Frequency Tracking for Speed Estimation

Martin Lindfors
This is a Swedish Licentiate’s Thesis.

Swedish postgraduate education leads to a Doctor’s degree and/or a Licentiate’s degree.
A Doctor’s Degree comprises 240 ECTS credits (4 years of full-time studies).
A Licentiate’s degree comprises 120 ECTS credits,
of which at least 60 ECTS credits constitute a Licentiate’s thesis.
To my family
Abstract

Estimating the frequency of a periodic signal, or tracking the time-varying frequency of an almost periodic signal, is an important problem that is well studied in literature. This thesis focuses on two subproblems where contributions can be made to the existing theory: frequency tracking methods and measurements containing outliers.

Maximum-likelihood-based frequency estimation methods are studied, focusing on methods which can handle outliers in the measurements. Katkovnik’s frequency estimation method is generalized to real and harmonic signals, and a new method based on expectation-maximization is proposed. The methods are compared in a simulation study in which the measurements contain outliers. The proposed methods are compared with the standard periodogram method.

Recursive Bayesian methods for frequency tracking are studied, focusing on the Rao-Blackwellized point mass filter (RBPMF). Two reformulations of the RBPMF aiming to reduce computational costs are proposed. Furthermore, the technique of variational approximate Rao-Blackwellization is proposed, which allows usage of a Student’s $t$ distributed measurement noise model. This enables recursive frequency tracking methods to handle outliers using heavy-tailed noise models in Rao-Blackwellized filters such as the RBPMF. A simulation study illustrates the performance of the methods when outliers occur in the measurement noise.

The framework above is applied to and studied in detail in two applications. The first application is on frequency tracking of engine sound. Microphone measurements are used to track the frequency of Doppler-shifted variants of the engine sound of a vehicle moving through an area. These estimates can be used to compute the speed of the vehicle. Periodogram-based methods and the RBPMF are evaluated on simulated and experimental data. The results indicate that the RBPMF has lower RMSE than periodogram-based methods when tracking fast changes in the frequency.

The second application relates to frequency tracking of wheel vibrations, where a car has been equipped with an accelerometer. The accelerometer measurements are used to track the frequency of the wheel axle vibrations, which relates to the wheel rotational speed. The velocity of the vehicle can then be estimated without any other sensors and without requiring integration of the accelerometer measurements. In situations with high signal-to-noise ratio (SNR), the methods perform well. To remedy situations when the methods perform poorly, an accelerometer input is introduced to the formulation. This input is used to predict changes in the frequency for short time intervals.
Populärvetenskaplig sammanfattning


Acknowledgments

I will now take the opportunity to thank the people who have enabled me to get to this point in my studies. Without their support, direct or indirect, this thesis could not have been written.

Naturally, I will start by thanking my supervisors: Rickard Karlsson, Gustaf Hendeby, and Fredrik Gustafsson. Rickard, thank you for your dedication in supervision, and for your strong knowledge in sensor fusion, both in theory and in relation to applications. Gustaf, thank you for your equally thorough knowledge and your attention to detail; in particular when it comes to \LaTeX. Fredrik, your visionary research approach and theoretical creativity is inspiring.

I would like to thank Svante Gunnarsson, the former head of division at Automatic Control, and Martin Enqvist, the current head of division. You have been sources of motivation and inspiration in challenging times. Thanks, too, to Ninna Stensgård for your administrative care.

Rickard Karlsson, Gustaf Hendeby, Svante Gunnarsson, Christian A. Naesseth, Tohid Ardeshiri, Kristoffer Bergman, Andreas Bergström, Per Boström-Rost, Du Ho Duc, Angela Fontan, Erik Hedberg, and Fredrik Ljungberg, have proofread all or parts of this thesis, and I am very grateful for it. Any remaining errors are my own. And yes. There will be cake.

\textit{NIRA Dynamics AB} and \textit{FOI} supplied data used in this thesis, for which I am thankful. This work was partially supported by the Wallenberg AI, Autonomous Systems and Software Program (\textit{wasp}) funded by the Knut and Alice Wallenberg Foundation. Their funding is gratefully acknowledged.

Thanks to my officemates. André Carvalho Bittencourt, thanks for answering all my weird questions when I started my PhD studies. George Mathai, I truly enjoyed our philosophical discussions. Niclas Evestedt, thanks for being an excellent co-driver in the Bay Area. Rafael Rui, thanks for the mornings in Campushallen, and please send some brigadeiros! And finally, Du Ho Duc, the trip to Vietnam was amazing.

The group at Automatic Control is truly great, and I appreciate being part of it. Thanks for the fika breaks, Christian A. Naesseth, Tohid Ardeshiri, Bram Dil, Inger Erlander-Klein, Anders Hansson, Erik Hedberg, Anders Helmersson, Ylva Jung, Sina Khoshfetra-Pakazad, Roger Larsson, Jonas Linder, Gustav Lindmark, Hanna Nyqvist, Zoran Sjanić, and Clas Veibäck, just to name a few. You know, if I would accept defeat in discussions more often, it would be less satisfying to you when I do!

In the academic world outside of Automatic Control, I genuinely appreciate the company and effort of all \textit{wasp} students who have participated in conferences and trips, and the \textit{wasp} seniors who have arranged them. Thanks to the \textit{MLRG} crew; our meetings gave much insight into machine learning theory.

Finally, thanks to my friends outside of academia. And of course, to my family. Your support means a lot.

\textit{Linköping, June 2018}

\textit{Martin Lindfors}
Contents

1 Introduction ................................................................. 1
   1.1 Applications ........................................................ 2
   1.2 Frequency Estimation and Tracking ................................ 3
   1.3 Publications and Contributions ................................... 5
   1.4 Outline .................................................................. 6

2 Theoretical Background ................................................... 7
   2.1 Probability Theory .................................................... 7
       2.1.1 Random Variables and Probability Density Functions .... 7
       2.1.2 Expectations .................................................... 8
       2.1.3 Jensen’s Inequality .............................................. 9
       2.1.4 Kullback-Leibler Divergence ................................. 9
       2.1.5 List of Probability Distributions ......................... 9
   2.2 Statistical Models ..................................................... 11
       2.2.1 State-Space Models ........................................... 12
       2.2.2 Conditionally Linear State-Space Models ................. 12
   2.3 Frequentist Inference ................................................ 13
       2.3.1 Maximum Likelihood Estimation ............................ 13
       2.3.2 Expectation Maximization ................................... 14
   2.4 Bayesian Inference .................................................... 15
       2.4.1 Bayesian Point Estimation ................................... 16
       2.4.2 Discretization of Bayesian Posteriors ..................... 16
       2.4.3 Rao-Blackwellization in Discretized Approximations ... 17
       2.4.4 Variational Bayes .............................................. 18
       2.4.5 Bayesian Filtering and Smoothing ......................... 20
   2.5 Robust Estimation ...................................................... 27

3 Maximum Likelihood Frequency Estimation ......................... 29
   3.1 Signal Model .......................................................... 29
   3.2 Frequency Estimation with Gaussian Noise Model ............... 31
       3.2.1 Maximum Likelihood Estimation ............................ 31
       3.2.2 Periodogram Estimation ..................................... 32
   3.3 Frequency Estimation with Non-Gaussian Noise Model .......... 34
## Notation

### Variables

<table>
<thead>
<tr>
<th>Notation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x$</td>
<td>Random variable</td>
</tr>
<tr>
<td>$Y$</td>
<td>Set or vector of measurements</td>
</tr>
<tr>
<td>$\theta$</td>
<td>Set or vector of parameters</td>
</tr>
<tr>
<td>$k$</td>
<td>Time index</td>
</tr>
<tr>
<td>$y_k$</td>
<td>Measurement with index $k$</td>
</tr>
<tr>
<td>$Y_k$</td>
<td>Set of measurements with index less than or equal to $k$</td>
</tr>
<tr>
<td>$N$</td>
<td>Number of total measurements (thus, $1 \leq k \leq N$)</td>
</tr>
<tr>
<td>$\hat{\theta}$</td>
<td>Estimate of $\theta$</td>
</tr>
<tr>
<td>$N_S$</td>
<td>Number of samples</td>
</tr>
<tr>
<td>$x_k$</td>
<td>State</td>
</tr>
<tr>
<td>$x^a_k$</td>
<td>Subcomponent of state inferred analytically</td>
</tr>
<tr>
<td>$x^n_k$</td>
<td>Subcomponent of state inferred numerically</td>
</tr>
<tr>
<td>$X_k$</td>
<td>State history</td>
</tr>
<tr>
<td>$Z$</td>
<td>Latent variables</td>
</tr>
<tr>
<td>$\omega$</td>
<td>Angular frequency</td>
</tr>
<tr>
<td>$m$</td>
<td>Harmonic index</td>
</tr>
<tr>
<td>$M$</td>
<td>Number of harmonics</td>
</tr>
<tr>
<td>$\alpha^{(m)}, \beta^{(m)}$</td>
<td>Phasor parameters</td>
</tr>
<tr>
<td>$Z$</td>
<td>Vector of phasor parameters</td>
</tr>
<tr>
<td>$C$</td>
<td>Vector of complex phasor parameters</td>
</tr>
<tr>
<td>$\tilde{y}_k$</td>
<td>Complex-valued measurement</td>
</tr>
<tr>
<td>$\lambda_k$</td>
<td>Latent variable</td>
</tr>
<tr>
<td>$\Delta \omega$</td>
<td>Maximum change in angular frequency between batches</td>
</tr>
<tr>
<td>$\phi_k$</td>
<td>Phase</td>
</tr>
<tr>
<td>$B_k$</td>
<td>Amplitude</td>
</tr>
<tr>
<td>$\hat{x}, \tilde{x}, P, \tilde{P}$</td>
<td>Parameters of filtering distributions</td>
</tr>
<tr>
<td>$w$</td>
<td>Particle weights</td>
</tr>
<tr>
<td>$\Delta$</td>
<td>Propagation delay</td>
</tr>
<tr>
<td>$\dot{\Delta}$</td>
<td>Doppler shift</td>
</tr>
<tr>
<td>$\overline{\text{RMSE}}$</td>
<td>RMSE against estimated reference frequency</td>
</tr>
</tbody>
</table>
## Functions and Operators

<table>
<thead>
<tr>
<th>Notation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p(x), q(x)$</td>
<td>Probability density function (PDF)</td>
</tr>
<tr>
<td>$E_x, E_p(x)$</td>
<td>Expectation of $x$ under the PDF $p(x)$</td>
</tr>
<tr>
<td>Var $x$</td>
<td>Variance of $x$</td>
</tr>
<tr>
<td>Cov $x$</td>
<td>Covariance of $x$</td>
</tr>
<tr>
<td>$D_{KL}(q(x)\mid p(x))$</td>
<td>Kullback-Leibler divergence</td>
</tr>
<tr>
<td>$N(x\mid \mu, \sigma^2)$</td>
<td>Gaussian PDF</td>
</tr>
<tr>
<td>$L(x\mid \mu, \sigma)$</td>
<td>Laplace PDF</td>
</tr>
<tr>
<td>$T(x\mid \mu, \sigma^2, \nu)$</td>
<td>Student’s $t$ PDF</td>
</tr>
<tr>
<td>$O$</td>
<td>Asymptotic rate, big-O notation</td>
</tr>
<tr>
<td>$Q(\theta\mid \theta^{(j)})$</td>
<td>Q function in EM</td>
</tr>
<tr>
<td>$\mathcal{L}(q(x); p(x))$</td>
<td>Evidence lower bound</td>
</tr>
<tr>
<td>$\mathcal{Y}(\omega)$</td>
<td>Fourier transform of $Y$</td>
</tr>
<tr>
<td>$A_\omega, A(\omega)$</td>
<td>Vector or matrix of sines and cosines</td>
</tr>
<tr>
<td>$D(\omega), F(\omega)$</td>
<td>Frequency regression matrices</td>
</tr>
<tr>
<td>$L(x)$</td>
<td>Log-likelihood function</td>
</tr>
<tr>
<td>Re$(x), \text{Im}(x)$</td>
<td>Real and imaginary part of complex number $x$</td>
</tr>
<tr>
<td>$J(\omega, C)$</td>
<td>Cost function for generalized Katkovnik</td>
</tr>
</tbody>
</table>

## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDF</td>
<td>Probability density function</td>
</tr>
<tr>
<td>CDF</td>
<td>Cumulative density function</td>
</tr>
<tr>
<td>SSM</td>
<td>State-space model</td>
</tr>
<tr>
<td>CLSSM</td>
<td>Conditionally linear state-space model</td>
</tr>
<tr>
<td>EM</td>
<td>Expectation maximization</td>
</tr>
<tr>
<td>ML</td>
<td>Maximum likelihood</td>
</tr>
<tr>
<td>KF</td>
<td>Kalman filter</td>
</tr>
<tr>
<td>EKF</td>
<td>Extended Kalman filter</td>
</tr>
<tr>
<td>PF</td>
<td>Particle filter</td>
</tr>
<tr>
<td>PMF</td>
<td>Point mass filter</td>
</tr>
<tr>
<td>RBPF</td>
<td>Rao-Blackwellized particle filter</td>
</tr>
<tr>
<td>RBPMPF</td>
<td>Rao-Blackwellized point mass filter</td>
</tr>
<tr>
<td>SHS</td>
<td>Sub-harmonic summation</td>
</tr>
<tr>
<td>DPM</td>
<td>Discrete periodogram maximization</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal to noise ratio</td>
</tr>
<tr>
<td>FFT</td>
<td>Fast Fourier transform</td>
</tr>
<tr>
<td>ATV</td>
<td>All-terrain vehicle</td>
</tr>
<tr>
<td>GPS</td>
<td>Global Positioning System</td>
</tr>
<tr>
<td>RMSE</td>
<td>Root mean square error</td>
</tr>
</tbody>
</table>
Introduction

Periodic signals occur everywhere in the world. In situations when the frequency of such signals is unknown, it may be of interest to infer it from measurements. The applications studied in this thesis are related to speed estimation of vehicles. It is evident that vehicles emit periodic signals, such as engine sound and wheel vibrations. These periodic, harmonic signals (such as the one illustrated in Figure 1.1) can be measured by inexpensive sensors and used in order to estimate relevant quantities, such as a vehicle’s absolute or relative speed. To reduce costs and unit size, small and inexpensive sensors are attractive to use. However, the measurements from these sensors are often noisy. The measurements must be carefully modeled, and suitable methods must be used, in order to extract as much as possible of the information available in the signal.

Figure 1.1: To the left, an almost periodic signal in the time domain is shown for a short duration. The data is acoustic, from Application 1. To the right, the signal’s energy content in the frequency domain is shown for a longer duration using a spectrogram. Darker shades correspond to more energy at that frequency and time.
1.1 Applications

In order to apply the frequency tracking theory studied in this thesis, two applications are studied. The applications encompass different physical domains and present different demands, but it will be seen that a common framework of frequency tracking can be used for both of them. For the acoustic engine sound tracking application, Application 1, it can be useful to track fast changes in the frequency, while the mechanical vibration tracking application, Application 2, has measurement noise with outliers.

Application 1: Frequency Tracking of Engine Sound

The first application relates to acoustic frequency tracking (Lindfors et al., 2017). The considered scenario involves a wheeled vehicle traversing an area, as illustrated in Figure 1.2 and discussed by Lindgren et al. (2015) and Deleskog et al. (2014). Microphones are measuring the sound of the vehicle (mainly the engine sound) with different Doppler shifts and embedded in noise. Figure 1.1 shows these measurements including the spectrogram. Using frequency tracking, accurate information about the measured frequency of the engine sound at the different microphone locations can be acquired. The measured Doppler shifts can later be used to estimate the speed and position of the vehicle (Lindgren et al., 2015). This is useful for surveillance of unknown vehicles. In this work, the focus is on frequency tracking of the sound.

