskunde*, 14: 237–253 (1998).
- H. de Jong. Modeling and simulation of genetic regulatory systems: a literature review. *J Comput Biol*, 9(1): 67–103 (2002).
- A. de la Fuente, N. Bing, I. Hoeschele, and P. Mendes. Discovery of meaningful associations in genomic data using partial correlation coefficients. *Bioinformatics (Oxford, England)*, 20(18): 3565–74 (2004).
- R. De Smet and K. Marchal. Advantages and limitations of current network inference methods. *Nature reviews. Microbiology*, 8(10): 717–29 (2010).
- Y. Deng, R. H. Singer, and W. Gu. Translation of ASH1 mRNA is repressed by Puf6p-Fun12p/eIF5B interaction and released by CK2 phosphorylation. *Genes & development*, 22(8): 1037–50 (2008).
- P. D’haeseleer, S. Liang, and R. Somogyi. Genetic network inference: from co-expression clustering to reverse engineering. *Bioinformatics*, 16(8): 707–726 (2000).
- J. H. Do and D.-K. Choi. Clustering approaches to identifying gene expression patterns from DNA microarray data. *Mol Cells*, 25(2): 279–288 (2008).
- B. Donckels, D. Depauw, B. Debaets, J. Maertens, and P. Vanrolleghem. An anticipatory approach to optimal experimental design for model discrimination. *Chemometrics and Intelligent Laboratory Systems*, 95(1): 53–63 (2009).
- F. Doyle and D. Lauffenburger. *WTEC panel report on International Research and Development in Systems Biology*, chapter Network In, 19–27. World Technology Evaluation Center, Inc., 2809 Boston Street, Suite 441, Baltimore, Maryland 21224, USA (2005).
- J. Doyle. Analysis of feedback systems with structured uncertainties. *Control Theory and Applications, IEE Proceedings D*, 129(6): 242–250 (1982).
- G. E. Dullerud and F. G. Paganini. *A course in robust control theory: a convex approach*. Springer-Verlag New York, Inc., New York, NY, USA (2000).
- M. B. Eisen, P. T. Spellman, P. O. Brown, and D. Botstein. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A*, 95(25): 14863–14868 (1998).

- T. Eissing, F. Allgöwer, and E. Bullinger. Robustness properties of apoptosis models with respect to parameter variations and intrinsic noise. *Syst Biol (Stevenage)*, 152(4): 221–228 (2005).
- S. M. Elbashir, J. Harborth, W. Lendeckel, A. Yalcin, K. Weber, and T. Tuschl. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature*, 411(6836): 494–498 (2001).
- F. Emmert-Streib, G. V. Glazko, G. Altay, and R. de Matos Simoes. Statistical inference and reverse engineering of gene regulatory networks from observational expression data. *Frontiers in genetics*, 3: 8 (2012).
- J. J. Faith, B. Hayete, J. T. Thaden, I. Mogno, J. Wierzbowski, G. Cottarel, S. Kasif, J. J. Collins, and T. S. Gardner. Large-Scale Mapping and Validation of *Escherichia coli* Transcriptional Regulation from a Compendium of Expression Profiles. *PLoS Biol*, 5(1): 54–66 (2007).
- D. Faller, U. Klingmüller, and J. Timmer. Simulation Methods for Optimal Experimental Design in Systems Biology. *Simulation*, 79(12): 717–725 (2003).
- J. Fan and Y. Fan. High Dimensional Classification Using Features Annealed Independence Rules. *Annals of statistics*, 36(6): 2605–2637 (2008).
- J. Fan and R. Li. Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties. *Journal of the American Statistical Association*, 96(456): 1348–1360 (2001).
- J. Fan and R. Li. Statistical challenges with high dimensionality: feature selection in knowledge discovery. In *Proceedings of the International Congress of Mathematicians*. European Mathematical Society, Madrid, Spain (2006).
- J. Fan and J. Lv. A Selective Overview of Variable Selection in High Dimensional Feature Space (Invited Review Article). *Statistica Sinica*, 20(1): 101–148 (2010).
- X. J. Feng, S. Hooshangi, D. Chen, G. Li, R. Weiss, and H. Rabitz. Optimizing genetic circuits by global sensitivity analysis. *Biophys J*, 87(4): 2195–2202 (2004).
- X. J. Feng and H. Rabitz. Optimal identification of biochemical reaction networks. *Biophys J*, 86(3): 1270–1281 (2004).
- X. J. Feng, H. Rabitz, G. Turinici, and C. Le Bris. A closed-loop identification protocol for nonlinear dynamical systems. *J Phys Chem A Mol Spectrosc Kinet Environ Gen Theory*, 110(25): 7755–7762 (2006).
- J. A. Ferreiro, N. G. Powell, N. Karabetsov, N. A. Kent, J. Mellor, and R. Waters. Cbf1p modulates chromatin structure, transcription and repair at the *Saccharomyces cerevisiae* MET16 locus. *Nucleic acids research*, 32(5): 1617–26 (2004).

- V. Filkov. Identifying gene regulatory networks from gene expression data. In S. Aluru, editor, *Handbook of computational molecular biology*, chapter 27, 27.1–27.29. Chapman & Hall/CRC Press (2005).
- R. A. Fisher. *The design of experiments*. Oliver & Boyd, London (1935).
- S. Fleige and M. W. Pfaffl. RNA integrity and the effect on the real-time qRT-PCR performance. *Molecular aspects of medicine*, 27(2-3): 126–39 (2006).
- R. D. Fleischmann, M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J. F. Tomb, B. A. Dougherty, and J. M. Merrick. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science*, 269(5223): 496–512 (1995).
- G. Franceschini and S. Macchietto. Model-based design of experiments for parameter precision: State of the art. *Chemical Engineering Science*, 63(19): 4846–4872 (2008).
- D. Freedman. A Note on Screening Regression Equations. *American Statistician*, 37(2): 152–155 (1983).
- S. H. Friedberg, A. J. Insel, and L. E. Spence. *Linear algebra*. Pearson Education, Ltd., Upper Saddle River, New Jersey 07458, USA, 4th edition (2003).
- J. Friedman, T. Hastie, and R. Tibshirani. Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal Of Statistical Software*, 33(1) (2010).
- S. Fu and M. C. Desmarais. Markov Blanket Based Feature Selection: a Review of Past Decade. In *Proceedings of the World Congress on Engineering 2010*, 321–328. IAENG International Association of Engineers, Hong Kong, China (2010).
- A. Fujita, A. G. Patriota, J. R. Sato, and S. Miyano. The impact of measurement errors in the identification of regulatory networks. *BMC Bioinformatics*, 10(1): 412 (2009).
- K. G. Gadkar, J. Varner, and F. J. Doyle. Model identification of signal transduction networks from data using a state regulator problem. *Syst Biol (Stevenage)*, 2(1): 17–30 (2005).
- T. S. Gardner, D. di Bernardo, D. Lorenz, and J. J. Collins. Inferring genetic networks and identifying compound mode of action via expression profiling. *Science*, 301(5629): 102–105 (2003).
- T. S. Gardner and J. J. Faith. Reverse-engineering transcription control networks. *Physics of Life Reviews*, 2(1): 65–88 (2005).
- E. I. George. The variable selection problem. *Journal of the American Statistical Association*, 95(452): 1304–1308 (2000).

- J. R. Gilbert. Predicting structures in sparse matrix computations. *SIAM Journal on Matrix Analysis and Applications*, 15(1): 62–79 (1994).
- J. Goncalves, R. Howes, and S. Warnick. Dynamical structure functions for the reverse engineering of LTI networks. *Decision and Control, 2007 46th IEEE Conference on*, 1516–1522 (2007).
- J. Goncalves and S. Warnick. Necessary and Sufficient Conditions for Dynamical Structure Reconstruction of LTI Networks. *Automatic Control, IEEE Transactions on*, 53(7): 1670–1674 (2008).
- G. B. Gonsalvez, K. A. Lehmann, D. K. Ho, E. S. Stanitsa, J. R. Williamson, and R. M. Long. RNA-protein interactions promote asymmetric sorting of the ASH1 mRNA ribonucleoprotein complex. *RNA (New York, N.Y.)*, 9(11): 1383–99 (2003).
- J. Goutsias and N. H. Lee. Computational and experimental approaches for modeling gene regulatory networks. *Current pharmaceutical design*, 13(14): 1415–36 (2007).
- C. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society* (1969).
- M. Green and D. J. N. Limebeer. *Linear robust control*. Prentice Hall information and system sciences series. Prentice Hall, Englewood Cliffs, N.J. (1995).
- Z. Griliches and J. A. Hausman. Errors in variables in panel data. *Journal of Econometrics*, 31(1): 93–118 (1986).
- J. L. Gross and J. Yellen. *Graph theory and its applications*. CRC Press, Boca Raton, Fla. (1999).
- P. Grünwald and S. De Rooij. Catching Up Faster by Switching Sooner: A Predictive Approach to Adaptive Estimation with an application to the AIC-BIC Dilemma. *Journal of the Royal Statistical Society - Series B: Statistical Methodology*, 74(3): 361–417 (2011).
- D.-W. Gu, P. H. Petkov, and M. M. Konstantinov. *Robust control design with Matlab*. Springer-Verlag London Limited, London (2005).
- R. N. Gutenkunst, J. J. Waterfall, F. P. Casey, K. S. Brown, C. R. Myers, and J. P. Sethna. Universally sloppy parameter sensitivities in systems biology models. *PLoS computational biology*, 3(10): 1871–78 (2007).
- I. Guyon and A. Elisseeff. An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research*, 3: 1157–1182 (2003).
- H. Hache, H. Lehrach, and R. Herwig. Reverse engineering of gene regulatory networks: a comparative study. *EURASIP journal on bioinformatics & systems biology*, 617281 (2009).

- H. Hache, C. Wierling, H. Lehrach, and R. Herwig. Reconstruction and validation of gene regulatory networks with neural networks. In *Proceedings of Foundations of Systems Biology in Engineering (FOSBE), 2nd Conference*, 319–324. Stuttgart, Germany (2007).
- J. Handl and J. Knowles. Feature subset selection in unsupervised learning via multiobjective optimization. *International Journal of Computational Intelligence Research*, 2(3): 217–238 (2006).
- R. B. O. Hara and M. J. Sillanp. A Review of Bayesian Variable Selection Methods: What, How and Which. *Bayesian Analysis*, 4(1): 85–118 (2009).
- N. Haruta, Y. Akamatsu, Y. Tsutsui, Y. Kurokawa, Y. Murayama, B. Arcangioli, and H. Iwasaki. Fission yeast Swi5 protein, a novel DNA recombination mediator. *DNA repair*, 7(1): 1–9 (2008).
- F. He, R. Balling, and A.-P. Zeng. Reverse engineering and verification of gene networks: principles, assumptions, and limitations of present methods and future perspectives. *Journal of biotechnology*, 144(3): 190–203 (2009).
- M. Hecker, S. Lambeck, S. Toepfer, E. van Someren, and R. Guthke. Gene regulatory network inference: data integration in dynamic models-a review. *Bio Systems*, 96(1): 86–103 (2009).
- R. Heinrich and S. Schuster. *The regulation of cellular systems*. Chapman & Hall, New York (1996).
- M. Hellgren, T. E. M. Nordling, E. W. Jacobsen, and J. O. Höög. Multi-level modelling of the parallel metabolism of ethanol and retinol, with implications for foetal alcohol syndrome. In *The 9th International Conference on Systems Biology (ICSB-2008) in Gothenburg (Sweden): Abstract book*. University Of Gothenburg, Curran Associates, Inc., Gothenburg, Sweden (2008).
- D. M. Hendrickx, M. M. W. B. Hendriks, P. H. C. Eilers, A. K. Smilde, and H. C. J. Hoefsloot. Reverse engineering of metabolic networks, a critical assessment. *Molecular bioSystems*, 7(2): 511–20 (2011).
- N. J. Higham. *Accuracy and Stability of Numerical Algorithms*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA (1996).
- R. Higuchi, G. Dollinger, P. S. Walsh, and R. Griffith. Simultaneous amplification and detection of specific DNA sequences. *Bio/technology (Nature Publishing Company)*, 10(4): 413–7 (1992).
- N. Hiroi, A. Funahashi, T. E. M. Nordling, and H. Kitano. In silico analysis for anchorage independent cell cycle start mechanisms. In *The 4th International Conference on Systems Biology (ICSB-2003) in St. Louis (USA): Abstract book* (2003).

- R. Hoffmann and A. Valencia. A gene network for navigating the literature. *Nature genetics*, 36(7): 664 (2004).
- N. S. Holter, A. Maritan, M. Cieplak, N. V. Fedoroff, and J. R. Banavar. Dynamic modeling of gene expression data. *Proc Natl Acad Sci U S A*, 98(4): 1693–1698 (2001).
- N. S. Holter, M. Mitra, A. Maritan, M. Cieplak, J. R. Banavar, and N. V. Fedoroff. Fundamental patterns underlying gene expression profiles: simplicity from complexity. *Proc Natl Acad Sci U S A*, 97(15): 8409–8414 (2000).
- X. Hong, R. J. Mitchell, S. Chen, C. J. Harris, K. Li, and G. W. Irwin. Model selection approaches for non-linear system identification: a review. *International Journal of Systems Science*, 39(10): 925–946 (2008).
- L. Hood and S. H. Friend. Predictive, personalized, preventive, participatory (P4) cancer medicine. *Nature reviews. Clinical oncology*, 8(3): 184–7 (2011).
- L. Hood, J. R. Heath, M. E. Phelps, and B. Lin. Systems biology and new technologies enable predictive and preventative medicine. *Science (New York, N. Y.)*, 306(5696): 640–3 (2004).
- R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge university press (1990).
- S. Huang, I. Ernberg, and S. Kauffman. Cancer attractors: a systems view of tumors from a gene network dynamics and developmental perspective. *Seminars in cell & developmental biology*, 20(7): 869–76 (2009).
- T. Huckle. Approximate sparsity patterns for the inverse of a matrix and preconditioning. *Appl. Numer. Math.*, 30(2-3): 291–303 (1999).
- J. Huggett, K. Dheda, S. Bustin, and A. Zumla. Real-time RT-PCR normalisation; strategies and considerations. *Genes Immun*, 6(4): 279–284 (2005).
- T. R. Hughes, M. J. Marton, A. R. Jones, C. J. Roberts, R. Stoughton, C. D. Armour, H. A. Bennett, E. Coffey, H. Dai, Y. D. He, M. J. Kidd, A. M. King, M. R. Meyer, D. Slade, P. Y. Lum, S. B. Stepaniants, D. D. Shoemaker, D. Gachotte, K. Chakraburttu, J. Simon, M. Bard, and S. H. Friend. Functional discovery via a compendium of expression profiles. *Cell*, 102(1): 109–126 (2000).
- T. Ideker, T. Galitski, and L. Hood. A new approach to decoding life: systems biology. *Annual review of genomics and human genetics*, 2: 343–72 (2001).
- T. Ideker and R. Sharan. Protein networks in disease. *Genome research*, 18(4): 644–52 (2008).
- T. E. Ideker, V. Thorsson, and R. M. Karp. Discovery of regulatory interactions through perturbation: inference and experimental design. *Pac Symp Biocomput*, 305–316 (2000).