Application 2: Frequency Tracking of Wheel Vibrations

In the second application, accelerometer measurements are used to estimate the rotational speed of the vehicle wheels (Lindfors et al., 2016, 2018). The speed (or wheel speed) of a vehicle is usually estimated using wheel speed sensors (WSS) or satellite systems such as the Global Positioning System (GPS). Dead-reckoning of accelerometer measurements can also be used, but this suffers from increasing
1.2 Frequency Estimation and Tracking

The objective of frequency estimation and tracking is to infer the unknown fundamental frequency or frequencies of a periodic or almost periodic signal from noisy measurements of it.

A periodic signal or harmonic signal can be described as a sum of sinusoids, with the frequency of each individual sinusoid given by a multiple of the fundamental frequency. In general, each sinusoid has a different amplitude and phase offset. To infer the fundamental frequency (frequency estimation), all these parameters must be estimated or marginalized out Gelman et al. (2014).
An almost periodic signal can also be described by a sum of sinusoids, but with a slowly varying fundamental frequency. The derivative of the instant phase of each sinusoid is its instant frequency, and in order to make the signal harmonic, it is required that the instant frequency of each sinusoid is equal to a multiple of the slowly varying fundamental frequency of the signal. The amplitude and phase offset of each sinusoid are also allowed to slowly vary with time. Estimation of the parameters of such an almost periodic signal is called frequency tracking.

Background

The least squares solution of the frequency estimation problem, or equivalently the maximum likelihood solution with a Gaussian noise model, yields the well-known periodogram estimator (Quinn and Hannan, 2001). In order to reduce computational costs, a large number of variations of this estimator have been proposed, including methods based on, for example correlation methods (de Cheveigne and Kawahara, 2002; Stoica et al., 1994; Talkin, 1995) and subspace methods (Roy and Kailath, 1989; Schmidt, 1982). Frequency estimation methods using non-Gaussian noise models have been proposed by Katkovnik (1998, 1999), who uses an iterative least squares method to compute maximum-likelihood frequency estimates using any noise distribution model. However, Katkovnik’s papers assume a complex signal and only use a single harmonic. Alternative approaches are given by Li (2010) and Gurubelli et al. (2014), using specific heavy-tailed noise models. These references do not perform a comprehensive review of frequency tracking with heavy-tailed noise models.

One way to perform frequency tracking is to apply frequency estimation techniques to a sequence of short batches. This is called batch-oriented frequency tracking in this thesis; for more information on this subject, see, for instance, the papers by de Cheveigne and Kawahara (2002) and Hermes (1988) or the book by Quinn and Hannan (2001). On the other hand, recursive frequency tracking algorithms, as exemplified by Kumar (1987), Dubois and Davy (2007), Kim et al. (2008), and Ng et al. (2009), typically process one measurement at a time and update their estimates continuously. These methods use state-space models and apply recursive filtering techniques in order to infer the frequency. These methods may be computationally intensive, and they do not handle heavy-tailed measurement models.

Research Questions

This thesis studies batch and recursive frequency tracking methods in relation to Applications 1 and 2. It focuses on the following research questions:

- How can measurement outliers in frequency estimation and tracking be handled?
- How can the computational complexity of frequency estimation and tracking methods be reduced?
- Which frequency tracking methods are suitable for the applications?
1.3 Publications and Contributions

This section lists the publications related to this thesis that the author has participated in writing. It then discusses the contributions of the thesis.

The thesis author’s contribution to the publications below have been mathematical derivations, computer code implementations, method evaluation on data, and writing the main part of the manuscripts. The co-authors provided ideas, gave input into the process, provided initial code and data, and participated in writing the manuscripts.

Publication 1

This publication concerns Application 1.

Martin Lindfors, Gustaf Hendeby, Fredrik Gustafsson, and Rickard Karlsson.


Frequency tracking methods including the Rao-Blackwellized point mass filter (RBPMF, see Chapter 4) are proposed and evaluated on the engine sound frequency tracking scenario (Application 1), as detailed in Chapter 5. The evaluation is done on both experimental and simulated data. Parts of the content are reused in this thesis, courtesy of IEEE.

Publication 2

In the following publication, Application 2 is studied.

Martin Lindfors, Gustaf Hendeby, Fredrik Gustafsson, and Rickard Karlsson.


Frequency tracking methods including the RBPMF (see Chapter 4) are proposed and evaluated on the wheel vibration data described in Chapter 6. The manuscript is submitted for publication.

Publication 3

A subset of Publication 2 has already been published as:

Martin Lindfors, Gustaf Hendeby, Fredrik Gustafsson, and Rickard Karlsson.


This is an introductory study of Application 2. The RBPMF is proposed for frequency tracking of wheel vibrations, including the variant discussed in Section 4.2.1. The paper presents a preliminary version of the evaluations in Publication 2 and the discussion in Chapter 6. Parts of the content are reused in this thesis, courtesy of IEEE.
Contributions

The contributions of this thesis are:

- Methods for frequency estimation with measurements that contain outliers as detailed in Chapter 3. More specifically, a generalization of the existing frequency estimation method with heavy-tailed noise models by Katkovnik (1998, 1999), and an expectation-maximization-based frequency estimation method.

- Methods for Rao-Blackwellized filtering with reduced computational cost and variational approximate Rao-Blackwellization for heavy-tailed Student’s t measurement noise models, in Chapter 4. This relates to Publications 1–3.

- A study of frequency tracking of engine sound (Application 1), in Chapter 5. This relates to Publication 1.

- A study of frequency tracking of wheel speed vibrations (Application 2), in Chapter 6. This relates to Publications 2 and 3.

1.4 Outline

In Chapter 2, a theoretical overview of selected basic concepts in statistical inference and filtering is given. In Chapter 3, maximum likelihood-based frequency estimation algorithms are reviewed. Frequency estimation algorithms with non-Gaussian noise models are discussed in particular. It is also discussed how frequency estimation algorithms can be used for frequency tracking. In Chapter 4, another approach to frequency tracking, using recursive methods, is covered. State-space models for frequency tracking are reviewed, and one model is selected. Then, a family of filtering algorithms suitable for the problem is proposed. In Chapter 5, frequency tracking methods are evaluated using acoustic frequency tracking data as discussed in Application 1. In Chapter 6, frequency tracking methods are evaluated using data from Application 2, wheel vibration frequency tracking. Finally, in Chapter 7, the thesis is concluded and ideas for future work are discussed.
In this chapter, the theoretical background required for the forthcoming chapters is presented. The chapter starts with a review of probability theory and continues with model structures and statistical methods.

2.1 Probability Theory

This section explains selected concepts in probability theory which will be of use in this thesis. More details are available in books on probability theory and statistics, such as (Cramér, 1948; van Trees, 1968; Yates and Goodman, 2005).

2.1.1 Random Variables and Probability Density Functions

In this thesis, random variables will be modeled by probability density functions (PDFs). Given a random variable \( x \), the PDF \( p_x(\cdot) \) satisfies

\[
P(x \in R) = \int_R p_x(x) \, dx,
\]

where capital \( P(E) \) is the probability of the event \( E \) and \( R \) is a subset of the event space \( V_x \), which is the space where \( x \) takes values. Note the abuse of notation where \( x \) is both a random variable and an integration variable. This will be used in order to simplify the notation for the rest of the thesis. Similarly, the subscript \( x \) in \( p_x \) will often be omitted, so that the PDF is simply written as \( p \), leaving the variable \( x \) in question to be inferred from the argument of \( p(x) \).

Assume that \( x \) is a vector, and split it into components \( x = [x_1^T, x_2^T]^T \). The PDF of \( x \) can be written as a function of the components,

\[
p(x) = p(x_1, x_2).
\]
The marginal distribution of $x_1$, $p(x_1)$, is given by

$$p(x_1) = \int p(x_1, x_2) dx_2. \quad (2.3)$$

The domain of integration is often omitted whenever it is the full event space $V_x$. The conditional distribution of $x_1$ given $x_2$ is given by

$$p(x_1|x_2) = \frac{p(x_1, x_2)}{p(x_2)}. \quad (2.4)$$

The notation $p(x_1|\theta)$ is also used if $\theta$ is not considered a random variable, but a deterministic (non-random) variable on which the distribution of $x_1$ depends.

Equation (2.4) can be reformulated to yield another expression for the joint PDF,

$$p(x_1, x_2) = p(x_1|x_2)p(x_2), \quad (2.5)$$

and swapping the roles of $x_1$ and $x_2$ in (2.5), and substituting into (2.4), yields Bayes’ theorem,

$$p(x_1|x_2) = \frac{p(x_2|x_1)p(x_1)}{p(x_2)}. \quad (2.6)$$

The components $x_1$ and $x_2$ are independent when the joint distribution can be written as a product of the marginal distributions,

$$p(x_1, x_2) = p(x_1)p(x_2), \quad (2.7)$$

or equivalently, when the conditional distribution is the same as the marginal distribution,

$$p(x_1|x_2) = p(x_1). \quad (2.8)$$

### 2.1.2 Expectations

The expectation of a function $f(x)$ of a random variable $x$ (with associated PDF $p_x$) is given by

$$\mathbb{E} f(x) = \mathbb{E}_{p_x} f(x) = \int f(x)p_x(x)dx. \quad (2.9)$$

The expression $\mathbb{E}$ is used instead of $\mathbb{E}_{p_x}$ whenever the probability distribution $p_x$ is clear from the context. The expected value of $x$ is defined as $\mathbb{E} x$, and the covariance of $x$ is defined as $\text{Cov} x = \mathbb{E}[(x - \mathbb{E} x)(x - \mathbb{E} x)^T]$. The covariance of $x$ also called the variance when $x$ is scalar.
2.1.3 Jensen's Inequality

Jensen’s inequality (Jensen, 1906; Rudin, 1987) is an inequality related to probability distributions and convex functions. One variant of the inequality states that

$$\int p(x)f(g(x))dx \geq f\left(\int p(x)g(x)dx\right),$$

(2.10)

or equivalently,

$$E_p(f(g(x))) \geq f(E_p(g(x))),$$

(2.11)

where \(f(x)\) is a convex function, \(g(x)\) is any continuous function, and \(p(x)\) is a probability distribution.

2.1.4 Kullback-Leibler Divergence

The Kullback-Leibler divergence (Cover and Thomas, 2012; Kullback and Leibler, 1951) is an asymmetric quantification of the difference between two probability distributions \(q(x)\) and \(p(x)\) over the same variable \(x\). It is given by the expectation of the log-likelihood ratio,

$$D_{\text{KL}}(q(x)||p(x)) = \int \log \frac{q(x)}{p(x)}q(x)dx = E_q\left[\log \frac{q(x)}{p(x)}\right].$$

(2.12)

The Kullback-Leibler divergence is nonnegative,

$$D_{\text{KL}}(q(x)||p(x)) \geq 0,$$

(2.13)

with equality only when \(q(x) = p(x)\) almost everywhere (Cover and Thomas, 2012). This can be shown by using Jensen’s inequality, see Section 2.1.3.

2.1.5 List of Probability Distributions

This section defines the probability distributions which will be of use in this thesis. For more information, see, e.g., Gelman et al. (2014), Johnson et al. (1995a), and Johnson et al. (1995b). In Figure 2.1, three of the distributions discussed below are shown.

Gaussian Distribution

The univariate normal or Gaussian distribution (Johnson et al., 1995a) is one of the most common continuous distributions. Its PDF is given by

$$\mathcal{N}(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma^2}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

(2.14)

supported on \(x \in \mathbb{R}\). The parameters are the mean \(E_x = \mu \in \mathbb{R}\), and variance \(\text{Var } x = \sigma^2 > 0\). Probability distributions are often also parameterized using
location and scale parameters. In this case, the location parameter coincides with the mean \( \mu \) and the scale parameter coincides with the variance \( \sigma^2 \).

The multivariate Gaussian distribution in \( n \) dimensions (Anderson, 1958) has the PDF

\[
\mathcal{N}(x|\mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} e^{-\frac{(x-\mu)^T \Sigma^{-1} (x-\mu)}{2}}
\]  

(2.15)
supported on \( x \in \mathbb{R}^n \). It is parameterized by the mean \( \mathbb{E}x = \mu \in \mathbb{R}^n \) and covariance \( \text{Cov} x = \Sigma > 0 \), where the inequality denotes a positive definite matrix.

**Gamma Distribution**

The Gamma distribution (Johnson et al., 1995a) is given by the PDF

\[
\mathcal{G}(x|a, b) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}
\]

(2.16)
supported on \( x \geq 0 \), with shape parameter \( a > 0 \), and rate (or inverse scale) parameter \( b > 0 \), where \( \Gamma(a) \) is the Gamma function (Abramowitz and Stegun, 1964). An alternative parameterization with a scale parameter \( 1/b \) also exists. The expected value is \( \mathbb{E}x = a/b \), the variance is \( \text{Var} x = a/b^2 \), and the expected value of \( \log x \) is \( \mathbb{E} \log x = \psi(a) - \log(b) \), where \( \psi(a) = \frac{d}{da} \log \Gamma(a) \).
Student’s $t$ Distribution

The Student’s $t$ distribution with location-scale parameters (Johnson et al., 1995b) is defined by the PDF

$$T(x|\mu, \sigma^2, \nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi\sigma^2}\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{(x-\mu)^2}{\nu\sigma^2}\right)^{-\frac{\nu+1}{2}}$$

with support on $x \in \mathbb{R}$, location parameter $\mu \in \mathbb{R}$, scale parameter $\sigma^2 > 0$, and shape parameter $\nu > 0$. The expected value is $\mathbb{E}x = \mu$ if $\nu > 1$, and undefined otherwise. Similarly, the variance $\text{Var}x = \nu\sigma^2/(\nu - 2)$ if $\nu > 2$, and infinite or undefined otherwise.

The Student’s $t$ PDF tends to the Gaussian as $\nu \to \infty$ (Gelman et al., 2014). This shows that the Student’s $t$ is a generalization of the Gaussian. The density decays polynomially, $T(x|\mu, \sigma^2, \nu) = \mathcal{O}(|x|^{-\nu})$ as $|x| \to \infty$, with $\nu$ finite. This can be put in contrast to the Gaussian distribution, which has exponential tail decay, $\mathcal{N}(x|\mu, \sigma^2) = \mathcal{O}(\exp(-x^2/\sigma^2))$ as $|x| \to \infty$.

The Student’s $t$ distribution can be represented as a mixture of Gaussian variables with Gamma distributed precision (inverse variance) $\lambda$ (Johnson et al., 1995b). Specifically, its PDF can be written as

$$T(x|\mu, \sigma^2, \nu) = \int \mathcal{N}(x|\mu, \sigma^2/\lambda) \mathcal{G}(\lambda|\nu/2, \nu/2)d\lambda.$$  \hspace{1cm} (2.18)

This representation is a key property that makes the Student’s $t$ distribution useful.

Delta Distribution

The Delta distribution, also known as the Dirac or degenerate distribution, is given by the PDF

$$p(x|\mu) = \delta(x - \mu),$$

where $\delta$ is the Dirac delta (Benedetto, 1996). The Dirac delta is defined through the integral relation

$$\int_a^b \delta(x) dx = \begin{cases} 1, & 0 \in [a, b], \\ 0, & \text{otherwise}. \end{cases}$$

The expected value is $\mathbb{E}x = \mu$, and the variance is $\text{Var}x = 0$.