- IHGSC. Initial sequencing and analysis of the human genome. *Nature*, 409(6822): 860–921 (2001).
- IHGSC. Finishing the euchromatic sequence of the human genome. *Nature*, 431(7011): 931–945 (2004).
- V. R. Iyer, C. E. Horak, C. S. Scafe, D. Botstein, M. Snyder, and P. O. Brown. Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature*, 409(6819): 533–8 (2001).
- E. W. Jacobsen. Identification for Control of Strongly Interactive Plants. In *AIChE Annual Meeting*. San Francisco, California, U.S.A. (1994).
- E. W. Jacobsen and T. E. M. Nordling. On identification of genetic and metabolic networks. 15th ERNSI Workshop on System Identification in Linköping (Sweden) (2006).
- A. Jain, R. Duin, and J. Mao. Statistical pattern recognition: A review. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(1): 4–37 (2000).
- K. Jaqaman and G. Danuser. Linking data to models: data regression. *Nature reviews. Molecular cell biology*, 7(11): 813–9 (2006).
- H. Jeong, S. P. Mason, A.-L. Barabási, and Z. N. Oltvai. Lethality and centrality in protein networks. *Nature*, 411(6833): 41–2 (2001).
- H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabási. The large-scale organization of metabolic networks. *Nature*, 407(6804): 651–4 (2000).
- M. Johnston, J. S. Flick, and T. Pexton. Multiple mechanisms provide rapid and stringent glucose repression of GAL gene expression in *Saccharomyces cerevisiae*. *Molecular and cellular biology*, 14(6): 3834–41 (1994).
- R. Jörnsten, T. Abenius, T. Kling, L. Schmidt, E. Johansson, T. E. M. Nordling, B. Nordlander, C. Sander, P. Gennemark, K. Funa, B. Nilsson, L. Lindahl, and S. Nelander. Network modeling of the transcriptional effects of copy number aberrations in glioblastoma. *Molecular systems biology*, 7(1): 486 (2011).
- A. Julià, J. Moore, L. Miquel, C. Alegre, P. Barceló, M. Ritchie, and S. Marsal. Identification of a two-loci epistatic interaction associated with susceptibility to rheumatoid arthritis through reverse engineering and multifactor dimensionality reduction. *Genomics*, 90(1): 6–13 (2007).
- A. Julius, M. Zavlanos, S. Boyd, and G. J. Pappas. Genetic network identification using convex programming. *IET systems biology*, 3(3): 155–66 (2009).
- T. Kailath. *Linear systems*. Prentice Hall, Englewood Cliffs, N.J., U.S.A. (1980).

- G. Karlebach and R. Shamir. Modelling and analysis of gene regulatory networks. *Nature reviews. Molecular cell biology*, 9(10): 770–80 (2008).
- Y. Karlen, A. McNair, S. Perseguers, C. Mazza, and N. Mermod. Statistical significance of quantitative PCR. *BMC bioinformatics*, 8(1): 131 (2007).
- S. Kay. *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice hall PTR, Upper Saddle River, New Jersey 07458, USA (1993).
- B. N. Kholodenko. Cell-signalling dynamics in time and space. *Nature reviews. Molecular cell biology*, 7(3): 165–76 (2006).
- A. I. Khuri. *Advanced calculus with applications in statistics*. John Wiley & Sons, Inc., Hoboken, NJ, USA, 2nd edition (2003).
- J. Kim, D. G. Bates, I. Postlethwaite, L. Ma, and P. A. Iglesias. Robustness analysis of biochemical network models. *Systems biology*, 153(3): 96–104 (2006).
- H. A. Kishore and T. K. Kundu. Transcription through chromatin — Dynamic organization of genes. *Resonance*, 8(11): 78–93 (2003).
- H. Kitano. Computational systems biology. *Nature*, 420(6912): 206–210 (2002).
- H. Kitano. Biological robustness. *Nat Rev Genet*, 5(11): 826–837 (2004).
- H. Kitano. A robustness-based approach to systems-oriented drug design. *Nat Rev Drug Discov*, 6(3): 202–210 (2007a).
- H. Kitano. Towards a theory of biological robustness. *Mol Syst Biol*, 3 (2007b).
- S. Knudsen. *Guide to analysis of DNA microarray data*. John Wiley & Sons, New York, 2nd edition (2004).
- R. Kohavi. *Wrappers for performance enhancement and oblivious decision graphs*. Phd thesis, Stanford university (1995).
- R. Kohavi and G. H. John. Wrappers for feature subset selection. *Artificial Intelligence*, 97(1-2): 273–324 (1997).
- R. Kohavi and F. Provost. Applications of data mining to electronic commerce. *Data Mining and Knowledge Discovery*, 5(1): 5–10 (2001).
- W. Kolch. Defining systems biology: through the eyes of a biochemist. *IET systems biology*, 2(1): 5–7 (2008).
- J. Koljonen, J. T. Alander, T. E. M. Nordling, A. Autere, P. Välisuo, and O. Kanniainen. The Effects of Erythema on Near-Infrared Absorption Spectra. In *The 13th International Conference on Near Infrared Spectroscopy (13th ICNIRS) in Umeå-Vasa (Sweden and Finland): Abstract book*. Umeå-Vasa, Sweden and Finland (2007a).

- J. Koljonen, J. T. Alander, T. E. M. Nordling, and P. Välisuo. A review on evolutionary optimisation and search methods in NIR spectroscopy. In *The 13th International Conference on Near Infrared Spectroscopy (13th ICNIRS) in Umeå-Vasa (Sweden and Finland): Abstract book*. Umeå-Vasa, Sweden and Finland (2007b).
- J. Koljonen, T. E. M. Nordling, and J. Alander. A review of genetic algorithms in near infrared spectroscopy and chemometrics: past and future. *Journal of Near Infrared Spectroscopy*, 16(3): 189 (2008).
- K. Konstantinides and K. Yao. Statistical analysis of effective singular values in matrix rank determination. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 36(5): 757–763 (1988).
- A. Kremling, S. Fischer, K. Gadkar, F. J. Doyle, T. Sauter, E. Bullinger, F. Allgöwer, and E. D. Gilles. A benchmark for methods in reverse engineering and model discrimination: problem formulation and solutions. *Genome Res*, 14(9): 1773–1785 (2004).
- C. Kreutz, M. M. Bartolome Rodriguez, T. Maiwald, M. Seidl, H. E. Blum, L. Mohr, and J. Timmer. An error model for protein quantification. *Bioinformatics (Oxford, England)*, 23(20): 2747–53 (2007).
- C. Kreutz and J. Timmer. Systems biology: experimental design. *FEBS J*, 276(4): 923–942 (2009).
- A. Krishnan, A. Giuliani, and M. Tomita. Indeterminacy of reverse engineering of gene regulatory networks: the curse of gene elasticity. *PLoS ONE*, 2 (2007).
- E. Kritikou, B. Pulverer, and A. Heinrichs. All systems go! *Nature Reviews Molecular Cell Biology*, 7(11): 801–801 (2006).
- M. Kubista, J. M. Andrade, M. Bengtsson, A. Forootan, J. Jonák, K. Lind, R. Sindelka, R. Sjöback, B. Sjögreen, L. Strömbom, A. Ståhlberg, and N. Zoric. The real-time polymerase chain reaction. *Molecular aspects of medicine*, 27(2-3): 95–125 (2006).
- L. Kuras, R. Barbey, and D. Thomas. Assembly of a bZIP-bHLH transcription activation complex: formation of the yeast Cbf1-Met4-Met28 complex is regulated through Met28 stimulation of Cbf1 DNA binding. *The EMBO journal*, 16(9): 2441–51 (1997).
- F. G. Kuruvilla, P. J. Park, and S. L. Schreiber. Vector algebra in the analysis of genome-wide expression data. *Genome Biol*, 3(3): RESEARCH0011 (2002).
- H. Lähdesmäki and I. Shmulevich. Learning the structure of dynamic Bayesian networks from time series and steady state measurements. *Machine Learning*, 71(2-3): 185–217 (2008).

- E. S. Lander. Initial impact of the sequencing of the human genome. *Nature*, 470(7333): 187–97 (2011).
- D. T. Larose. *Data Mining Methods and Models*. John Wiley & Sons, Inc., Hoboken, NJ, USA (2005).
- P. Lecca, T.-P. Nguyen, C. Priami, and P. Quaglia. Network inference from time-dependent Omics data. *Methods in molecular biology (Clifton, N.J.)*, 719: 435–55 (2011).
- H. Leeb and B. M. Pötscher. Can one estimate the conditional distribution of post-model-selection estimators? *The Annals of Statistics*, 34(5): 2554–2591 (2006).
- H. Leeb and B. M. Pötscher. Can One Estimate The Unconditional Distribution of Post-Model-Selection Estimators? *Econometric Theory*, 24(02): 338–376 (2008).
- H. Li, J. Sun, X. Fan, X. Sui, L. Zhang, Y. Wang, and Z. He. Considerations and recent advances in QSAR models for cytochrome P450-mediated drug metabolism prediction. *Journal of Computer-Aided Molecular Design*, 22(11): 843–855 (2008a).
- H. Li, J. Xuan, Y. Wang, and M. Zhan. Inferring regulatory networks. *Front Biosci*, 13: 263–275 (2008b).
- D. Liao. *Collinearity Diagnostics for Complex Survey Data*. Phd thesis, University of Maryland, USA (2010).
- J. C. Liao, R. Boscolo, Y.-L. Yang, L. M. Tran, C. Sabatti, and V. P. Roychowdhury. Network component analysis: reconstruction of regulatory signals in biological systems. *Proc Natl Acad Sci U S A*, 100(26): 15522–15527 (2003).
- J. D. Lieb, X. Liu, D. Botstein, and P. O. Brown. Promoter-specific binding of Rap1 revealed by genome-wide maps of protein-DNA association. *Nature genetics*, 28(4): 327–34 (2001).
- F. Liljeros, C. R. Edling, L. A. Amaral, H. E. Stanley, and Y. Aberg. The web of human sexual contacts. *Nature*, 411(6840): 907–8 (2001).
- H. Lilliefors. On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association*, 62(318): 399–402 (1967).
- G. Lima-Mendez and J. van Helden. The powerful law of the power law and other myths in network biology. *Molecular bioSystems*, 5(12): 1482–93 (2009).
- D. K. Lindner. *Introduction to signals and systems*. WCB/McGraw-Hill, Boston (1999).

- J. W. Little, D. P. Shepley, and D. W. Wert. Robustness of a gene regulatory circuit. *The EMBO journal*, 18(15): 4299–307 (1999).
- H. Liu, H. Motoda, R. Setiono, and Z. Zhao. Feature Selection: An Ever Evolving Frontier in Data Mining. In N. Lawrence, editor, *JMLR Workshop and Conference Proceedings Volume 10: Feature Selection in Data Mining*, 4–13. JMLR, Hyderabad, India (2010).
- L. Ljung. *System identification*. Prentice hall PTR, Upper Saddle River, New Jersey 07458, USA, 2nd edition (1999).
- D. R. Lorenz, C. R. Cantor, and J. J. Collins. A network biology approach to aging in yeast. *Proceedings of the National Academy of Sciences of the United States of America*, 106(4): 1145–50 (2009).
- P. M. Lukacs, K. P. Burnham, and D. R. Anderson. Model selection bias and Freedman’s paradox. *Annals of the Institute of Statistical Mathematics*, 62(1): 117–125 (2009).
- G. MacBeath and S. L. Schreiber. Printing Proteins as Microarrays for High-Throughput Function Determination. *Science*, 289(5485): 1760–1763 (2000).
- T. Maiwald, C. Kreutz, A. C. Pfeifer, S. Bohl, U. Klingmüller, and J. Timmer. Dynamic pathway modeling: feasibility analysis and optimal experimental design. *Annals of the New York Academy of Sciences*, 1115(Reverse Engineering Biological Networks: Opportunities and Challenges in Computational Methods for Pathway Inference): 212–20 (2007).
- D. Marbach, R. J. Prill, T. Schaffter, C. Mattiussi, D. Floreano, and G. Stolovitzky. Revealing strengths and weaknesses of methods for gene network inference. *Proceedings of the National Academy of Sciences of the United States of America*, 107(14): 6286–91 (2010).
- A. A. Margolin and A. Califano. Theory and limitations of genetic network inference from microarray data. *Ann N Y Acad Sci*, 1115: 51–72 (2007).
- I. Markovsky and S. Van Huffel. Overview of total least-squares methods. *Signal Processing*, 87(10): 2283–2302 (2007).
- F. Markowetz and R. Spang. Inferring cellular networks—a review. *BMC bioinformatics*, 8 Suppl 6(Suppl 6): S5 (2007).
- H. H. McAdams and A. Arkin. Simulation of prokaryotic genetic circuits. *Annual review of biophysics and biomolecular structure*, 27: 199–224 (1998).
- V. Measday, H. McBride, J. Moffat, D. Stillman, and B. Andrews. Interactions between Pho85 cyclin-dependent kinase complexes and the Swi5 transcription factor in budding yeast. *Molecular Microbiology*, 35(4): 825–834 (2000).

- J. Mellor, W. Jiang, M. Funk, J. Rathjen, C. A. Barnes, T. Hinz, J. H. Hegemann, and P. Philippsen. CPF1, a yeast protein which functions in centromeres and promoters. *The EMBO journal*, 9(12): 4017–26 (1990).
- B. Mélykúti, E. August, A. Papachristodoulou, and H. El-Samad. Discriminating between rival biochemical network models: three approaches to optimal experiment design. *BMC systems biology*, 4(1): 38 (2010).
- C. D. Meyer. *Matrix analysis and applied linear algebra*. SIAM (2000).
- C. Michalik, M. Stuckert, and W. Marquardt. Optimal Experimental Design for Discriminating Numerous Model Candidates: The AWDC Criterion. *Industrial & Engineering Chemistry Research*, 49(2): 913–919 (2010).
- A. Mogilner, R. Wollman, and W. F. Marshall. Quantitative modeling in cell biology: what is it good for? *Developmental cell*, 11(3): 279–87 (2006).
- L. Molina, L. Belanche, and A. Nebot. Feature selection algorithms: a survey and experimental evaluation. In *2002 IEEE International Conference on Data Mining, 2002. Proceedings.*, 306–313. IEEE Comput. Soc (2002).
- V. R. Nadador, A. Ben-Zvi, and S. L. Shah. Inferring Gene Networks using Robust Statistical Techniques. *Statistical Applications in Genetics and Molecular Biology*, 10(1) (2011).
- K. Nalley, S. A. Johnston, and T. Kodadek. Proteolytic turnover of the Gal4 transcription factor is not required for function in vivo. *Nature*, 442(7106): 1054–7 (2006).
- R. Nash, S. Weng, B. Hitz, R. Balakrishnan, K. R. Christie, M. C. Costanzo, S. S. Dwight, S. R. Engel, D. G. Fisk, J. E. Hirschman, E. L. Hong, M. S. Livstone, R. Oughtred, J. Park, M. Skrzypek, C. L. Theesfeld, G. Binkley, Q. Dong, C. Lane, S. Miyasato, A. Sethuraman, M. Schroeder, K. Dolinski, D. Botstein, and J. M. Cherry. Expanded protein information at SGD: new pages and proteome browser. *Nucleic acids research*, 35(Database issue): D468–71 (2007).
- S. G. Nash and A. Sofer. *Linear and nonlinear programming*. McCraw-Hill, Singapore, internatio edition (1996).
- T. O. Nielsen, R. B. West, S. C. Linn, O. Alter, M. A. Knowling, J. X. O’Connell, S. Zhu, M. Fero, G. Sherlock, J. R. Pollack, P. O. Brown, D. Botstein, and M. van de Rijn. Molecular characterisation of soft tissue tumours: a gene expression study. *Lancet*, 359(9314): 1301–1307 (2002).
- B. Ninness and G. C. Goodwin. Estimation of model quality. *Automatica*, 31(12): 1771–1797 (1995).