2.2 Statistical Models

Probability theory is often used for constructing statistical models (Bishop, 2006). These models relate unknown quantities (parameters) $\theta$ to known quantities (measurements) $Y$. One approach is to consider $\theta$ as deterministic, and $Y$ as random. This means that the statistical model can be written as a likelihood

$$p(Y|\theta).$$  \hspace{1cm} (2.21)
Another approach is to model both $\theta$ and $Y$ as random. The statistical model then can be written as a joint distribution $p(Y, \theta)$, which commonly is factorized as
\[
p(Y, \theta) = p(Y|\theta)p(\theta),
\]
into the likelihood $p(Y|\theta)$ and the prior $p(\theta)$.

### 2.2.1 State-Space Models

Probabilistic state-space models (SSMs (Arnold, 1974; Kay, 1993)) are an important class of statistical models. For SSMs, the parameter of interest can be the current state, $\theta = x_k$, with time index $k$. The available measurements are often $Y = Y_k = (y_j)_{j=1}^k$, where each $y_j$ is a measurement from time index $j \leq k$. An SSM describes the time evolution of the state $x_k$, and the relation between the state, $x_k$, and the measurement at that point in time, $y_k$. One representation is the functional representation
\[
x_{k+1} = f(x_k, v_k), \quad v_k \sim p_v(v_k),
\]
\[
y_k = h(x_k, e_k), \quad e_k \sim p_e(e_k),
\]
where the process noise $v_k$ and the measurement noise $e_k$ introduce stochasticity into the SSM through the functions $f$ and $h$. Another way to describe SSMs is with the distributional representation
\[
x_{k+1}|x_k \sim p(x_{k+1}|x_k),
\]
\[
y_k|x_k \sim p(y_k|x_k).
\]
The SSMs in this thesis can be formulated both in distributional and functional form, and both will be used interchangeably. Typically, SSMs will be formulated using the functional representation, and the noise will be modeled as additive,
\[
f(x_k, v_k) = f(x_k) + v_k,
\]
\[
h(x_k, e_k) = h(x_k) + e_k,
\]
with slight abuse of notation for $f$ and $h$. In this case, it is straightforward to translate the functional representation into a distributional one, since
\[
p(x_{k+1}|x_k) = p_v(x_{k+1} - f(x_k)),
\]
\[
p(y_k|x_k) = p_e(y_k - h(x_k)).
\]

### 2.2.2 Conditionally Linear State-Space Models

A special case of interest, see (Doucet et al., 2000; Lindsten, 2011; Schön et al., 2005, 2006), is when subcomponents of the state $x_k$ in a SSM occur linearly, conditioned on the rest of the states. In this case, the state $x_k$ may be split into two parts,
\[
x_k = \begin{bmatrix} x^n_k \\ x^a_k \end{bmatrix},
\]
where $x^a_k$ is the linear or analytical part and $x^n_k$ is the nonlinear or numerical part. Then, the state-space equations (2.23) can be written as

\begin{align}
  x^n_{k+1} &= f^n(x^n_k) + F^n(x^n_k)x^a_k + G^n(x^n_k)v^n_k, \
  x^a_{k+1} &= f^a(x^n_k) + F^a(x^n_k)x^a_k + G^a(x^n_k)v^a_k, \
  y_k &= h(x^n_k) + H(x^n_k)x^a_k + e_k.
\end{align}

(2.28a) 
(2.28b) 
(2.28c)

Note that all three equations are linear with respect to $x^a_k$ when $x^n_k$ is fixed; aconditionally linear ssm (CLSSM).

### 2.3 Frequentist Inference

Frequentist inference aims to compute estimates $\hat{\theta}$ of a parameter $\theta$ using measurements $Y$. The parameter $\theta$ is assumed to be unknown but deterministic, with a true value of $\theta^*$ (Cramér, 1948; van Trees, 1968; Yates and Goodman, 2005). The estimate $\hat{\theta} = \hat{\theta}(Y)$ is a function of the measurements $Y$, assumed to be related to the parameter $\theta$ through the likelihood $p(Y|\theta)$.

#### 2.3.1 Maximum Likelihood Estimation

The maximum likelihood (ML) method (Cramér, 1948; Fisher, 1925; Yates and Goodman, 2005) is a well known approach for computing estimates $\hat{\theta}$. It considers the likelihood function, which is the conditional distribution $p(Y|\theta)$ of the measurements given the parameters. Assuming that the measurements $Y = (y_j)_{j=1}^N$ are conditionally independent, this likelihood can be written as

\[ p(Y|\theta) = \prod_{k=1}^N p(y_k|\theta), \]

(2.29)

or with a logarithmic transformation,

\[ \log p(Y|\theta) = \sum_{k=1}^N \log p(y_k|\theta). \]

(2.30)

The ML estimate $\hat{\theta}$ is then given by

\[ \hat{\theta}^{\text{ML}} = \hat{\theta}^{\text{ML}}(Y) = \arg \max_{\theta} \log p(Y|\theta) = \arg \max_{\theta} \sum_{k=1}^N \log p(y_k|\theta), \]

(2.31)

since the logarithm is monotonically increasing.

The ML estimate is asymptotically consistent and efficient when the number of data points $N$ grows large under certain regularity and observability conditions (Cramér, 1948; Kay, 1993). This means that with the dimension of $\theta$ kept constant, the ML estimate $\hat{\theta}^{\text{ML}}$ satisfies

\[ \hat{\theta}^{\text{ML}} \overset{d}{\to} \theta^*, \]

(2.32)

\[ \text{cov}(\hat{\theta}^{\text{ML}}) \to J(\theta^*)^{-1}, \]

(2.33)
when \( N \to \infty \). Above, \( \xrightarrow{d} \) denotes convergence in distribution, and \( \mathcal{J}(\theta^*) \) is the Fisher information matrix, with elements
\[
(\mathcal{J}(\theta^*))_{i,j} = - \mathbb{E}_{p(Y|\theta)} \left[ \frac{d^2 \log p(Y|\theta)}{d\theta_i d\theta_j} \right]_{\theta = \theta^*}.
\] (2.34)

Under certain regularity conditions (Cramér, 1948), the limit (2.33) is a lower bound on the covariance for unbiased estimators, i.e., estimators \( \hat{\theta}(Y) \) satisfying
\[
\mathbb{E}_{p(Y|\theta^*)} \hat{\theta}(Y) = \theta^*.
\] (2.35)

### 2.3.2 Expectation Maximization

The expectation maximization (EM) method (Dempster et al., 1977) is a method that can be used to compute the maximum likelihood solution, see Section 2.3.1, when the optimization problem is difficult to solve directly.

Assume the likelihood \( p(Y|\theta) \) to be unavailable or unsuitable for direct maximization, but that it can be expressed as
\[
p(Y|\theta) = \int p(Z,Y|\theta) dZ,
\] (2.36)
where \( Z \) is a vector of unobserved latent variables. The EM method optimizes \( p(Y|\theta) \) indirectly by repeatedly optimizing a lower bound of \( \log p(Y|\theta) \). The method requires an initial guess \( \theta^{(0)} \) of the parameters \( \theta \) as an input. At iteration \( i \geq 1 \), the method then proceeds by considering the distribution
\[
p(Z|Y, \theta^{(i-1)})
\]
of the latent variables, \( Z \), given the previous value of the parameters, \( \theta^{(i-1)} \), and the measurements, \( Y \). The log likelihood \( \log p(Y|\theta) \) can then be bounded from below as
\[
\log p(Y|\theta) = \log \int p(Z, Y|\theta) dZ
\]
\[
= \log \int p(Z|Y, \theta^{(i-1)}) \frac{p(Z, Y|\theta)}{p(Z|Y, \theta^{(i-1)})} dZ
\]
\[
\geq \int p(Z|Y, \theta^{(i-1)}) \log \frac{p(Z, Y|\theta)}{p(Z|Y, \theta^{(i-1)})} dZ
\]
\[
= \int p(Z|Y, \theta^{(i-1)}) \log p(Z, Y|\theta) dZ
\]
\[
- \int p(Z|Y, \theta^{(i-1)}) \log p(Z|Y, \theta^{(i-1)}) dZ
\]
\[
= \int p(Z|Y, \theta^{(i-1)}) \log p(Z, Y|\theta) dZ + \text{const},
\] (2.37)
where Jensen’s inequality (Section 2.1.3) was used, and the constant is independent of \( \theta \).
Algorithm 2.1 The expectation maximization (EM) method (Dempster et al., 1977)

Require:
Initial parameter guess $\theta^{(0)}$.
Set $i = 0$.

while not converged do
  Set $i = i + 1$.
  E step: Set $Q(\theta|\theta^{(i-1)}) = \mathbb{E}_{p(Z|Y,\theta^{(i-1)})} \log p(Y, Z|\theta)$.
  M step: Set $\theta^{(i)} = \arg\max_\theta Q(\theta|\theta^{(i-1)})$.
end while
return $\hat{\theta}^{EM} = \theta^{(i)}$.

The expectation step, or E step, consists of computing the integral in the right hand side of (2.37) to get the $Q$ function,

$$Q(\theta|\theta^{(i-1)}) = \int p(Z|Y, \theta^{(i-1)}) \log p(Z, Y|\theta) dZ = \mathbb{E}_{p(Z|Y,\theta^{(i-1)})} \log p(Z, Y|\theta). \tag{2.38}$$

The update equation for $\theta$ is given by the maximization step, or M step,

$$\theta^{(i)} = \arg\max_\theta Q(\theta|\theta^{(i-1)}). \tag{2.39}$$

Since the $Q$ function is the only component dependent of $\theta$ on the right hand side of (2.37), this choice will increase the lower bound in (2.37).

The EM method consists of iterating (2.38) and (2.39) until convergence, as quantified by a sufficiently low value of the parameter difference $\|\theta^{(i)} - \theta^{(i-1)}\|$. Under technical conditions discussed by Wu (1983), each iteration can be shown not to decrease the likelihood. This indicates that the final value $\hat{\theta}^{EM}$ is a local maximizer of $p(Y|\theta)$. A summary is available in Algorithm 2.1.

2.4 Bayesian Inference

Bayesian inference works with probability distributions over parameters (Gelman et al., 2014), in contrast to frequentist inference which assumes that parameters have an unknown deterministic true value. The model used in Bayesian inference is the joint probabilistic model $p(Y, \theta) = p(Y|\theta)p(\theta)$ over the measurements $Y$ and the parameters $\theta$. Bayes’ rule, (2.6), can then be used to form the posterior

$$p(\theta|Y) = \frac{p(Y|\theta)p(\theta)}{p(Y)} \propto p(Y|\theta)p(\theta), \tag{2.40}$$

where the proportionality sign signifies that the marginal likelihood of the data,

$$p(Y) = \int p(Y|\theta)p(\theta) d\theta, \tag{2.41}$$

does not depend on $\theta$. 
One of the main issues with the Bayesian approach is that computation of the posterior in an analytically tractable form can be difficult for many models occurring in practice. Even if \( p(Y|\theta) \) and \( p(\theta) \) are PDFs which can be easily expressed, the marginal likelihood \( p(Y) \) is often analytically intractable.

### 2.4.1 Bayesian Point Estimation

Given a Bayesian posterior \( p(\theta|Y) \), it may be of interest to compute point estimates. This is typically achieved by defining a loss function, \( l(\theta', \theta) \), and selecting \( \hat{\theta} \) as

\[
\hat{\theta} = \arg \min_{\theta'} E_{p(\theta|Y)} l(\theta', \theta).
\]  

(2.42)

A standard choice, which yields an analytic expression for \( \theta' \), is to set \( l(\theta', \theta) = \|\theta - \theta'\|^2 \) as the squared Euclidean distance. This yields the minimum mean square error estimate, which has an explicit expression in the posterior mean,

\[
\hat{\theta} = E_{p(\theta|Y)} \theta.
\]  

(2.43)

This choice yields that \( \hat{\theta} \xrightarrow{P} \hat{\theta}^{\text{ML}} \) when \( N \to \infty \) (Van der Vaart, 1998), where \( \xrightarrow{P} \) denotes convergence in probability.

### 2.4.2 Discretization of Bayesian Posteriors

Discretization of the domain of \( p(\theta|Y) \) is one approach which yields a tractable approximation of the posterior. The approximation is given by

\[
p(\theta|Y) \approx \sum_{j=1}^{N_S} w^j \delta (\theta - \theta^j),
\]

in the form of a mixture of delta distributions. The parameters \( \theta^j \) are discretized values on a grid over the parameter space of \( \theta \), and \( w^j \) are associated discretization weights. To make the approximation a valid PDF, one must require \( \sum_{j=1}^{N_S} w^j = 1 \) and \( w^j \geq 0 \) for all \( j \).

The discretized values \( \theta^j \) can be chosen deterministically, e.g., as equidistant values on a grid over the significant support of \( p(\theta|Y) \). The unnormalized weights are then chosen as \( w^j \propto p(\theta^j, Y) \) (Bucy and Senne, 1971), where the common proportionality constant is chosen to ensure \( \sum_{j=1}^{N_S} w^j = 1 \).

The value \( \theta^j \) can also be sampled. Using approaches such as rejection sampling or Markov chain Monte Carlo (MCMC) (Bishop, 2006), it is possible to sample directly from \( p(\theta|Y) \). Then, one can set \( w^j = \frac{1}{N_S} \) for all \( j \). Another choice is importance sampling; in this approach, one samples from an approximation \( q(\theta) \) of \( p(\theta|Y) \) (Bishop, 2006). To compensate for this, the weights must be set to \( w^j \propto p(\theta^j|Y)/q(\theta^j) \), again normalized so that \( \sum_{j=1}^{N_S} w^j = 1 \), in order to obtain an unbiased sample (Bishop, 2006). As \( N_S \to \infty \), it can be shown that the sampling
approximation will approach the true distribution \( p(\theta|Y) \) (Robert and Casella, 2004).

The described approximations all suffer from the curse of dimensionality. In the worst case scenario, a \( d \)-dimensional parameter \( \theta \) requires a number of discretization points \( N_S \) that is exponential in \( d \) in order to achieve a given error size (Davis and Rabinowitz, 2007; Snyder et al., 2008).

### 2.4.3 Rao-Blackwellization in Discretized Approximations

Rao-Blackwellization is a general technique for variance reduction (Blackwell, 1947). It can be used in order to get an analytic solution for a factor of the posterior when the full posterior is not analytically available.

Consider a statistical model with two subsets of parameters, \( \theta = [\theta^a, \theta^n]^T \). Factorize the posterior as

\[
p(\theta^a, \theta^n|Y) = p(\theta^a|\theta^n, Y)p(\theta^n|Y)
\]  

(2.44)

using conditional probability, where the conditional distribution \( p(\theta^a|\theta^n, Y) \) is assumed to be analytically available (hence the superscript \( a \)). A consequence of the Rao-Blackwell theorem (Blackwell, 1947) is that it is more accurate to approximate the marginal density \( p(\theta^n|Y) \) in the right hand side of (2.44), than to approximate the full density \( p(\theta^a, \theta^n|Y) \) in the left hand side of (2.44). If \( (\theta^{n,j})_{j=1}^{N_S} \) are samples from \( p(\theta^n|Y) \), then it can be shown (Casella and Robert, 1996; Doucet et al., 2000) that

\[
\frac{1}{N_S} \sum_{j=1}^{N_S} p(\theta^a|\theta^{n,j}, Y) \delta(\theta^n - \theta^{n,j})
\]  

(2.45)

is a more accurate approximation of \( p(\theta^a, \theta^n|Y) \) than

\[
\frac{1}{N_S} \sum_{j=1}^{N_S} \delta(\theta^n - \theta^{a,j})
\]  

(2.46)

where \( (\theta^{n,j}, \theta^{a,j}) \) are corresponding samples from the joint posterior \( p(\theta^a, \theta^n|Y) \).