- T. E. M. Nordling, N. Hiroi, and A. Funahashi. Dynamic modes in biological systems: Why should a biologist care? The 8th International Conference on Systems Biology (ICSB-2007), Long Beach (U.S.A.) (2007a).
- T. E. M. Nordling, N. Hiroi, A. Funahashi, and H. Kitano. Deduction of intracellular sub-systems from a topological description of the network. *Molecular bioSystems*, 3(8): 523–9 (2007b).
- T. E. M. Nordling, N. Hiroi, A. Funahashi, and H. Kitano. Detection of functional modules in graphs depend on a linearization of the biochemical network. In *The 15th Annual International Conference on Intelligent Systems for Molecular Biology (ISMB) & 6th European Conference on Computational Biology (ECCB) in Vienna (Austria): Abstract book*. International Society for Computational Biology, Vienna, Austria (2007c).
- T. E. M. Nordling and E. W. Jacobsen. Experiment design for optimal excitation of gene regulatory networks. In *The 7th International Conference on Systems Biology (ICSB-2006) in Yokohama (Japan): Abstract book*. The Systems Biology Institute (SBI) (2006a).
- T. E. M. Nordling and E. W. Jacobsen. Experiment design for systematic excitation of gene regulatory networks. International Workshop on Systems Biology in Maynooth (Ireland) (2006b).
- T. E. M. Nordling and E. W. Jacobsen. Iterative experimental design and identification of gene regulatory networks. In *The 13th Nordic Process Control Workshop (NPCW'06) in Lyngby (Denmark): Abstract book*, 27. Department of Chemical Engineering, Technical University of Denmark, Lyngby, Denmark (2006c).
- T. E. M. Nordling and E. W. Jacobsen. Experiment Design for Proper Excitation of Gene Regulatory Networks. In *Proceedings of Foundations of Systems Biology in Engineering (FOSBE), 2nd Conference*. Fraunhofer IRB Verlag, Postfach 800469, 70504 Stuttgart, Germany (2007).
- T. E. M. Nordling and E. W. Jacobsen. Ill-conditioning - a property of bio-networks. The 2nd annual q-bio conference on cellular information, Santa Fe (U.S.A.) (2008a).
- T. E. M. Nordling and E. W. Jacobsen. Inference of interampatte gene regulatory networks - with application to apoptosis signalling. The 9th International Conference on Systems Biology (ICSB-2008), Gothenburg (Sweden) (2008b).
- T. E. M. Nordling and E. W. Jacobsen. Interampattiness—a generic property of biochemical networks. *IET Syst Biol*, 3(5): 388–403 (2009a).
- T. E. M. Nordling and E. W. Jacobsen. Invalidating models of gene regulatory networks - the implications of characteristic and weak modes for network inference. Engineering Principles in Biological systems conference, Hinxton (UK) (2009b).

- T. E. M. Nordling and E. W. Jacobsen. Design of perturbations is the key to inference of tumour specific gene regulation. MGH-KI-Cell press Days of Molecular Medicine, Stockholm (Sweden) (2010a).
- T. E. M. Nordling and E. W. Jacobsen. Sparsity is a means and not an aim in inference of gene regulatory networks. In *The 11th International Conference on Systems Biology (ICSB-2010) in Edinburgh (UK): Abstract book*. Edinburgh, UK (2010b).
- T. E. M. Nordling and E. W. Jacobsen. On Sparsity As a Criterion in Reconstructing Biochemical Networks. In B. Sergio, A. Cenedese, and S. Zampieri, editors, *Proceedings of the 18th International Federation of Automatic Control (IFAC) World Congress, 2011*, 11672–11678. The International Federation of Automatic Control, Milano, Italy (2011).
- T. E. M. Nordling, J. Koljonen, J. T. Alander, and P. Geladi. Genetic Algorithms as a Tool for Wavelength Selection. In J. T. Alander, P. Ala-Siuru, and H. Hyötyniemi, editors, *Proceedings of the 11th Finnish Artificial Intelligence Conference (STeP 2004) in Vantaa (Finland), Volume 3*, volume 3, 99–113. Finnish Artificial Intelligence Society (FAIS), Vantaa, Finland (2004).
- T. E. M. Nordling, J. Koljonen, J. Nyström, I. Bodén, B. Lindholm-Sethson, P. Geladi, and J. T. Alander. Wavelength selection by genetic algorithms in near infrared spectra for melanoma diagnosis. In *IFMBE Proceedings, Volume 11, 3rd European Medical & Biological Engineering Conference (EMBECC'05) in Prague (Czech Republic)*. Prague, Czech Republic (2005).
- A. Olowoyeye and C. I. Okwundu. Gene therapy for sickle cell disease. *Cochrane database of systematic reviews (Online)*, (8): CD007652 (2010).
- L. Omberg, G. H. Golub, and O. Alter. A tensor higher-order singular value decomposition for integrative analysis of DNA microarray data from different studies. *Proc Natl Acad Sci U S A*, 104(47): 18371–18376 (2007).
- S. H. Orkin and D. R. Higgs. Medicine. Sickle cell disease at 100 years. *Science (New York, N. Y.)*, 329(5989): 291–2 (2010).
- P. D. Owen. General-to-Specific Modelling Using PcGets. *Journal of Economic Surveys*, 17(4): 609–628 (2003).
- K. Pearson. X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine Series 5*, 50(302): 157–175 (1900).
- J. Peccoud and C. Jacob. Theoretical uncertainty of measurements using quantitative polymerase chain reaction. *Biophysical journal*, 71(1): 101–8 (1996).

- C. A. Penfold and D. L. Wild. How to infer gene networks from expression profiles, revisited. *Interface Focus*, 1(6): 857–870 (2011).
- H. Peng, F. Long, and C. Ding. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE transactions on pattern analysis and machine intelligence*, 27(8): 1226–38 (2005).
- C. L. Peterson and J. L. Workman. Promoter targeting and chromatin remodeling by the SWI/SNF complex. *Current opinion in genetics & development*, 10(2): 187–92 (2000).
- M. Piazza, X.-J. Feng, J. D. Rabinowitz, and H. Rabitz. Diverse metabolic model parameters generate similar methionine cycle dynamics. *Journal of theoretical biology*, 251(4): 628–39 (2008).
- A. A. Prinz, D. Bucher, and E. Marder. Similar network activity from disparate circuit parameters. *Nature neuroscience*, 7(12): 1345–52 (2004).
- L. Pronzato. Optimal experimental design and some related control problems. *Automatica*, 44(2): 303–325 (2008).
- J. Quackenbush. Computational analysis of microarray data. *Nat Rev Genet*, 2(6): 418–427 (2001).
- J. Quackenbush. Computational approaches to analysis of DNA microarray data. *Yearb Med Inform*, 45 Suppl 1: 91–103 (2006).
- C. Rangel, J. Angus, Z. Ghahramani, M. Lioumi, E. Sothoran, A. Gaiba, D. L. Wild, and F. Falciani. Modeling T-cell activation using gene expression profiling and state-space models. *Bioinformatics (Oxford, England)*, 20(9): 1361–72 (2004).
- C. R. Rao and H. Toutenburg. *Linear models: least squares and alternatives*. Springer-Verlag New York, Inc., 175 Fifth Avenue, New York, USA, 2nd edition (1999).
- J. Reeb and S. Leavengood. Using the graphical method to solve linear programs. *Performance excellence in the wood products industry*, (October) (1998).
- W. Reinelt, A. Garulli, and L. Ljung. Comparing different approaches to model error modeling in robust identification. *Automatica*, 38(5): 787–803 (2002).
- B. Ren, F. Robert, J. J. Wyrick, O. Aparicio, E. G. Jennings, I. Simon, J. Zeitlinger, J. Schreiber, N. Hannett, E. Kanin, T. L. Volkert, C. J. Wilson, S. P. Bell, and R. A. Young. Genome-wide location and function of DNA binding proteins. *Science (New York, N.Y.)*, 290(5500): 2306–9 (2000).