The specific meaning of “more accurate” is that approximations of

\[
E_{\theta^a, \theta^n|Y} f(\theta^a, \theta^n),
\]  

(2.47)

for bounded functions \( f(\theta^a, \theta^n) \), have lower variance when expectations are approximated using (2.45) than when using (2.46). The practical meaning of this result is that for a given required accuracy of the posterior approximation, a Rao-Blackwellized sampling method needs a lower value of \( N_S \) than a full sampling method does.
2.4.4 Variational Bayes

The variational Bayes method is about approximating a posterior \( p(\theta | Y) \) by an analytic approximate posterior \( q(\theta) \). The approximation \( q(\theta) \) has a functional form which can be either parameterized or derived using a component-wise independence assumption for \( q(\theta) \) (Bishop, 2006; Fox and Roberts, 2012; Jordan et al., 1999; Tzikas et al., 2008). This method is explained in more detail because it will be needed later on.

Variational Decomposition of the Log-Marginal Likelihood

Assume that one wishes to compute an approximate distribution

\[
q(\theta) \approx p(\theta | Y),
\]

(2.48)

where \( p(\theta | Y) \) is not available analytically. Since \( q(\theta) \) is a PDF, \( \int q(\theta) d\theta = 1 \). Thus, the logarithm of the marginal data likelihood, \( \log p(Y) \), can be written as

\[
\log p(Y) = \log p(Y) \int q(\theta) d\theta = \int q(\theta) \log p(Y) d\theta = \int q(\theta) \log \frac{p(\theta, Y)}{p(\theta | Y)} d\theta,
\]

(2.49)

where the definition of conditional probability, \( p(\theta, Y) = p(Y)p(\theta | Y) \), has been used in the last equality. Multiplying both numerator and denominator in the fraction by \( q(\theta) \), and expanding the logarithm, yields the sum of two integrals,

\[
\log p(Y) = \int q(\theta) \log \frac{p(\theta, Y)q(\theta)}{q(\theta)p(\theta | Y)} d\theta
\]

\[
= \int q(\theta) \log \frac{p(\theta, Y)}{q(\theta)} d\theta + \int q(\theta) \log \frac{q(\theta)}{p(\theta | Y)} d\theta,
\]

(2.50)

where the left summand defines the evidence lower bound \( \mathcal{L}(q(\theta); p(\theta, Y)) \), and the right summand is the Kullback-Leibler divergence \( D_{\text{KL}}(q(\theta) || p(\theta | Y)) \).

Several observations can be made. The left hand side of (2.50), \( \log p(Y) \), does not depend on \( q(\theta) \). Meanwhile, the summands on the right hand side do. The second summand is the Kullback-Leibler divergence from the exact posterior \( p(\theta | Y) \) to the approximate posterior \( q(\theta) \). In Section 2.1.4 it was seen that

\[
D_{\text{KL}}(q(\theta) || p(\theta | Y)) \geq 0,
\]

(2.51)

with equality only when \( q(\theta) = p(\theta | Y) \). This indicates that the Kullback-Leibler divergence is a suitable choice of minimization criterion.

The posterior distribution \( p(\theta | Y) \) occurs in the expression \( D_{\text{KL}}(q(\theta) || p(\theta | Y)) \), so it is typically intractable to minimize it analytically. However, the minimization problem can be rewritten using (2.50) as

\[
\arg \min_{q(\theta)} D_{\text{KL}}(q(\theta) || p(\theta | Y)) = \arg \min_{q(\theta)} \left( \log p(Y) - \mathcal{L}(q(\theta); p(\theta, Y)) \right)
\]

\[
= \arg \max_{q(\theta)} \mathcal{L}(q(\theta); p(\theta, Y)),
\]

(2.52)
where \( \log p(Y) \) was disregarded since it is independent of \( q(\theta) \). The evidence lower bound \( \mathcal{L}(q(\theta); p(\theta, Y)) \) can be evaluated for suitable \( q(\theta) \) since neither the posterior nor the marginal likelihood are needed.

If \( q(\theta) \) is allowed to be any PDF, the optimal \( q(\theta) \) is the exact posterior;

\[
\arg \max_{q(\theta) \text{ any PDF}} \mathcal{L}(q(\theta); p(\theta, Y)) = p(\theta|Y),
\]

as seen above in (2.51). Since \( p(\theta|Y) \) is analytically unavailable, this does not help in computing a tractable approximation of it. Constraints must thus be imposed on \( q(\theta) \) in order to get a tractable solution. One choice is to assume a factorized approximation

\[
\arg \max_{q(\theta) = q(\theta_1)\delta(\theta_2 - \hat{\theta}_2)} \mathcal{L}(q(\theta); p(\theta, Y)),
\]

where \( \theta = [\theta_1^T, \theta_2^T]^T \). This is equivalent to the EM method described in Section 2.3.2 (Tzikas et al., 2008), with \( X \) substituted for \( \theta_1 \) and \( \theta \) substituted for \( \theta_2 \), when one chooses a prior \( p(\theta_2) \propto 1 \). In the sequel, a more general factorization assumption called mean-field variational Bayes is described.

**Mean-Field Variational Approximation**

Split \( \theta \) into two components, \( \theta = [\theta_1^T, \theta_2^T]^T \), and assume posterior independence between these components. This means that \( q(\theta) \) can be factorized, so that

\[
q(\theta) = q_1(\theta_1)q_2(\theta_2).
\]

In many cases, this factorization or mean-field assumption is sufficient to make the expectations in (2.56) analytical. Note that the parameter could be split into more than two subcomponents, but for simplicity, only the case for two subcomponents is discussed in detail.

It can be shown using variational calculus (Jordan et al., 1999) that the optimal \( q^*(\theta) = q^*(\theta_1)q^*(\theta_2) \) satisfies

\[
\log q^*_2(\theta_2) = \mathbb{E}_{q_1^*(\theta_1)} [\log p(\theta_1, \theta_2, Y)] + \text{const},
\]

\[
\log q^*_1(\theta_1) = \mathbb{E}_{q_2^*(\theta_2)} [\log p(\theta_1, \theta_2, Y)] + \text{const},
\]

where the additive constants in the equations above are chosen such that the PDFs normalize. Since the equations (2.56) are implicit, they must be solved. One way to solve them is to iterate them, and it can be shown that carrying out one of the updates above is guaranteed not to decrease \( \mathcal{L} \). This suggests an iterative coordinate ascent scheme which can be used until convergence to compute the approximating distribution \( q^*(\theta) = q_2^*(\theta_1)q_2^*(\theta_2) \) (Jordan et al., 1999), as described in Algorithm 2.2. Convergence can be quantified by a sufficiently small value of \( \mathcal{L}^{(i)} - \mathcal{L}^{(i-1)} \) or by a sufficiently small change in summary statistics such as \( \mathbb{E}_{q^{(i)}(\theta)} \theta \).

This method will converge to a local maximum of the evidence lower bound (Lange, 2013). Wang and Blei (2017) show that under certain conditions, Bayesian point estimators extracted from the variational Bayes solution are consistent.
Algorithm 2.2 Mean-field variational Bayes with two variational factors (Bishop, 2006)

Require: Initialization $q_1^{(0)}(\theta_1)$ as a suitable density.
Set $i = 1$, $\mathcal{L}^{(i-1)} = -\infty$.

while not converged do
    Compute $q_2^{(i)}(\theta_2)$ using (2.56a).
    Compute $q_1^{(i)}(\theta_1)$ using (2.56b).
    Set $q^{(i)}(\theta) = q_1^{(i)}(\theta_1)q_2^{(i)}(\theta_2)$.
    Compute $L^{(i)} = \mathbb{E}_{q^{(i)}(\theta)}\left[\log \frac{p(\theta, Y)}{q^{(i)}(\theta)}\right]$.
    Set $i = i + 1$ and continue.
end while

return $q^*(\theta) = q^{(i)}(\theta)$ and $L^{(i)}$.

2.4.5 Bayesian Filtering and Smoothing

Consider Bayesian inference in a state-space model, as discussed in Section 2.2, where the available measurements $Y = Y_k = (y_j)_{j=1}^k$, and the parameter of interest $\theta$ is the current state of the dynamical system, $\theta = x_k$. This is called Bayesian filtering (Jazwinski, 1970; Särkkä, 2013). It is of particular interest to achieve a recursive formulation so that the computational complexity of the filter at each time step is the same. In order to recursively solve the Bayesian filtering problem, the following PDFs

$$
p(x_k|Y_{k-1}) = \int p(x_k|x_{k-1})p(x_{k-1}|Y_{k-1}) \, dx_{k-1}, \quad (2.57a)
$$

$$
p(x_k|Y_k) = \frac{p(y_k|x_k)p(x_k|Y_{k-1})}{\int p(y_k|x_k)p(x_k|Y_{k-1}) \, dx_k}, \quad (2.57b)
$$

must be computed. The first and second equation, respectively, correspond to the time update or prediction, and the measurement update or correction. In Bayesian smoothing, one wishes to compute the posterior distribution

$$
p(x_k|Y_N), \quad (2.58)
$$

with $N > k$. This problem can be recursively solved using the distributions (2.57) from the filter and the backward iteration

$$
p(x_k|Y_N) = p(x_k|Y_k) \int \frac{p(x_{k+1}|x_k)p(x_{k+1}|Y_N)}{p(x_{k+1}|Y_k)} \, dx_{k+1}. \quad (2.59)
$$

For more information, see the book by Särkkä (2013).

Kalman Filter

The Kalman filter (KF) is an analytic solution to the Bayesian filtering problem in the special case when the state-space model is linear and the noise is Gaussian.
Algorithm 2.3 Standard Kalman filter (Kalman, 1960)

Require: $p(x_0) = p(x_0|Y_0) = \mathcal{N}(x_0|\hat{x}_0, P_{0|0})$.

Require: Model parameters $F_k, H_k, Q_k, R_k$.

for $k = 1, \ldots, N$ do

Time update:

\[
p(x_k|Y_{k-1}) = \mathcal{N}(\hat{x}_{k|k-1}, P_{k|k-1}),
\]
\[
\hat{x}_{k|k-1} = F_{k-1} \hat{x}_{k-1|k-1},
\]
\[
P_{k|k-1} = F_{k-1} P_{k-1|k-1} F_{k-1}^T + Q_{k-1}.
\]

Measurement update:

\[
p(x_k|Y_k) = \mathcal{N}(\hat{x}_k, P_k),
\]
\[
\hat{x}_k = \hat{x}_{k|k-1} + K_k (y_k - H_k \hat{x}_{k|k-1}),
\]
\[
P_k = (I - K_k H_k) P_{k|k-1},
\]
\[
K_k = P_{k|k-1} H_k^T S_k^{-1},
\]
\[
S_k = H_k P_{k|k-1} H_k^T + R_k.
\]

end for

Consider the structure (2.25) where $f(x_k) = F_k x_k$ and $h(x_k) = H_k x_k$ are linear in $x_k$. Furthermore, assume that $v_k \sim \mathcal{N}(0, Q_k)$ and $e_k \sim \mathcal{N}(0, R_k)$ are Gaussian, mutually independent, and independent in time. In this case, the Kalman filter (Kailath, 1980; Kalman, 1960; Särkkä, 2013) can be applied. Analytical evaluation of the Bayesian recursion (2.57) yields the exact solution shown in Algorithm 2.3.

Rauch-Tung-Striebel Smoother

The Rauch-Tung-Striebel smoother is a recursive formulation of the full Bayesian smoothing problem; where one wishes to compute $p(x_k|Y_N)$, with $N > k$. Like the KF, it applies to linear-Gaussian dynamical systems as detailed above. The method is detailed in Algorithm 2.4; for more information, see the book by Särkkä (2013).

Particle Filter

The idea of the particle filter (PF) (Doucet, 1998; Gordon et al., 1993) is to compute an approximate solution

\[
p(x_k|Y_k) \approx \sum_{j=1}^{N_S} w_j^k \delta(x_k - x_j^k) \quad (2.60)
\]

of (2.57) using sampling methods, as discussed in Section 2.4.2. In particular, it uses importance sampling with the proposal $q(x_k|x_{k-1}, y_k)$, with weights derived...
Algorithm 2.4 Rauch-Tung Striebel smoother (Särkkä, 2013)

Require: Predictive and filtering distributions $p(x_k|y_{k-1}) = \mathcal{N}(\hat{x}_{k|k-1}, P_{k|k-1})$, $p(x_k|y_k) = \mathcal{N}(\hat{x}_{k|k}, P_{k|k})$ from the Kalman filter (Algorithm 2.3).

Require: Model parameters $F_k$, $H_k$, $Q_k$, $R_k$.

for $k = N-1, \ldots, 1$ do

Backward recursion:

$$p(x_k|y_N) = \mathcal{N}(\hat{x}_{k|N}, P_{k|N}),$$

$$\hat{x}_{k|N} = \hat{x}_{k|k} + G_k(\hat{x}_{k+1|N} - \hat{x}_{k+1|k}),$$

$$P_{k|N} = P_{k|k} + G_k(P_{k+1|N} - P_{k+1|k})G_k^T,$$

$$G_k = P_{k|k}F_k^TP_{k+1|k}. $$

end for

from the ssm distributions $p(x_k|x_{k-1})$ and $p(y_k|x_k)$, to update $x_k^j$ and $w_k^j$. After sampling, the algorithm resamples the approximation (2.60). A summary of the bootstrap particle filter, where the bootstrap proposal $q(x_k|x_{k-1}, y_k) = p(x_k|x_{k-1})$ and multinomial resampling (Evans et al., 2000) are used, is available in Algorithm 2.5. Under conditions detailed by, e.g., Chopin (2004), the PF will converge to the exact Bayesian solution of (2.57) when $N_S \to \infty$. Furthermore, note that the computational complexity of the PF is linear, $O(N_S)$.

Point Mass Filter

The point mass filter (PMF) or discrete grid filter (Bucy and Senne, 1971; Sorenson, 1974) evaluates the Bayesian filtering recursions (2.57) using a deterministic discretization $(x_j^k)^{N_S}_{j=1}$ over the state space of $x_k$, as discussed in Section 2.4.2. The method is given in Algorithm 2.6. The computational complexity of the PMF at each time step is quadratic, $O(N_S^2)$, due to the sum in the prediction. This can be contrasted to the linear complexity of the PF.

Rao-Blackwellized Particle Filter

The Rao-Blackwellized particle filter (RBPF) (Casella and Robert, 1996; Doucet et al., 2000; Schön et al., 2005, 2006) uses Rao-Blackwellization, see Section 2.4.3, in conjunction with the PF. The factorization corresponding to (2.44) is given by

$$p(X^n_k, x^n_k|y_k) = p(X^n_k|y_k)p(x^n_k|X^n_k, y_k),$$

where the full state history $\theta^n = X^n_k = (x^n_j)^k_{j=1}$ must be included in order to have an analytical conditional expression for $\theta^n = x^n_k$. The distribution $p(X^n_k|y_k)$ is approximated using a PF where $x^n_k$ has been analytically marginalized out, and the conditional distribution $p(x^n_k|X^n_k, y_k)$ is computed using KFs. See Algorithm 2.7 for a description of the filter.
The basic idea is similar to the Bayesian Inference.

**Algorithm 2.5** Bootstrap particle filter (PF) (Doucet, 1998; Gordon et al., 1993)

Require: \( p(x_0) = p(x_0|Y_0) \approx \sum_{j=1}^{N_S} w_0^j \delta(x_0 - x_0^j) \).

Require: Model \( p(x_k|x_{k-1}), p(y_k|x_k) \).

for \( k = 1, \ldots, N \) do
  for \( j = 1, \ldots, N_S \) do
    Time update: Draw \( \tilde{x}_k^j \sim p(x_k|\tilde{x}_{k-1}^j) \)
    Measurement update: \( \tilde{w}_k^j \propto w_{k-1}^j p(y_k|x_k^j) \)
  end for
  for \( j = 1, \ldots, N_S \) do
    Resample: Set \( x_k^j = \tilde{x}_k^j \) with probability \( \tilde{w}_k^j \) for \( \ell = 1, \ldots, N_S \).
    Set \( w_k^j = \frac{1}{N_S} \).
  end for
Filtering distribution at time \( k \): \( p(x_k|Y_k) \approx \sum_{j=1}^{N_S} w_k^j \delta(x_k - x_k^j) \).
end for

**Algorithm 2.6** Standard point mass filter (PMF) (Bucy and Senne, 1971; Sorenson, 1974)

Require: Prior distribution \( p(x_0) \approx \sum_{j=1}^{N_S} w_0^j \delta(x_0 - x_0^j) \).

Require: Model \( p(x_k|x_{k-1}), p(y_k|x_k) \).

for \( k = 1, \ldots, N \) do
  for \( j = 1, \ldots, N_S \) do
    Optionally, update grid \( x_k^j \), or set \( x_k^j = x_{k-1}^j \).
    Time update: \( p(x_k^j|Y_{k-1}) = \sum_{j=1}^{N_S} w_{k|k-1}^i \delta(x_k - x_k^j) \), with
    \( w_{k|k-1}^j = \sum_{i=1}^{N_S} w_{i|k-1}^i w_{k|k-1}^i p(x_k^i|x_{k-1}^i) / \sum_{i=1}^{N_S} w_{i|k-1}^i p(x_k^i|x_{k-1}^i) \).
    Measurement update: \( p(x_k|Y_k) = \sum_{j=1}^{N_S} w_{k|k}^j \delta(x_k - x_k^j) \), with
    \( w_{k|k}^j = w_{k|k-1}^j p(y_k|x_k = x_k^j) \).
  end for
Filtering distribution at time \( k \): \( p(x_k|Y_k) \approx \sum_{j=1}^{N_S} w_k^j \delta(x_k - x_k^j) \).
end for

**Rao-Blackwellized Point-Mass Filter**

The RBPMF is a Bayesian filtering method originally proposed by Smidl and Gasperin (2013), based on the Rao-Blackwellized marginal particle filter by (Lindsten, 2011). It will be explained in more detail, which will be useful later on. The basic idea is similar to the RBPF, but the RBPMF deals with the marginal distribution \( p(x_k^n, x_k^m|Y_k) \) instead of the full posterior distribution of the discretized component, \( p(X_k^n, x_k^m|Y_k) \). Hence, in contrast to (2.61), \( p(x_k^n|x_k^m, Y_k) \) is not linear-Gaussian. Instead, it is approximated as linear-Gaussian.
Algorithm 2.7 Bootstrap Rao-Blackwellized particle filter (RBPF) (Casella and Robert, 1996; Doucet et al., 2000; Schön et al., 2005, 2006)

Require: $p(x_0) = p(x_0|Y_0) = \sum_{j=1}^{N_S} w^0_j \delta(x^n_0 - x^{n,j}_0)N(x^n_0|\hat{x}^{a,j}_0, P^{a,j}_0)$.

Require: Model $p(x_k|x_{k-1})$, $p(y_k|x_k)$.

for $k = 1, \ldots, N$ do

for $j = 1, \ldots, N_S$ do

Draw: $\tilde{x}^{n,j}_k \sim p(x^n_k|x^{n,j}_{k-1})$.

Linear update: Compute $p(x_k|X^n_{k-1}, \tilde{x}^{n,j}_k, Y_k) = N(x_k|\hat{x}^{a,j}_k, \hat{P}^{a,j}_k)$

using the KF prediction and measurement update in Algorithm 2.3, with both $\tilde{x}^{n,j}_k$ and $y_k$ as measurements using (2.28a) and (2.28c).

Nonlinear measurement update: $\tilde{w}^j_k \propto w^j_{k-1} p(y_k|X^n_{k-1}, \tilde{x}^{n,j}_k, Y_{k-1})$.

end for

for $j = 1, \ldots, N_S$ do

Resample: $X^n_{k,j} = (X^n_{k-1}, \tilde{x}^{n,j}_k)$, with probability $\tilde{w}^j_k$ for $\ell = 1, \ldots, N_S$.

Set $w^j_k = \frac{1}{N_S}$.

Set $\hat{x}^{a,j}_k = \tilde{x}^{a,j}_k$, and $\hat{P}^{a,j}_k = P^{a,j}_k$.

end for

Filtering distribution at time $k$:

$p(x^n_k, x^n_{k}|Y_k) \approx \sum_{i=1}^{N_S} w^i_{k|k} N(x^n_k|\hat{x}^{a,i}_k, \hat{P}^{a,i}_k) \delta(x^n_k - x^{n,i}_k)$

end for

Assume that an existing conditionally linear-Gaussian approximation

$$p(x_k|Y_{k-1}) = p(x_k|X^n_{k-1}, Y_{k-1}) \propto \sum_{i=1}^{N} w^i_{k-1|k-1} N(x^n_k|\hat{x}^{a,i}_{k-1}, P^{a,i}_{k-1|k-1}) \delta(x^n_k - x^{n,i}_{k-1}),$$  \hspace{1cm} (2.62)

is available at time index $k-1$ using $N_S$ grid points. Now, a method to compute $p(x_k|Y_{k-1})$ with the same structure will be described. In order to compute $p(x_k|Y_{k-1})$, first define a grid $(x^{n,i}_{k-1})_{i=1}^{N_S}$. The joint distribution $p(x^n_k, x^n_{k}|Y_{k-1})$ can then be computed from (2.62), and $x^n_{k-1}$ can be marginalized away.

$$p(x^n_k, x^n_{k}|Y_{k-1}) \approx \sum_{i,j=1}^{N_S} w^{i,j}_{k-1|k-1} N(x^n_k|\hat{x}^{a,i,j}_{k-1}, P^{a,i,j}_{k-1|k-1}) \delta(x^n_k - x^{n,j}_{k-1}).$$  \hspace{1cm} (2.63)

A Gaussian distribution in $x^n_k$ is required to handle every combination of $x^{n,i}_{k-1}$ and $x^{n,j}_{k-1}$ (Lindsten, 2011; Smidl and Gasperin, 2013). Thus, $N^2_S$ KF means and covariance matrices must be saved and updated using $y_k$. If these parameters were all retained, the RBPMF would have exponential complexity $O(N^2_S)$, at time step $k$. Thus, this is unsuitable for implementation. To remedy this, Lindsten (2011) and Smidl and Gasperin (2013) merge means and covariance matrices through moment
matching in order to reduce the number of parameters. The moment matching step is given by

\[
\begin{align*}
    w_j^{k | k} &= \sum_{i=1}^{N_S} w_i^{i,j | k}, \\ 
    \hat{x}_a,j^{k | k} &= \sum_{i=1}^{N_S} w_i^{i,j | k} \hat{x}_a,i,j^{k | k}, \\ 
    P_{a,j}^{k | k} &= \sum_{i=1}^{N_S} w_i^{i,j | k} \left( P_{a,i,j}^{k | k} + (\hat{x}_a,j^{k | k} - \hat{x}_a,i,j^{k | k})(\hat{x}_a,j^{k | k} - \hat{x}_a,i,j^{k | k})^T \right),
\end{align*}
\]

(2.64a) (2.64b) (2.64c)

which corresponds to the approximation

\[
\begin{align*}
p(x_a^k, x_n^k | Y_{k-1}) &\approx \sum_{i,j=1}^{N_S} w_i^{i,j | k} \mathcal{N}(x_a^k | \hat{x}_a,i,j^{k | k}, P_{a,i,j}^{k | k}) \delta(x_n^k - x_n^{i,j}) \\ &\approx \sum_{j=1}^{N_S} w_j^{j | k} \mathcal{N}(x_a^k | \hat{x}_a,j^{k | k}, P_{a,j}^{k | k}) \delta(x_n^k - x_n^{j | k}).
\end{align*}
\]

(2.65)

This closes the recursion and returns an approximation on the same form as (2.62). See Algorithm 2.8 for all details.
Algorithm 2.8 The standard Rao-Blackwellized point mass filter (RBPMF) (Smidl and Gasperin, 2013)

Require: \( p(x^n_0, x^a_0| Y_0) \approx \sum_{i=1}^{N_S} w^i_0 N(x^n_0|x^n_0, P^n_0)\delta(x^n_0 - x^n_0) \)

for \( k = 1, \ldots, N \) do

Optionally, update grid \( x^{n,j}_k \), or set \( x^{n,j}_k = x^{n,j}_{k-1} \).

for \( j = 1, \ldots, N_S \) do

for \( i = 1, \ldots, N_S \) do

Linear time update: Compute the linear prediction
\[
p(x^a_k|x^{n,i}_{k-1}, x^{n,j}_k, Y_{k-1}) = N(x^a_k|x^{a,i}\hat{x}_k, P_{a,i}\hat{x}_k) \]
for each combination of \( x^{p,j}_k \) and \( x^{p,i}_{k-1} \) using the KF prediction in Algorithm 2.3.

Nonlinear time update: Compute \( w^{i,j}_k \propto w^{i}_{k-1} p(x^{n,j}_k|x^{n,i}_{k-1}) \).

Linear measurement update: Compute \( \bar{x}_{a,i,j}^k \) and \( \bar{P}_{a,i,j}^k \) for each combination of \( x^{p,j}_k \) and \( x^{p,i}_{k-1} \) using the KF measurement update in Algorithm 2.3 with both \( x^{n,j}_k \) and \( y_k \) as measurements using (2.28a) and (2.28c).

Nonlinear measurement update: \( w^{i,j}_k \propto w^{i}_{k-1} p(y_k|x^{n,j}_k, x^{n,i}_{k-1}) \).

end for

end for

end for

for \( j = 1, \ldots, N_S \) do

Merging step: Compute \( w^j_k \propto \frac{w^j_{k-1} p(y_k|x^{n,j}_k, x^{n,i}_{k-1})}{w^j_{k-1} p(x^{n,j}_k|x^{n,i}_{k-1})} \).

end for

Filtering distribution at time \( k \):
\[
p(x^n_k, x^a_k| Y_k) \approx \sum_{j=1}^{N_S} w^j_k N(x^n_k|x^{a,j}\bar{x}_k, P_{a,j}\bar{x}_k)\delta(x^n_k - x^n_k) \]

end for
2.5 Robust Estimation

The objective of robust estimation is to design estimators which are insensitive to deviations in the model assumptions (Huber, 1981). One phenomenon which often affects robustness in estimators is the occurrence of outliers in the measurements. For instance, when a Gaussian likelihood $p(Y|\theta)$ is assumed, robustness issues may appear in the presence of outliers (Insua and Ruggeri, 2012).

The influence function is one tool of robust statistics, which quantifies how much an additional data point with a given value will influence the parameter estimate (Huber, 1981). Consider an estimator $\hat{\theta} = \hat{\theta}(Y)$. In order to understand how a small change in the measurements changes the value of the estimate, consider the empirical PDF of $Y = (y_j)_{j=1}^N$,

$$p_Y(y) = \frac{1}{N} \sum_{j=1}^N \delta(y - y_j).$$  \hspace{1cm} (2.66)

The estimator $\hat{\theta}$ may be seen as a function of the empirical PDF $p_Y(y)$ instead of $Y$, (Huber, 1981). The influence function is then defined as the limit

$$r(x, p_Y, \hat{\theta}) = \lim_{s \to 0} \frac{\hat{\theta}((1-s)p_Y(y) + s\delta(y-x)) - \hat{\theta}(p_Y(y))}{s}.$$  \hspace{1cm} (2.67)

The value of the enumerator at $s = \frac{1}{N+1}$ is the change of $\hat{\theta}$ when an additional measurement at $y = x$ is added. The limit $s \to 0$ yields an asymptotic expression which often has a simpler form. For maximum-likelihood estimators, the influence function is given by

$$r(x, p_Y, \hat{\theta}) \propto - \frac{d}{d\theta} \log p(x|\theta) \bigg|_{\theta=\hat{\theta}(Y)},$$  \hspace{1cm} (2.68)

where the proportionality constant depends on parameters such as the number of data points $N$ (Huber, 1981).

A special case is when $\theta$ is a scalar location parameter,

$$p(y_j|\theta) = p_e(y_j - \theta).$$  \hspace{1cm} (2.69)

In this case, the influence function is proportional to the derivative of the log-likelihood. These derivatives are shown in Figure 2.2. The Student’s $t$ distribution has an influence function similar to the Gaussian one at values close to 0, but the Student’s $t$ influence function is bounded, representing the fact that an arbitrarily large measurement will only effect an estimate to a bounded extent in the large-data limit; in fact, the Student’s $t$ influence function redescends to 0, meaning that the larger the outliers, the less the estimate will be affected.

Bayesian point estimators may also be analyzed in this manner. The interpretation of robustness to outlier measurements can be extended to them.
Figure 2.2: The influence functions of the Gaussian and Student’s t distributions (the latter with two different values of the parameter $\nu$). Location and scale parameters are set to $\mu = 0$ and $\sigma^2 = 1$, respectively.
This chapter considers algorithms for maximum likelihood frequency estimation of a harmonic signal. Maximum likelihood is a popular approach for frequency estimation; see, e.g., the work by Stoica et al. (1994), Roy and Kailath (1989), and Williamson et al. (1994). Alternatives include subspace methods (Schmidt, 1982) and correlation methods (de Cheveigne and Kawahara, 2002). An overview of algorithms for frequency estimation is available in the book by Quinn and Hannan (2001). The extent of this chapter is limited to methods closely related to maximum likelihood estimation (described in general in Section 2.3.1) and the related periodogram estimation methods.

Frequency estimation algorithms based on Gaussian noise models are presented, and the well known periodogram estimator of frequency is derived. The chapter then proceeds to methods which assume non-Gaussian noise models. Katkovnik’s frequency estimation method (Katkovnik, 1998, 1999) is extended to harmonic signals, and a variant of it is proposed for real valued signals. An EM-based frequency estimation algorithm using a Student’s $t$ noise model is proposed.

### 3.1 Signal Model

The signal $y_k$ is assumed to have fundamental frequency $\omega$. There are $M$ harmonics with frequency $m\omega$, $m = 1, \ldots, M$. This chapter generally uses the signal model

\[
y_k = \sum_{m=1}^{M} \text{Re} \left( e^{im\omega k T} (\alpha^{(m)} + i\beta^{(m)}) \right) + e_k
\]

\[
= \sum_{m=1}^{M} \left( \alpha^{(m)} \cos(m\omega k T) - \beta^{(m)} \sin(m\omega k T) \right) + e_k, \quad k = 1, \ldots, N, \quad (3.1)
\]
where $k$ is the discrete time index, $T$ is the sample time, $y_k$ are measurements, and $e_k$ denotes the noise. The parameters $\alpha^{(m)}$ and $\beta^{(m)}$ are the real and imaginary components of the phasor $\alpha^{(m)} + i\beta^{(m)}$, together representing the phase offset and amplitude of each harmonic $m = 1, \ldots, M$. Assume that $e_k$ is independent and identically distributed (i.i.d.) according to a known distribution

$$e_k \sim p_e(e_k), \quad (3.2)$$

where $E e_k = 0$.