- M. A. J. Roberts, E. August, A. Hamadeh, P. K. Maini, P. E. McSharry, J. P. Armitage, and A. Papachristodoulou. A model invalidation-based approach for elucidating biological signalling pathways, applied to the chemotaxis pathway in *R. sphaeroides*. *BMC systems biology*, 3(1): 105 (2009).
- K. A. Robinson and J. M. Lopes. SURVEY AND SUMMARY: *Saccharomyces cerevisiae* basic helix-loop-helix proteins regulate diverse biological processes. *Nucleic acids research*, 28(7): 1499–505 (2000).
- D. H. Roukos. Novel clinico-genome network modeling for revolutionizing genotype-phenotype-based personalized cancer care. *Expert review of molecular diagnostics*, 10(1): 33–48 (2010a).
- D. H. Roukos. Systems medicine: a real approach for future personalized oncology? *Pharmacogenomics*, 11(3): 283–7 (2010b).
- D. H. Roukos. Trastuzumab and beyond: sequencing cancer genomes and predicting molecular networks. *The pharmacogenomics journal*, 11(2): 81–92 (2011).
- O. Roy and M. Vetterli. The Effective Rank: A Measure of Effective Dimensionality. In *European signal processing conference (EUSIPCO)*, 606–610 (2007).
- J. M. Ruijter, C. Ramakers, W. M. H. Hoogaars, Y. Karlen, O. Bakker, M. J. B. van den Hoff, and A. F. M. Moorman. Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. *Nucleic acids research*, 37(6): e45 (2009).
- R. G. Rutledge and D. Stewart. Critical evaluation of methods used to determine amplification efficiency refutes the exponential character of real-time PCR. *BMC molecular biology*, 9: 96 (2008).
- R. G. Rutledge and D. Stewart. Assessing the performance capabilities of LRE-based assays for absolute quantitative real-time PCR. *PloS one*, 5(3): e9731 (2010).
- Y. Saeys, I. n. Inza, and P. Larrañaga. A review of feature selection techniques in bioinformatics. *Bioinformatics (Oxford, England)*, 23(19): 2507–17 (2007).
- M. G. Safonov. Stability margins of diagonally perturbed multivariable feedback systems. *Control Theory and Applications, IEE Proceedings D*, 129(6): 251–256 (1982).
- R. K. Saiki, S. Scharf, F. Faloona, K. B. Mullis, G. T. Horn, H. A. Erlich, and N. Arnheim. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science (New York, N.Y.)*, 230(4732): 1350–4 (1985).
- Y. Sako. Imaging single molecules in living cells for systems biology. *Molecular systems biology*, 2: 56 (2006).

- Y. Sako, S. Minoghchi, and T. Yanagida. Single-molecule imaging of EGFR signalling on the surface of living cells. *Nature cell biology*, 2(3): 168–72 (2000).
- H. M. Sauro and B. N. Kholodenko. Quantitative analysis of signaling networks. *Prog Biophys Mol Biol*, 86(1): 5–43 (2004).
- M. A. Savageau. Biochemical systems analysis—II. The steady-state solutions for an n-pool system using a power-law approximation. *Journal of Theoretical Biology*, 25(3): 370–379 (1969).
- M. A. Savageau. *Biochemical systems analysis: a study of function and design in molecular biology*. Addison-Wesley, Reading, MA (1976).
- M. A. Savageau. Recasting nonlinear differential equations as S-systems: a canonical nonlinear form. *Mathematical Biosciences*, 87(1): 83–115 (1987).
- J. Schäfer and K. Strimmer. An empirical Bayes approach to inferring large-scale gene association networks. *Bioinformatics (Oxford, England)*, 21(6): 754–64 (2005).
- J. H. Schefe, K. E. Lehmann, I. R. Buschmann, T. Unger, and H. Funke-Kaiser. Quantitative real-time RT-PCR data analysis: current concepts and the novel "gene expression's C ( T ) difference" formula. *J Mol Med*, 84(11): 901–910 (2006).
- M. Schena, D. Shalon, R. W. Davis, and P. O. Brown. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, 270(5235): 467–470 (1995).
- H. Schmidt, K. H. Cho, and E. W. Jacobsen. Identification of small scale biochemical networks based on general type system perturbations. *FEBS J*, 272(9): 2141–2151 (2005).
- T. D. Schmittgen and K. J. Livak. Analyzing real-time PCR data by the comparative CT method. *Nature Protocols*, 3(6): 1101–1108 (2008).
- G. J. Schütz, G. Kada, V. P. Pastushenko, and H. Schindler. Properties of lipid microdomains in a muscle cell membrane visualized by single molecule microscopy. *The EMBO journal*, 19(5): 892–901 (2000).
- M. Schwaab, J. L. Monteiro, and J. C. Pinto. Sequential experimental design for model discrimination Taking into account the posterior covariance matrix of differences between model predictions. *Chemical Engineering Science*, 63(9): 2408–2419 (2008).
- M. Schwaab, F. Silva, C. Queipo, A. J. Barreto, M. Nele, and J. Pinto. A new approach for sequential experimental design for model discrimination. *Chemical Engineering Science*, 61(17): 5791–5806 (2006).

- G. Schwarz. Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2): 461–464 (1978).
- H. Schwender, K. Ickstadt, and J. Rahnenführer. Classification with high-dimensional genetic data: assigning patients and genetic features to known classes. *Biometrical journal. Biometrische Zeitschrift*, 50(6): 911–26 (2008).
- C. A. Sellick, R. N. Campbell, and R. J. Reece. Galactose metabolism in yeast-structure and regulation of the leloir pathway enzymes and the genes encoding them. *International review of cell and molecular biology*, 269: 111–150 (2008).
- P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*, 13(11): 2498–2504 (2003).
- Z. Shen, N. Paquin, A. Forget, and P. Chartrand. Nuclear shuttling of She2p couples ASH1 mRNA localization to its translational repression by recruiting Loc1p and Puf6p. *Molecular biology of the cell*, 20(8): 2265–75 (2009).
- D. Skanda and D. Lebiedz. An optimal experimental design approach to model discrimination in dynamic biochemical systems. *Bioinformatics (Oxford, England)*, 26(7): 939–45 (2010).
- S. Skogestad and I. Postlethwaite. *Multivariable feedback control*. John Wiley & Sons, Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England, 1st edition (1996).
- D. K. Slonim. From patterns to pathways: gene expression data analysis comes of age. *Nat Genet*, 32 Suppl: 502–508 (2002).
- S. Smit, H. C. J. Hoefsloot, and A. K. Smilde. Statistical data processing in clinical proteomics. *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences*, 866(1-2): 77–88 (2008).
- T. Söderström. Errors-in-variables methods in system identification. *Automatica*, 43(6): 939–958 (2007).
- E. Sontag. Molecular Systems Biology and Control. *European Journal of Control*, 11(4-5): 396–435 (2005).
- P. T. Spellman, G. Sherlock, M. Q. Zhang, V. R. Iyer, K. Anders, M. B. Eisen, P. O. Brown, D. Botstein, and B. Futcher. Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol Biol Cell*, 9(12): 3273–3297 (1998).

- C. Stark, B.-J. Breitkreutz, A. Chatr-Aryamontri, L. Boucher, R. Oughtred, M. S. Livstone, J. Nixon, K. Van Auken, X. Wang, X. Shi, T. Reguly, J. M. Rust, A. Winter, K. Dolinski, and M. Tyers. The BioGRID Interaction Database: 2011 update. *Nucleic acids research*, 39(Database issue): D698–704 (2011).
- J. P. Steibel, R. Poletto, P. M. Coussens, and G. J. M. Rosa. A powerful and flexible linear mixed model framework for the analysis of relative quantification RT-PCR data. *Genomics*, 94(2): 146–52 (2009).
- D. M. Steinberg and W. G. Hunter. Experimental-design - review and comment. *Technometrics*, 26(2): 71–97 (1984).
- F. Steinke, M. Seeger, and K. Tsuda. Experimental design for efficient identification of gene regulatory networks using sparse Bayesian models. *BMC Syst Biol*, 1(1): 51 (2007).
- G. W. Stewart. Collinearity and Least Squares Regression. *Statistical Science*, 2(1): 68–84 (1987).
- G. W. Stewart. Perturbation Theory for the Singular Value Decomposition. *SVD and Signal Processing, II: Algorithms, Analysis and Applications*, 99–109 (1990).
- P. Stoica and Y. Selen. Model-order selection: a review of information criterion rules. *IEEE Signal Processing Magazine*, 21(4): 36–47 (2004).
- G. Stolovitzky, P. Kahlem, and A. Califano. The Challenges of Systems Biology: Community Efforts to Harness Biological Complexity. *Annals of the New York Academy of Sciences*, 1158 (2009).
- M. P. Styczynski and G. Stephanopoulos. Overview of computational methods for the inference of gene regulatory networks. *Computers & Chemical Engineering*, 29(3): 519–534 (2005).
- O. Suslov and D. A. Steindler. PCR inhibition by reverse transcriptase leads to an overestimation of amplification efficiency. *Nucleic Acids Res*, 33(20) (2005).
- T. Suzuki-Fujimoto, M. Fukuma, K. I. Yano, H. Sakurai, A. Vonika, S. A. Johnston, and T. Fukasawa. Analysis of the galactose signal transduction pathway in *Saccharomyces cerevisiae*: interaction between Gal3p and Gal80p. *Mol Cell Biol*, 16(5): 2504–2508 (1996a).
- T. Suzuki-Fujimoto, M. Fukuma, K. I. Yano, H. Sakurai, A. Vonika, S. A. Johnston, and T. Fukasawa. Analysis of the galactose signal transduction pathway in *Saccharomyces cerevisiae*: interaction between Gal3p and Gal80p. *Molecular and cellular biology*, 16(5): 2504–8 (1996b).
- M. S. Szalay, I. A. Kovács, T. Korcsmáros, C. Böde, and P. Csermely. Stress-induced rearrangements of cellular networks: Consequences for protection and drug design. *FEBS Lett* (2007).

- J. Tegnér and J. Björkegren. Perturbations to uncover gene networks. *Trends Genet*, 23(1): 34–41 (2007).
- J. Tegnér, M. K. Yeung, J. Hasty, and J. J. Collins. Reverse engineering gene networks: integrating genetic perturbations with dynamical modeling. *Proc Natl Acad Sci U S A*, 100(10): 5944–5949 (2003).
- O. Thellin, B. ElMoualij, E. Heinen, and W. Zorzi. A decade of improvements in quantification of gene expression and internal standard selection. *Biotechnology Advances*, 27(4): 323–333 (2009).
- D. Thieffry, A. M. Huerta, E. Pérez-Rueda, and J. Collado-Vides. From specific gene regulation to genomic networks: a global analysis of transcriptional regulation in *Escherichia coli*. *BioEssays*, 20(5): 433–440 (1998).
- R. Tibshirani. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society*, 58(1): 267–288 (1996).
- A. Tichopad, M. Dilger, G. Schwarz, and M. W. Pfaffl. Standardized determination of real-time PCR efficiency from a single reaction set-up. *Nucleic Acids Res*, 31(20) (2003).
- A. Tjörnberg, T. E. M. Nordling, M. Studham, and E. L. L. Sonnhammer. Optimal sparsity criteria for network inference. *Journal of Computational Biology*, 20(5) (2013).
- H. L. Trentelman, A. A. Stoorvogel, and M. Hautus. *Control Theory for Linear Systems*. Springer (2001).
- N. G. Tsvetanova, D. M. Klass, J. Salzman, and P. O. Brown. Proteome-wide search reveals unexpected RNA-binding proteins in *Saccharomyces cerevisiae*. *PloS one*, 5(9) (2010).
- S. Vajda. Structural equivalence of linear systems and compartmental models. *Mathematical Biosciences*, 55(1-2): 39–64 (1981).
- E. P. van Someren, L. F. A. Wessels, E. Backer, and M. J. T. Reinders. Genetic network modeling. *Pharmacogenomics*, 3(4): 507–25 (2002).
- J. Vandesompele, K. De Preter, F. Pattyn, B. Poppe, N. Van Roy, A. De Paepe, and F. Speleman. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol*, 3(7) (2002).
- J. C. Venter, M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, M. Yandell, C. A. Evans, R. A. Holt, J. D. Gocayne, P. Amanatides, R. M. Ballew, D. H. Huson, J. R. Wortman, Q. Zhang, C. D. Kodira, X. H. Zheng, L. Chen, M. Skupski, G. Subramanian, P. D. Thomas, J. Zhang, G. L. Gabor Miklos, C. Nelson, S. Broder, A. G. Clark, J. Nadeau, V. A. McKusick,

- N. Zinder, A. J. Levine, R. J. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A. Delcher, I. Dew, D. Fasulo, M. Flanigan, L. Florea, A. Halpern, S. Hannenhalli, S. Kravitz, S. Levy, C. Mobarry, K. Reinert, K. Remington, J. Abu-Threideh, E. Beasley, K. Biddick, V. Bonazzi, R. Brandon, M. Cargill, I. Chandramouliswaran, R. Charlab, K. Chaturvedi, Z. Deng, V. Di Francesco, P. Dunn, K. Eilbeck, C. Evangelista, A. E. Gabrielian, W. Gan, W. Ge, F. Gong, Z. Gu, P. Guan, T. J. Heiman, M. E. Higgins, R. R. Ji, Z. Ke, K. A. Ketchum, Z. Lai, Y. Lei, Z. Li, J. Li, Y. Liang, X. Lin, F. Lu, G. V. Merkulov, N. Milshina, H. M. Moore, A. K. Naik, V. A. Narayan, B. Neelam, D. Nusskern, D. B. Rusch, S. Salzberg, W. Shao, B. Shue, J. Sun, Z. Wang, A. Wang, X. Wang, J. Wang, M. Wei, R. Wides, C. Xiao, C. Yan, A. Yao, J. Ye, M. Zhan, W. Zhang, H. Zhang, Q. Zhao, L. Zheng, F. Zhong, W. Zhong, S. Zhu, S. Zhao, D. Gilbert, S. Baumhueter, G. Spier, C. Carter, A. Cravchik, T. Woodage, F. Ali, H. An, A. Awe, D. Baldwin, H. Baden, M. Barnstead, I. Barrow, K. Beeson, D. Busam, A. Carver, A. Center, M. L. Cheng, L. Curry, S. Danaher, L. Davenport, R. Desilets, S. Dietz, K. Dodson, L. Doup, S. Ferriera, N. Garg, A. Gluecksmann, B. Hart, J. Haynes, C. Haynes, C. Heiner, S. Hladun, D. Hostin, J. Houck, T. Howland, C. Ibegwam, J. Johnson, F. Kalush, L. Kline, S. Koduru, A. Love, F. Mann, D. May, S. McCawley, T. McIntosh, I. McMullen, M. Moy, L. Moy, B. Murphy, K. Nelson, C. Pfannkoch, E. Pratts, V. Puri, H. Qureshi, M. Rardon, R. Rodriguez, Y. H. Rogers, D. Romblad, B. Ruhfel, R. Scott, C. Sitter, M. Smallwood, E. Stewart, R. Strong, E. Suh, R. Thomas, N. N. Tint, S. Tse, C. Vech, G. Wang, J. Wetter, S. Williams, M. Williams, S. Windsor, E. Winn-Deen, K. Wolfe, J. Zaveri, K. Zaveri, J. F. Abril, R. Guigo, M. J. Campbell, K. V. Sjolander, B. Karlak, A. Kejariwal, H. Mi, B. Lazareva, T. Hatton, A. Narechania, K. Diemer, A. Muruganujan, N. Guo, S. Sato, V. Bafna, S. Istrail, R. Lippert, R. Schwartz, B. Walenz, S. Yooseph, D. Allen, A. Basu, J. Baxendale, L. Blick, M. Caminha, J. Carnes-Stine, P. Caulk, Y. H. Chiang, M. Coyne, C. Dahlke, A. Mays, M. Dombroski, M. Donnelly, D. Ely, S. Esparham, C. Fosler, H. Gire, S. Glanowski, K. Glasser, A. Glodek, M. Gorokhov, K. Graham, B. Gropman, M. Harris, J. Heil, S. Henderson, J. Hoover, D. Jennings, C. Jordan, J. Jordan, J. Kasha, L. Kagan, C. Kraft, A. Levitsky, M. Lewis, X. Liu, J. Lopez, D. Ma, W. Majoros, J. McDaniel, S. Murphy, M. Newman, T. Nguyen, N. Nguyen, M. Nodell, S. Pan, J. Peck, M. Peterson, W. Rowe, R. Sanders, J. Scott, M. Simpson, T. Smith, A. Sprague, T. Stockwell, R. Turner, E. Venter, M. Wang, M. Wen, D. Wu, M. Wu, A. Xia, A. Zandieh, and X. Zhu. The sequence of the human genome. *Science*, 291(5507): 1304–1351 (2001).
- R. Visintin, K. Craig, E. S. Hwang, S. Prinz, M. Tyers, and A. Amon. The phosphatase Cdc14 triggers mitotic exit by reversal of Cdk-dependent phosphorylation. *Molecular cell*, 2(6): 709–18 (1998).
- E. Voit. *Computational Analysis of Biochemical Systems : A Practical Guide for Biochemists and Molecular Biologists*. Cambridge University Press, Cambridge, UK (2000).