Collect the phasor components into the vector

$$Z = [\alpha^{(1)}, \beta^{(1)}, \ldots, \alpha^{(M)}, \beta^{(M)}]^T, \quad (3.3)$$

and the sinusoidal terms into

$$A_k(\omega) = \begin{bmatrix} \cos(\omega k T) - \sin(\omega k T) \\ \vdots \\ \cos(M\omega k T) - \sin(M\omega k T) \end{bmatrix}, \quad (3.4a)$$

$$A(\omega) = \begin{bmatrix} A_1(\omega) \\ \vdots \\ A_N(\omega) \end{bmatrix}. \quad (3.4b)$$

The equation (3.1) can then be rewritten in shorthand as

$$y_k = A_k(\omega)Z + e_k, \quad (3.5)$$

or in vectorized form as

$$Y = A(\omega)Z + E, \quad (3.6)$$

where $Y = [y_1, \ldots, y_N]^T$ and $E = [e_1, \ldots, e_N]^T$.

The maximum likelihood problem for frequency estimation can now be formulated. The joint maximum likelihood estimate of the phasor parameters $Z$ and the fundamental frequency $\omega$ is given by maximizing the log-likelihood of the data $Y$ (Quinn and Hannan, 2001);

$$\hat{\omega}, \hat{Z} = \arg \max_{\omega, Z} \log p(Y|\omega, Z) \quad (3.7a)$$

where the log-likelihood function is

$$\log p(Y|\omega, Z) = \sum_{k=1}^{N} \log p(y_k|\omega, Z) = \sum_{k=1}^{N} \log p_e(y_k - A_k(\omega)Z). \quad (3.7b)$$

Several observations can be made. It can be seen that $Z$ enters linearly in the mean $A_k(\omega)Z$ of the distribution $p(y_k|\omega, Z)$, and that the frequency $\omega$ enters in a nonlinear manner. The optimization problem (3.7a) will in general turn out to have many local maxima. This brings forth the need to use algorithms which are able to handle this problem. Since $Z$ enters linearly into the mean, it should be possible to estimate it conditionally on $\omega$. Then, $\omega$ must be estimated using an algorithm which is resistant to getting stuck in local maxima, such as grid-based search algorithms. This estimate may then be refined using a local search.
3.2 Frequency Estimation with Gaussian Noise Model

This section describes frequency estimation with a Gaussian noise model for $e_k$, i.e., when $p_e(e_k) = \mathcal{N}(e_k|0, \sigma^2)$. In this case,

$$\log p_e(e_k) = -\frac{e_k^2}{2\sigma^2} + \text{const.} \quad (3.8)$$

The maximum likelihood problem (3.7) can then be written as a nonlinear least squares problem

$$(\hat{\omega}, \hat{Z}) = \arg \min_{\omega, Z} \sum_{k=1}^{N} (y_k - A_k(\omega)Z)^2. \quad (3.9)$$

This problem has been well studied; for an overview, see the book by Quinn and Hannan (2001). Below, the numerical solution of this problem is discussed.

3.2.1 Maximum Likelihood Estimation

The maximum likelihood problem (3.9) with a Gaussian noise model is now further studied, and a partially analytic solution is derived. The phasor parameter vector $\hat{Z}(\omega)$ given $\omega$ can be computed as the solution to a linear least squares problem as discussed by, e.g., Quinn and Hannan (2001). The problem (3.9) can be rewritten in a vectorized form using (3.6) as

$$(\hat{\omega}, \hat{Z}) = \arg \min_{\omega, Z} \sum_{k=1}^{N} (y_k - A_k(\omega)Z)^2 = \arg \min_{\omega, Z} (Y - A(\omega)Z)^T(Y - A(\omega)Z). \quad (3.10)$$

Conditionally on $\omega$, the maximum likelihood estimator $\hat{Z}(\omega)$ is given by

$$\hat{Z}(\omega) = \arg \min_Z (Y - A(\omega)Z)^T(Y - A(\omega)Z)$$

$$= \left( A(\omega)^TA(\omega) \right)^{-1} A(\omega)^TY$$

$$= D(\omega)^{-1} F(\omega), \quad (3.11)$$

where the matrices

$$D(\omega) = A(\omega)^TA(\omega), \quad (3.12a)$$

$$F(\omega) = A(\omega)^TY, \quad (3.12b)$$
are defined in order to simplify the notation. Substitution of this solution into (3.10) yields the nonlinear least squares problem for $\omega$ given by

$$\hat{\omega} = \arg \min_{\omega} (Y - A(\omega) \hat{Z}(\omega))^T (Y - A(\omega) \hat{Z}(\omega))$$

$$= \arg \min_{\omega} Y^T Y - F(\omega)^T D(\omega)^{-1} F(\omega)$$

$$= \arg \max_{\omega} F(\omega)^T D(\omega)^{-1} F(\omega). \quad (3.13)$$

The first step is to rewrite the objective function in terms of $D(\omega)$ and $F(\omega)$, and the second one is to disregard the additive constant $Y^T Y$, and rewrite the problem as a maximization problem. This expression is a univariate optimization problem, but there are many local optima, which makes it difficult to find a global optimum without conducting a grid search.

### 3.2.2 Periodogram Estimation

Periodogram estimation is a way to approximately solve the Gaussian maximum likelihood frequency estimation problem (3.13) in a computationally efficient way that involves the periodogram. This presentation follows Quinn and Hannan (2001), where further details can be found. Under general conditions (a sufficient condition being the Nyquist criterion $M\omega T \leq \pi$ (Quinn and Hannan, 2001)), the matrix $D(\omega)$ approaches a scaled identity, $\frac{2}{N} D(\omega) \to I$ as $N \to \infty$. The matrix elements behave asymptotically as

$$(D(\omega))_{m_1,m_2} = \begin{cases} 
N/2 + \mathcal{O}(1), & m_1 = m_2, \\
\mathcal{O}(1), & \text{otherwise} 
\end{cases} \quad (3.14)$$

where $\mathcal{O}(1)$ represents a function of $N$ which is bounded. Thus, (3.13) can be approximately minimized when $N$ is large by finding the maximum

$$\hat{\omega} = \arg \max_{\omega} \frac{2}{N} F(\omega)^T F(\omega), \quad (3.15)$$

which avoids the matrix inversion in (3.13). Furthermore, note that

$$F(\omega) = A(\omega)^T Y = \begin{bmatrix}
\sum_{k=1}^{N} y_k \cos(\omega k T) \\
\sum_{k=1}^{N} y_k \sin(\omega k T) \\
\vdots \\
\sum_{k=1}^{N} y_k \cos(M\omega k T) \\
\sum_{k=1}^{N} y_k \sin(M\omega k T)
\end{bmatrix}. \quad (3.16)$$
3.2 Frequency Estimation with Gaussian Noise Model

The objective function can then be simplified as follows (Quinn and Hannan, 2001) using the rule $x^2 + y^2 = |x + iy|^2$,

$$\frac{2}{N} F(\omega)^T F(\omega) = \frac{2}{N} \sum_{m=1}^{M} \left( \left( \sum_{k=1}^{N} y_k \cos(m \omega_k T) \right)^2 + \left( \sum_{k=1}^{N} -y_k \sin(m \omega_k T) \right)^2 \right)$$

$$= \frac{2}{N} \sum_{m=1}^{M} \left| \sum_{k=1}^{N} y_k e^{-im \omega_k T} \right|^2$$

$$= \frac{2}{N} \sum_{m=1}^{M} |Y(m \omega)|^2,$$

(3.17)

where each summand involves the truncated discrete time Fourier transform (DTFT), $Y(\omega) = \sum_{k=1}^{N} y_k e^{-ik \omega T}$, so that $|Y(\omega)|^2$ is the periodogram of $Y$. This is an intuitively appealing result since for periodic signals, periodogram peaks are well known to correspond to frequencies of the periodic signal.

**Periodogram Discretization using Fast Fourier Transform**

The function $Y(\omega)$ in the objective function (3.17) may be efficiently computed using the fast Fourier transform (FFT) (Quinn and Hannan, 2001). The FFT provides the function values $Y(\omega_j)$ over a grid $(\omega_j)_{j=1}^{N}$ of size $N$ over $\omega$. Using these values, a search for a global maximum over the discrete grid is then performed. The complete estimation procedure has the computational cost $O(N \log N)$ due to the Fourier transform. In contrast, evaluating the objective function (3.13) at $N$ points has a computational cost of $O(N^2 M^3)$, where $M$ is the number of harmonics (Golub and Van Loan, 2012). Thus, it is straightforward to see that the periodogram approximation is more computationally efficient. A short description of the method is available in Algorithm 3.1.

**Subharmonic Summation (SHS)**

The method of subharmonic summation (SHS) (Hermes, 1988) has been proposed as a variation of the discretized periodogram frequency estimator. The FFT output has a linear frequency scale. In order to compute the objective function (3.17) with higher frequency resolution, the FFT output may be interpolated. Hermes (1988) also proposes using a log-frequency scale in order to more efficiently compute multiples $m \omega$ of the frequency.

Change the frequency variable to $u = \log \omega$, and interpolate $Y(\omega_j)$ to the log-frequency scale $\tilde{Y}(u)$ using linear interpolation. Then,

$$\sum_{m=1}^{M} |Y(m \omega)|^2 \approx \sum_{m=1}^{M} |\tilde{Y}(u + \log m)|^2.$$

(3.18)

Hence, SHS computes the magnitude spectrum of the signal and interpolates the frequency to a logarithmic scale. The objective function is then given by a sum of
Algorithm 3.1 Discretized periodogram frequency estimation (Hermes, 1988; Quinn and Hannan, 2001)

Require: Measurements $Y$, number of harmonics $M$, and frequency grid $(\omega_j)_{j=1}^{NF}$.

- Compute $(Y(\omega_j))_{j=1}^{NF}$ using the FFT.
- Set $\hat{\omega}$ as the maximizer of (3.17) or (3.18).
- Optionally, refine $\hat{\omega}$ using (3.13).

return $\hat{\omega}$

translated versions of the interpolated spectrum. The estimate is given by

$$\hat{\omega} = \arg \max_{\omega = e^u} \sum_{m=1}^{M} \left| \tilde{Y}(u + \log m) \right|^2. \quad (3.19)$$

See Algorithm 3.1.

3.3 Frequency Estimation with Non-Gaussian Noise Model

In the case when the true distribution is non-Gaussian, it is still possible to apply methods derived assuming Gaussian noise. This is equivalent to using the least squares principle for estimation (van Trees, 1968). However, the performance may be notably worse when the noise is heavy-tailed (Huber, 1981; Ljung, 1999). If the true distribution, or a distribution close to the true one, is used, the theoretical properties of maximum likelihood method (see Section 2.3.1) indicate that estimation error is reduced. Frequency estimation for non-Gaussian noise distributions is discussed in, for example, the work by Katkovnik (1999) and Li (2010).

3.3.1 Katkovnik’s Method with Extensions

Katkovnik (1998, 1999) proposes a framework in order to consider maximum likelihood frequency estimation with general noise distributions. In the papers, Katkovnik considers single-tone frequency estimation (number of harmonics $M = 1$), with a complex signal model

$$\tilde{y}_k = C^{(1)}e^{ik\omega T} + \tilde{e}_k, \quad (3.20)$$

where $C^{(1)}$ corresponds to $\alpha^{(1)} + i\beta^{(1)}$ in (3.1). The real part $\text{Re}(\tilde{e}_k)$ and the imaginary part $\text{Im}(\tilde{e}_k)$ of the estimation error are modeled as independent and identically distributed according to

$$\text{Re}(\tilde{e}_k) \overset{\text{i.i.d.}}{\sim} p_e(e), \quad (3.21)$$

$$\text{Im}(\tilde{e}_k) \overset{\text{i.i.d.}}{\sim} p_e(e), \quad (3.22)$$
3.3 Frequency Estimation with Non-Gaussian Noise Model

with corresponding negative log-likelihood

\[ L(e) = -\log p_e(e). \tag{3.23} \]

This section proposes two extensions of Katkovnik’s method. In order to obtain
the original method of Katkovnik (1999), set \( M = 1 \) in the method for complex
harmonic signals described below.

**Complex Harmonic Signals**

A generalization of Katkovnik’s method to multiple-tone frequency estimation
(number of harmonics \( M > 1 \)) for complex signals is now described. The complex
periodic harmonic signal \( \tilde{y}_k \) is now modeled as

\[ \tilde{y}_k = \sum_{m=1}^{M} C^{(m)} e^{ikm\omega T} + \bar{\epsilon}_k, \tag{3.24} \]

where \( C^{(m)} \) corresponds to \( \alpha^{(m)} + i\beta^{(m)} \) in (3.1). The maximum likelihood solution
is given by

\[ (\hat{\omega}, \hat{C}) = \arg \min_{\omega, C} J(\omega, C), \tag{3.25a} \]

where \( C = [C^{(1)}, C^{(2)}, \ldots, C^{(M)}]^{\top} \) is a complex vector of the phasors \( C^{(m)} \) of the
different harmonics, corresponding to the real vector \( Z \) in (3.3). The cost function
is the sum of negative log-likelihoods of the prediction errors,

\[ J(\omega, C) = \sum_{k=1}^{N} \left( L(\text{Re}(\epsilon_k(\omega, C))) + L(\text{Im}(\epsilon_k(\omega, C))) \right), \tag{3.25b} \]

\[ \epsilon_k(\omega, C) = \tilde{y}_k - \sum_{m=1}^{M} C^{(m)} e^{ikm\omega T}, \tag{3.25c} \]

where \( L \) is given in (3.23).

Katkovnik (1999) proposes an iterative method for solving the minimization
problem (3.25) with respect to \( C \) for a fixed \( \omega \) when \( M = 1 \). The proposed
generalization for \( M > 1 \) is described below. The objective function is rewritten
as

\[ J(\omega, C) = \sum_{k=1}^{N} \gamma_k(\omega, \epsilon_k(\omega, C))|\epsilon_k(\omega, C)|^2, \tag{3.26a} \]

\[ \gamma_k(\omega, \epsilon_k) = \frac{L(\text{Re}(\epsilon_k)) + L(\text{Im}(\epsilon_k))}{|\epsilon_k|^2}. \tag{3.26b} \]

An iterative algorithm is carried out as follows. At each iteration \( j \), set

\[ \gamma_k^{(j)}(\omega) = \frac{L(\text{Re}(\epsilon_k^{(j-1)}(\omega))) + L(\text{Im}(\epsilon_k^{(j-1)}(\omega)))}{|\epsilon_k^{(j-1)}(\omega)|^2} \tag{3.27} \]
as constants depending on the error in the previous iteration

$$
\hat{\varepsilon}_k^{(j-1)}(\omega) = \hat{y}_k - \sum_{m=1}^{M} C_{j-1}^{(m)} e^{ikm\omega T}.
$$

A quadratic approximation of the objective function (3.26a) at iteration $j$ can now be formulated as

$$
J^{(j)}(\omega, C) = \sum_{k=1}^{N} \gamma_k^{(j)} |\hat{\varepsilon}_k(\omega, C)|^2.
$$

The value of $C$ minimizing (3.29) is then given by

$$
\arg\min_{\hat{C}} J^{(j)}(\omega, C) = C_j(\omega),
$$

$$
C_j(\omega) = D_j(\omega)^{-1} F_j(\omega),
$$

$$
(D_j(\omega))_{m_1, m_2} = \sum_{k=1}^{N} \gamma_k^{(j)} e^{i(m_2 - m_1)\omega k T},
$$

$$
(F_j(\omega))_m = \sum_{k=1}^{N} \gamma_k^{(j)} \hat{y}_k e^{-im\omega k T},
$$

where $m$, $m_1$ and $m_2$ are indices ranging from 1 to $M$.

The generalization of Katkovnik (1999)'s method is achieved by iteratively computing (3.30), (3.28) and (3.27) until convergence, yielding a vector of phasors $\hat{C}(\omega)$. The iteration is initialized by setting $\gamma_k^{(1)} = 1$ for all $k$. Convergence properties are still to be analyzed, but empirical results indicate that the method converges. The iterations are done on a discrete grid of points $(\omega_j)_{j=1}^{N_F}$ over the frequency $\omega$, so that the final estimate $\hat{\omega}$ is given by

$$
\hat{\omega} = \arg\min_{\omega \in (\omega_j)} J(\omega, \hat{C}(\omega)).
$$

The method is summarized in Algorithm 3.2.

If the available measurements are not in complex form (3.24), the analytic signal may be computed using the discrete Hilbert transform (DHT) (Kak, 1977), which is a linear transformation approximating an analytic continuation $\hat{y}_k$ of $y_k$. It is numerically computed using discrete Fourier transforms (Kak, 1977),

$$
\hat{y}_k = \frac{1}{N} \sum_{j=1}^{N} e^{2\pi ij k/N} s_j \sum_{\ell=1}^{N} e^{-2\pi ij \ell/N} y_\ell,
$$

with the scaling

$$
s_j = \begin{cases} 
1, & j = 1, \quad j = \frac{N}{2} + 1, \\
2, & 1 < j \leq \frac{N}{2}, \\
0, & \frac{N}{2} + 1 < j \leq N.
\end{cases}
$$
3.3 Frequency Estimation with Non-Gaussian Noise Model

This quantity may be used as an approximation of measurements which truly follow (3.24), but a linear transformation from a space of dimension $N$ to one of dimension $2N$, with $N$ real components and $N$ imaginary components, does not preserve the i.i.d. property of the noise.

**Real Harmonic Signals**

The complex harmonic method outlined above does not inherently consider real valued signals. Real valued measurements can be extended to complex valued ones using the Hilbert transform, but this does not preserve the i.i.d. property of the noise. This section proposes a variation of the method using the real valued harmonic signal model

$$y_k = A_k(\omega)Z + \epsilon_k,$$  \hspace{1cm} (3.33)

with notation as defined in Section 3.1. The maximum likelihood solution for $(\omega, Z)$ is given by

$$(\hat{\omega}, \hat{Z}) = \arg \min_{\omega, Z} J(\omega, Z),$$  \hspace{1cm} (3.34a)

where $J(\omega, Z)$ is the sum of log-likelihoods of the prediction errors,

$$J(\omega, Z) = \sum_{k=1}^{N} L(\epsilon_k(\omega, Z)),\hspace{1cm} (3.34b)$$

$$\epsilon_k(\omega, Z) = y_k - A_k(\omega)Z.$$  \hspace{1cm} (3.34c)

Analogously to (3.26), the objective function is rewritten in a least squares form as

$$J(\omega, Z) = \sum_{k=1}^{N} \gamma_k(\omega, \epsilon_k(\omega, Z))\epsilon_k(\omega, Z)^2,$$ \hspace{1cm} (3.35a)

$$\gamma_k(\omega, \epsilon_k) = \frac{L(\epsilon_k)}{\epsilon_k^2}.$$ \hspace{1cm} (3.35b)

An iterative algorithm is now outlined, in analogy with the complex variation. At each iteration $j$, set

$$\gamma_k^{(j)}(\omega) = \frac{L(\text{Re}(\epsilon_k^{(j-1)}(\omega))) + L(\text{Im}(\epsilon_k^{(j-1)}(\omega)))}{|\epsilon_k^{(j-1)}(\omega)|^2}$$ \hspace{1cm} (3.35c)

as constants, where

$$\epsilon_k^{(j-1)}(\omega) = \tilde{y}_k - \sum_{m=1}^{M} Z_j^{(m)}e^{ik\omega T}.$$ \hspace{1cm} (3.35d)

The quadratic approximation of the objective function (3.35a) at iteration $j$ is now

$$J^{(j)}(\omega, Z) = \sum_{k=1}^{N} \gamma_k^{(j)}|\epsilon_k(\omega, Z)|^2.$$ \hspace{1cm} (3.35e)
Algorithm 3.2 Generalizations of Katkovnik (1999)’s method to complex or real valued harmonic signals

Require: Complex measurements \( \{\tilde{y}_k\}_{k=1}^N \) or real measurements \( \{y_k\}_{k=1}^N \).

Require: Number of harmonics \( M \), grid \( (\omega_{\ell})_{\ell=1}^{N_F} \), negative log-likelihood \( L(\varepsilon) \).

Initialize \( \gamma^{(1)}(\omega_{\ell}) = 1 \).

for \( j = 1, \ldots, N_F \) do
  while \( C \) or \( Z \) not converged do
    Complex variation: Sequentially update \( C_j(\omega_{\ell}), \varepsilon^{(j)}(\omega_{\ell}), \) and \( \gamma^{(j)}(\omega_{\ell}) \) using (3.30), (3.28), and (3.27).
    Real variation: Sequentially update \( Z_j(\omega_{\ell}), \varepsilon^{(j)}(\omega_{\ell}), \) and \( \gamma^{(j)}(\omega_{\ell}) \) using (3.36), (3.35d), and (3.35c).
  end while

end for

Set \( \hat{\omega} \) using (3.31) or (3.37).

Optionally, refine the estimate using (3.25) or (3.34).

return \( \hat{\omega} \)

The solution for \( Z \) minimizing \( J^{(j)}(\omega, Z) \), when the values \( \gamma_k \) are kept constant, is

\[
\arg\min_{Z} J^{(j)}(\omega, Z) = Z_j(\omega), Z_j(\omega) = D_j(\omega)^{-1} F_j(\omega), \quad (3.36a)
\]

\[
D_j(\omega) = A_j(\omega)^T G_{\gamma^{(j)}} A_j(\omega), \quad (3.36b)
\]

\[
F_j(\omega) = A_j(\omega)^T G_{\gamma^{(j)}} Y, \quad (3.36c)
\]

where \( G_{\gamma^{(j)}} \) is a diagonal matrix with elements \((G_{\gamma})_{k,k} = \gamma^{(j)}_k\). As in Section 3.3.1, the equations (3.36), (3.34c) and (3.35b) are iterated until convergence, with \( \gamma^{(1)}_k \) initialized to 1, over a grid \( (\omega_j)_{j=1}^{N_F} \) over \( \omega \). The algorithm proceeds to set

\[
\hat{\omega} = \arg\min_{\omega \in (\omega_j)} J(\omega, \hat{Z}(\omega)), \quad (3.37)
\]

similarly to (3.31). A summary of the method is available in Algorithm 3.2.

3.3.2 Expectation Maximization with Student’s t Noise Model

This section proposes a frequency estimation method using a heavy-tailed noise model based on the EM method, see Section 2.3.2. The objective of the algorithm is to handle outliers in a more computationally efficient way than the generalized Katkovnik methods described in Section 3.3.1. It assumes that \( e_k \) originates from the Student’s t distribution described in Section 2.1.5, \( p_e(e_k) = T(e_k|0, \sigma^2, \nu) \). The method exploits the latent formulation of a Student’s t variable as a mixture of Gaussians with Gamma distributed precision factors, as discussed in Section 2.1.5.
Since this factorization contains a Gaussian component, the computationally inexpensive periodogram maximization method (Section 3.2.2) can be used as a subcomponent, while being able to accommodate heavy-tailed noise.

The Student’s $t$ distribution can be written as a mixture of Gaussians, as discussed in Section 2.1.5. Augment a latent variable $\lambda_k$ for each measurement in order to write each measurement likelihood as such a mixture of Gaussians, yielding

$$p(y_k|\omega, Z) = \int p(y_k|\lambda_k, \omega, Z)p(\lambda_k|\omega, Z)dz,$$

$$p(y_k|\lambda_k, \omega, Z) = \mathcal{N}(y_k|A_k(\omega)Z, \sigma^2/\lambda_k),$$

$$p(\lambda_k|\omega, Z) = \mathcal{G}(\lambda_k|\nu/2, \nu/2).$$

with notation for $A_k(\omega)$ and $Z$ as defined in Section 3.1. The likelihood of the measurements $Y$ and the latents $\Lambda = (\lambda_k)_{k=1}^N$ can then be written as the product of the densities

$$p(Y, \Lambda|\omega, Z) = p(Y|\Lambda, \omega, Z)p(\Lambda|\omega, Z),$$

$$p(Y|\Lambda, \omega, Z) = \prod_{k=1}^N \mathcal{N}(y_k|A_k(\omega)Z, \sigma^2/\lambda_k),$$

$$p(\Lambda|\omega, Z) = \prod_{k=1}^N \mathcal{G}(\lambda_k|\nu/2, \nu/2).$$

The EM algorithm, as described in Section 2.3.2, may then be used to compute the maximum likelihood estimate for $(\omega, Z)$. There exists similar approaches for inferring Student’s $t$ distribution parameters, such as the work by Liu and Rubin (1995), but the method presented here uses a different statistical model for the mean in order to perform frequency estimation. The EM algorithm iteratively computes

$$Q(\omega, Z|\omega^{j-1}, \hat{Z}^{j-1}) = \mathbb{E}_{p(\Lambda|Y, \omega^{j-1}, \hat{Z}^{j-1})}\log p(Y, \Lambda|\omega, Z)$$

$$\hat{Z}^j, \hat{\omega}^j = \arg \max_{\omega, Z} Q(\omega, Z|\omega^{j-1}, \hat{Z}^{j-1})$$

until convergence. It is now shown that this formulation yields a nonlinear least squares problem to be solved in each iteration. The conditional distribution of $\Lambda$ given $Y$, $\omega$, and $Z$ is

$$p(\Lambda|Y, \omega, Z) \propto p(Y, \Lambda|\omega, Z)$$

$$\propto \prod_{k=1}^N e^{-\lambda_k(y_k-A_k(\omega)Z)^2/2\sigma^2} \lambda_k^{1/2} e^{-\lambda_k\nu/2} \lambda_k^{\nu/2-1},$$

where the proportionality sign is with respect to $\lambda_k$. This can be identified as a Gamma distribution,

$$p(\Lambda|Y, \omega, Z) = \prod_{k=1}^N \mathcal{G}(\lambda_k|a_k, b_k)$$
with parameters
\[
a_k = \frac{\nu + 1}{2}, \quad (3.43c)
\]
\[
b_k = \frac{\nu}{2} + \frac{1}{2\sigma^2}(y_k - A_k(\omega))Z^2. \quad (3.43d)
\]

**E step**

First, the \(Q\) function (3.42a) must be computed. It is given by
\[
Q(\omega, Z|\omega^{j-1}, \hat{Z}^{j-1}) = E_{p(\Lambda|Y, \hat{\omega}^{j-1}, \hat{Z}^{j-1})} \log p(Y, \Lambda|\omega, Z)
\]
\[
= -\frac{1}{2}(Y - A(\omega)Z)^T G_\Lambda (Y - A(\omega)Z), \quad (3.44)
\]
where \(G_\Lambda\) is a diagonal matrix with diagonal elements given by the expectations
\[
(G_\Lambda)_{k,k} = E_{p(\lambda_k|Y, \hat{\omega}^{j-1}, \hat{Z}^{j-1})} \frac{\lambda_k}{\sigma^2} = \frac{a_k}{b_k\sigma^2}. \quad (3.45)
\]
This can be seen by inspection of (3.43) and usage of the expectation of the Gamma distribution (see Section 2.1.5).

**M step**

The M step consists of solving (3.42b). This formulation is similar to the Gaussian case shown in (3.10), except for the inclusion of \(G_\Lambda\). Indeed, one may still estimate \(Z\) conditionally on \(\omega\), yielding
\[
\hat{Z}(\omega) = D(\omega)^{-1} F(\omega), \quad (3.46a)
\]
\[
D(\omega) = A(\omega)^T G_\Lambda A(\omega), \quad (3.46b)
\]
\[
F(\omega) = A(\omega)^T G_\Lambda Y. \quad (3.46c)
\]
The marginal least squares problem with respect to \(\omega\) is then given by
\[
\hat{\omega}^j = \arg \max_{\omega} F(\omega)^T D(\omega)^{-1} F(\omega), \quad (3.47)
\]
analogously to the Katkovnik equations (3.36), but with \(G_\Lambda\) instead of \(G_\gamma\). These two matrices can both be seen as measurement weighting matrices. The maximization problem for \(\omega^j\) can be shown to be connected to a scaled periodogram approach when \(N\) is large, since \(\frac{2}{N} D(\omega) \xrightarrow{P} \frac{\nu}{\nu + 1} I\) as \(N \to \infty\), where \(\xrightarrow{P}\) denotes convergence in probability. A proof is outlined below. The result suggests that a similar approach as in the Gaussian case described in Section 3.2.2 can be done, yielding an approximate solution for \(\hat{\omega}\) given by
\[
\hat{\omega}^j \approx \arg \max_{\omega} F(\omega)^T F(\omega) = \arg \max_{\omega} \sum_{m=1}^{M} |F(G_\Lambda Y)(m\omega)|^2. \quad (3.48)
\]
Thus, the only difference in the M step, as compared to the Gaussian periodogram solution, is that the measurements \(Y\) are adaptively rescaled using the diagonal matrix \(G_\Lambda\) before being Fourier transformed. Note that the approximation is poor when \(N\) is small, but it gets better when \(N\) increases.
Algorithm 3.3 Expectation-maximization frequency estimation with Student’s $t$ noise model

**Require:** Measurements $Y$, number of harmonics $M$, frequency grid $(\omega_j)_{j=1}^{N_F}$.

Initialize $a_k = b_k = \frac{\nu+1}{2}$.

**while** not converged **do**

Update $\hat{\omega}$ using (3.47) or (3.48).

Update $\hat{Z}$ using (3.46).

Update $p(\Lambda|Y, \hat{\omega}, \hat{Z})$ using (3.43).

**end while**

return $\hat{\omega}$

Algorithm summary and discussion

The proposed algorithm is arrived at by iterating the E and M steps detailed above until the method converges; see Algorithm 3.3. There are similarities between the EM and Katkovnik approaches, since they both consider maximum-likelihood frequency estimation using a heavy-tailed noise model by iteratively estimating the phasor parameters $Z$ or $C$, and estimating outlier weights $\gamma$ or $\lambda$. However, the EM algorithm is more computationally efficient. This is because it does not need to perform every iteration for a grid over $\omega$, but only considers a fixed $\omega$ at each iteration. This is because of the connection to the Gaussian periodogram approach using the mixture of Gaussians formulation of the Student’s $t$ distribution, (3.41).

Sketch of proof of convergence for the Student’s $t$ EM periodogram approximation

This section outlines a proof that the matrix $D(\omega)$ defined in (3.46b) satisfies

$$\frac{2}{N} D(\omega) \xrightarrow{P} \frac{\nu}{\nu+1} I$$

as $N \to \infty$, where $\xrightarrow{P}$ denotes convergence in probability. The objective is to show that $D(\omega)$ approaches a scaled identity matrix, so that the periodogram approximation (3.48) can be used in the Student’s $t$ EM algorithm.