- L. von Bertalanffy. *General system theory: foundations, development, applications*. Braziller, New York, rev. ed., edition (2006).
- G. von Dassow, E. Meir, E. M. Munro, and G. M. Odell. The segment polarity network is a robust developmental module. *Nature*, 406(6792): 188–92 (2000).
- F. O. Walker. Huntington’s disease. *Lancet*, 369(9557): 218–28 (2007).
- Y. Wang, D. J. Miller, and R. Clarke. Approaches to working in high-dimensional data spaces: gene expression microarrays. *British journal of cancer*, 98(6): 1023–8 (2008).
- J. D. Watson and F. H. Crick. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, 171(4356): 737–738 (1953).
- J. C. Way and P. A. Silver. Systems engineering without an engineer: Why we need systems biology. *Complexity*, 13(2): 22–29 (2007).
- S. Weng, Q. Dong, R. Balakrishnan, K. Christie, M. Costanzo, K. Dolinski, S. S. Dwight, S. Engel, D. G. Fisk, E. Hong, L. Issel-Tarver, A. Sethuraman, C. Theesfeld, R. Andrada, G. Binkley, C. Lane, M. Schroeder, D. Botstein, and J. Michael Cherry. Saccharomyces Genome Database (SGD) provides biochemical and structural information for budding yeast proteins. *Nucleic acids research*, 31(1): 216–8 (2003).
- D. B. Wheeler, A. E. Carpenter, and D. M. Sabatini. Cell microarrays and RNA interference chip away at gene function. *Nature genetics*, 37 Suppl: S25–30 (2005).
- J. Wildenhain and E. J. Crampin. Reconstructing gene regulatory networks: from random to scale-free connectivity. *Syst Biol (Stevenage)*, 153(4): 247–256 (2006).
- O. Wolkenhauer, D. Fell, P. De Meyts, N. Blüthgen, H. Herzel, N. Le Novère, T. Höfer, K. Schürrie, and I. van Leeuwen. SysBioMed report: advancing systems biology for medical applications. *IET Syst Biol*, 3(3): 131–136 (2009).
- O. Wolkenhauer, S. N. Sreenath, P. Wellstead, M. Ullah, and K. H. Cho. A systems- and signal-oriented approach to intracellular dynamics. *Biochem Soc Trans*, 33(Pt 3): 507–515 (2005a).
- O. Wolkenhauer, M. Ullah, P. Wellstead, and K. H. Cho. The dynamic systems approach to control and regulation of intracellular networks. *FEBS Lett*, 579(8): 1846–1853 (2005b).
- World Health Organization. *The global burden of disease: 2004 update*. WHO Press, World Health Organization, 20 Avenue Appia, 1211 Geneva 27, Switzerland (2008).
- X. Wu and T. G. Dewey. From microarray to biological networks: Analysis of gene expression profiles. *Methods Mol Biol*, 316: 35–48 (2006).

- Z. Wu and Z. Wu. Exploration, visualization, and preprocessing of high-dimensional data. *Methods in molecular biology (Clifton, N.J.)*, 620: 267–84 (2010).
- Z. Xiaobo, Z. Jiewen, M. J. W. Povey, M. Holmes, and M. Hanpin. Variables selection methods in near-infrared spectroscopy. *Analytica chimica acta*, 667(1-2): 14–32 (2010).
- Y. Yang. Can the strengths of AIC and BIC be shared? A conflict between model identification and regression estimation. *Biometrika*, 92(4): 937–950 (2005).
- Y. Yang. Prediction estimation with simple linear models: is it really that simple? *Econometric Theory*, 23(01): 1–36 (2007).
- M. K. Yeung, J. Tegnér, and J. J. Collins. Reverse engineering gene networks using singular value decomposition and robust regression. *Proc Natl Acad Sci U S A*, 99(9): 6163–6168 (2002).
- T. M. Yi, Y. Huang, M. I. Simon, and J. Doyle. Robust perfect adaptation in bacterial chemotaxis through integral feedback control. *Proc Natl Acad Sci U S A*, 97(9): 4649–4653 (2000).
- C. Yoo and G. F. Cooper. A computer-based microarray experiment design-system for gene-regulation pathway discovery. *AMIA ... Annual Symposium proceedings / AMIA Symposium. AMIA Symposium*, 733–7 (2003).
- C. You, L. Yang, X.-Q. Cheng, and G. Li. Survey and taxonomy of feature selection algorithms in intrusion detection system. In *Information Security and Cryptology - Second SKLOIS Conference, Inscrypt 2006, Lecture notes in computer science*, 153–167. Springer (2006).
- G. A. Young and R. L. Smith. *Essentials of statistical inference*. Cambridge University Press, Cambridge CB2 2RU, UK (2005).
- J. Yu, V. A. Smith, P. P. Wang, A. J. Hartemink, and E. D. Jarvis. Advances to Bayesian network inference for generating causal networks from observational biological data. *Bioinformatics (Oxford, England)*, 20(18): 3594–603 (2004).
- Y. Yuan, G.-B. Stan, S. Warnick, and J. Goncalves. Robust dynamical network structure reconstruction. *Automatica*, 47(6): 1230–1235 (2011).
- S. C. Yusta. Different metaheuristic strategies to solve the feature selection problem. *Pattern Recognition Letters*, 30(5): 525–534 (2009).
- D. E. Zak, G. E. Gonye, J. S. Schwaber, and F. J. Doyle. Importance of input perturbations and stochastic gene expression in the reverse engineering of genetic regulatory networks: insights from an identifiability analysis of an in silico network. *Genome research*, 13(11): 2396–405 (2003).

- J. Zeng, S. Zhu, and H. Yan. Towards accurate human promoter recognition: a review of currently used sequence features and classification methods. *Briefings in bioinformatics*, 10(5): 498–508 (2009).
- L.-Q. Zhang, J. C. Collins, and P. H. King. Indistinguishability and identifiability analysis of linear compartmental models. *Mathematical Biosciences*, 103(1): 77–95 (1991).
- P. Zhao and B. Yu. On model selection consistency of Lasso. *The Journal of Machine Learning Research*, 7: 2541–2563 (2006).
- K. Zhou and J. C. Doyle. *Essentials of robust control*. Prentice Hall (1998).
- K. Zhou, J. C. Doyle, and K. Glover. *Robust and optimal control*. Prentice Hall, Englewood Cliffs, N.J., U.S.A. (1996).
- H. Zhu, M. Bilgin, R. Bangham, D. Hall, A. Casamayor, P. Bertone, N. Lan, R. Jansen, S. Bidlingmaier, T. Houfek, T. Mitchell, P. Miller, R. A. Dean, M. Gerstein, and M. Snyder. Global analysis of protein activities using proteome chips. *Science (New York, N.Y.)*, 293(5537): 2101–5 (2001).

“It’s almost as if a demon might have passed from one host to another.”

John Forbes Nash, Jr.