It is assumed that

$$\hat{\omega} = \omega, \quad (3.49a)$$

$$\hat{Z} = Z, \quad (3.49b)$$

and that the noise parameters $(\sigma^2, \nu)$ are true. In this case,

$$y_k - A_k(\hat{\omega})\hat{Z} = e_k, \quad (3.50)$$

$$p(e_k) = T(e_k|0, \sigma^2, \nu). \quad (3.51)$$

By using (3.45) and (3.43), it can be seen that the diagonal elements of $G_{\Lambda}$ are given by

$$\frac{a_k}{b_k\sigma^2} = \frac{(\nu+1)/2}{(\nu/2 + e_k^2/(2\sigma^2))\sigma^2} = \frac{\nu + 1}{\nu\sigma^2 + e_k^2}. \quad (3.52)$$
The expected value $E_{p(e_k)} \frac{a_k}{b_k \sigma^2}$, as a function of the prediction error $d_k = A_k(\omega)Z - A_k(\hat{\omega})\hat{Z}$, for four values of $\nu$, when $\sigma^2 = 1$.

The objective of the assumption (3.49) is to ensure that $\frac{a_k}{b_k \sigma^2}$ has the same expected value for every $k$, where the expectation is with respect to $e_k$. Symbolic integration of the expected value in Mathematica yields $E_{p(e_k)} \frac{a_k}{b_k \sigma^2} = \frac{1}{\sigma^2}$. In order to numerically examine how realistic the assumption (3.49) is, the value of $E_{p(e_k)} \frac{a_k}{b_k \sigma^2}$ was computed numerically for a prediction error bias $d_k = A_k(\omega)Z - A_k(\hat{\omega})\hat{Z}$ using numerical integration, as seen in Figure 3.1. The value $E_{p(e_k)} \frac{a_k}{b_k \sigma^2}$ does not change more than 15% when varying $d_k$ within $-1 \leq d_k \leq 1$. The interval is chosen to represent realistic values of the model error. This may serve to reduce concerns that the assumption of $\hat{\omega} = \omega$ and $\hat{Z} = Z$ is overly restrictive.

The elements of $D(\omega)$ are given by

\begin{align*}
(D(\omega))_{2m_1-1,2m_2-1} &= \sum_{k=1}^{N} \cos(m_1 k \omega T) \cos(m_2 k \omega T) \frac{a_k}{b_k \sigma^2}, \\
(D(\omega))_{2m_1,2m_2-1} &= \sum_{k=1}^{N} -\sin(m_1 k \omega T) \cos(m_2 k \omega T) \frac{a_k}{b_k \sigma^2}, \\
(D(\omega))_{2m_1-1,2m_2} &= \sum_{k=1}^{N} -\cos(m_1 k \omega T) \sin(m_2 k \omega T) \frac{a_k}{b_k \sigma^2}, \\
(D(\omega))_{2m_1,2m_2} &= \sum_{k=1}^{N} \sin(m_1 k \omega T) \sin(m_2 k \omega T) \frac{a_k}{b_k \sigma^2}.
\end{align*}

Asymptotically, the different terms in (3.53) behave analogously, so the cosine-cosine case (3.53a) is considered in detail. In order to prove convergence, an
argument similar to the proof of the weak law of large numbers (Kallenberg, 2006) is used to extend the periodogram approximation proof of Quinn and Hannan (2001). Taking an expectation over the noise in (3.53a) yields

\[
\mathbb{E}_{p(e_k)} \frac{2}{N} (D(\omega))_{2m_1-1,2m_2-1} = \frac{2}{N} \sum_{k=1}^{N} \cos(m_1 k \omega T) \cos(m_2 k \omega T) \mathbb{E}_{p(e_k)} \frac{a_k}{b_k \sigma^2} \\
= \frac{1}{N} \sum_{k=1}^{N} (\cos((m_1 - m_2) k \omega T) + \cos((m_1 + m_2) k \omega T)) \frac{1}{\sigma^2} \\
= \begin{cases} 
\frac{1}{\sigma^2} + \mathcal{O}(1/N) & m_1 = m_2 \\
\mathcal{O}(1/N) & m_1 \neq m_2.
\end{cases}
\]  

(3.54)

One way to motivate the last equality is to approximate the sum as an integral, and observe that the integrand is periodic when \(m_1 \neq m_2\); this causes cancellation, yielding the asymptotic result. The equality holds when \((m_1 + m_2) \omega T \not\equiv 0 \pmod{2\pi}\), e.g., when all harmonics are below the Nyquist frequency, \(M \omega < \pi/T\); for more details, see Quinn and Hannan (2001).

Thus far, the derivation is similar to the Gaussian case, except for the addition of the expectation. However, it must be verified that the expression itself converges, not only its expectation. The variance is

\[
\text{Var}_{p(e_k)} \frac{2}{N} (D(\omega))_{2m_1-1,2m_2-1} = \frac{4}{N^2} \sum_{k=1}^{N} \cos^2(m_1 k \omega T) \cos^2(m_2 k \omega T) \text{Var}_{p(e_k)} \frac{a_k}{b_k \sigma^2} \\
= \frac{1}{N^2} \mathcal{O}(N) = \mathcal{O}(1/N)
\]  

(3.55)

because the random variables \(\frac{a_k}{b_k \sigma^2}\) are independent with finite variance, and the cosines are bounded. This argument also works with sine functions instead of cosine functions. Chebyshev’s inequality (Cramér, 1948) then yields

\[
P \left( \left| \frac{2}{N} (D(\omega))_{2m_1-1,2m_2-1} - \frac{1}{\sigma^2} [[m_1 = m_2]] + \mathcal{O}(1/N) \right| \geq \varepsilon \right) \leq \frac{1}{\varepsilon^2 \mathcal{O}(N)}
\]  

(3.56)

for all \(\varepsilon > 0\), where

\[
[[m_1 = m_2]] = \begin{cases} 
1 & m_1 = m_2, \\
0 & m_1 \neq m_2.
\end{cases}
\]  

(3.57)

Now, it is clear that given an \(\varepsilon > 0\), \(N\) can be chosen such that the right hand side of the inequality (3.56) is arbitrarily small. This establishes convergence in
probability,

\[
\frac{2}{N} (D(\omega))_{2m_1-1,2m_2-1} \xrightarrow{P} \frac{1}{\sigma^2} [[m_1 = m_2]]
\]  

(3.58)
as \(N \to \infty\). For the sine and sine-cosine terms in (3.53), it is straightforward to see that, in general, it also holds when \(N \to \infty\) that

\[
\frac{2}{N} (D(\omega))_{m_1,m_2} \xrightarrow{P} \frac{1}{\sigma^2} [[m_1 = m_2]]
\]  

(3.59)

so that

\[
\frac{2}{N} D(\omega) \xrightarrow{P} \frac{1}{\sigma^2} I.
\]  

(3.60)

This concludes the outline of the proof.

A numerical illustration of the speed of convergence is shown in Figure 3.2. The illustration shows sample values of the ratio of Frobenius norms,

\[
\frac{\|D(\omega)^{-1} - \frac{2\sigma^2}{N} I\|}{\|D(\omega)^{-1}\|},
\]  

(3.61)
as a function of \(N\). \(D(\omega)\) is computed using (3.46b) and (3.52), \(M = 3\), \(\nu = 3\), and \(\omega T = 0.3\). The relative error is small when \(N > 100\), but higher when \(N\) is smaller.
3.4 Simulated Example of Frequency Estimation with Outliers

In order to evaluate the performance of the methods in an idealized setting, a simulation study of the frequency estimation methods discussed in this chapter is carried out. Four frequency estimation methods are studied more closely:

- Periodogram frequency estimation (Algorithm 3.1)
- Generalized complex Katkovnik (Algorithm 3.2)
- Generalized real Katkovnik (Algorithm 3.2)
- Expectation-maximization frequency estimation (Algorithm 3.3)

3.4.1 Simulation Setup

For the purpose of the study, the measurements are generated using the model

\[ y_k = 0.5 \sin(\omega_k T + \pi) + 0.5 \sin(2\omega_k T + \pi/2) + 0.5 \sin(3\omega_k T) + e_k, \quad (3.62) \]

where the amplitudes of each harmonic are the same, but the phases differ by \( \pi/2 \) rad each. The angular frequency \( \omega \) is constant. The noise model is

\[ p(e_k) = (1 - d) N(e_k|0, \sigma^2_{\text{gen}}) + d N(e_k|0, K^2 \sigma^2_{\text{gen}}), \quad (3.63) \]

where \( d \) is the probability of an outlier and \( K \) is the factor by which the standard deviation of the noise is increased in an outlier (in short, outlier scale). The values of the model parameters are given in Table 3.1, while the variance \( \sigma^2_{\text{gen}} \) is varied in order to understand the performance of the methods for varying signal-to-noise ratios.

The frequency estimation methods were applied to a sequence of \( N = 200 \) data points. The used noise model parameters were estimated by simulating \( 10^4 \) noise variates from the true noise model, and computing the maximum-likelihood estimate of the noise model parameters for Gaussian and Student’s \( t \) noise models respectively. This is done in order to get an equitable comparison between the methods. Experimentally, the performance was reduced when these parameter values were perturbed.

3.4.2 Simulation Results and Discussion

In Figure 3.3, the RMSE and median absolute error for \( \omega \) is shown as a function of \( \sigma^2_{\text{gen}} \) for the tested methods. Figure 3.4 shows the same quantities as a function of \( K \). The higher the value of \( \sigma^2_{\text{gen}} \), the lower the signal-to-noise ratio (SNR). The two different measures of error are both illustrated, since the RMSE is highly influenced by large, uncommon estimation errors, while the median error is not. From the RMSE plot, it can be seen that the generalized Katkovnik methods have the lowest RMSE. For low values of \( \sigma^2_{\text{gen}} \), the EM method has equal performance. It also requires less computation time, as can be seen in Table 3.2.
In the median absolute error plot in Figure 3.3, the periodogram method, which does not model outliers, has the highest errors. The EM method appears comparatively better in this plot. This may be because the EM method works well most of the time; however, sometimes it finds an inaccurate local maximum. These poor estimates largely influence the RMSE plot but not the median error plot. The complex Katkovnik method has poor estimation accuracy when $\sigma^2_{\text{gen}}$ is small. This may be partially explained by the Hilbert transform approximation, see Section 3.3.1. In isolation, maximum likelihood-derived methods are expected to yield an estimation error tending to zero as the noise variance tends to zero. The complex Katkovnik method does not operate on the original signal, however, but on the Hilbert transformed one; thus, maximum likelihood-related properties do not apply.

In conclusion, it has been seen on this small simulated example that the frequency estimation methods discussed above can improve performance when the noise contains outliers. Variants of the generalized Katkovnik’s method have the lowest error. In higher SNR cases, the EM frequency estimation method performs equally well, while having lower computational cost.

**Table 3.1:** Parameters used for simulation and estimation in Section 3.4.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega$</td>
<td>30</td>
</tr>
<tr>
<td>$T$</td>
<td>0.01</td>
</tr>
<tr>
<td>$d$</td>
<td>0.1</td>
</tr>
<tr>
<td>$M$</td>
<td>3</td>
</tr>
<tr>
<td>$N$</td>
<td>200</td>
</tr>
</tbody>
</table>

**Table 3.2:** Average computational time for one execution of the frequency estimation methods EM (Algorithm 3.3), Periodogram (Algorithm 3.1), Real Katkovnik, and Katkovnik (Algorithm 3.2).

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Computational Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Periodogram</td>
<td>0.07 s</td>
</tr>
<tr>
<td>Generalized complex Katkovnik</td>
<td>0.58 s</td>
</tr>
<tr>
<td>Generalized real Katkovnik</td>
<td>0.49 s</td>
</tr>
<tr>
<td>EM</td>
<td>0.08 s</td>
</tr>
</tbody>
</table>
3.4 Simulated Example of Frequency Estimation with Outliers

Figure 3.3: RMSE (left) and median absolute error (right) of frequency estimation methods as a function of $\sigma^2_{\text{gen}}$ when $K = 30$, see (3.63). The frequency estimation methods are Periodogram (Algorithm 3.1), generalized complex and real Katkovnik (Algorithm 3.2), and EM (Algorithm 3.3).

Figure 3.4: RMSE (left) and median absolute error (right) of frequency estimation methods as a function of $K$ when $\sigma^2_{\text{gen}} = 10^{-2}$, see (3.63). The frequency estimation methods are Periodogram (Algorithm 3.1), generalized complex and real Katkovnik (Algorithm 3.2), and EM (Algorithm 3.3).
3.5 Batch-Oriented Frequency Tracking

The methods discussed in Sections 3.2–3.3 are, in practice, not only used for estimating the frequency of perfectly periodic signals. They can also be used to track a slowly varying frequency by grouping the measurements into batches of length $N_b$ or duration $T_b$. The fundamental frequency in each batch can then be estimated using one of the frequency estimation algorithms described in this chapter. The main advantage of the batch-oriented frequency tracking approach is that it is straightforward to understand and requires minimal additional complexity. However, there are a few disadvantages; clearly, any possible frequency variation within each batch is ignored, since the frequency estimation algorithms assume the frequency to be is constant. In addition, the frequency estimation algorithms will not consider phase and amplitude coherence between different batches. To partially remedy this problem, one may let the batches overlap. It is also suitable to apply a windowing function to the data (Quinn and Hannan, 2001).

One modification to the batched frequency tracking approach may be used if the estimated frequency $\hat{\omega}_k$ is known not to change more than $\Delta \omega$ between each batch $k$. It is then possible to search only for a frequency maximum within the range defined by this tolerance. This is described in Algorithm 3.4.

Further information on the batched approach is available from Hermes (1988), Lindgren et al. (2015), and Quinn and Hannan (2001).

---

**Algorithm 3.4** Batch-oriented frequency tracking

<table>
<thead>
<tr>
<th>Require:</th>
<th>Measurements $(y_k)_{k=1}^N$, frequency estimator $\hat{\omega}(y)$, batch length $N_b &lt; N$, optionally: frequency tolerance $\Delta \omega$.</th>
</tr>
</thead>
<tbody>
<tr>
<td>for $k = N_b, \ldots, N$ do</td>
<td>Compute $\hat{\omega}<em>k = \hat{\omega}((y_j)</em>{j=k-N_b+1}^k)$, optionally, with limits $</td>
</tr>
</tbody>
</table>
| end for | }
This chapter discusses models and algorithms which consider an inherently time-varying frequency between each sample instant. Three state-space models for this task are discussed, and one model suitable for further use is chosen. These models can be used for frequency tracking in any nonlinear filtering algorithm. Approaches used in the literature include the EKF, PF, and RBPF (Bittanti and Savaresi, 2000; Kim et al., 2008; La Scala and Bitmead, 1996; Ng et al., 2009), see Section 2.4. The focus here, however, is on the Rao-Blackwellized point mass filter (RBPMF). Variations and an extension of the RBPMF, suitable for frequency tracking, are proposed and discussed.

4.1 State-Space Models for Frequency Tracking

This section describes state-space models used for frequency tracking. The notation is similar to the one used in Chapter 3, but with subscript index $k$ representing the time dependence of the variables. Thus, $\omega_k$ is the fundamental frequency at time index $k$, and the measurements are still denoted $y_k$.

Any $v_k$ with superscript denotes a process noise on the corresponding state; for instance, the variable $v^\omega_k$ constitutes a process noise exciting the state $\omega_k$. Superscripts $v^{\omega,(l)}_k$ denote elements $l$ of a process noise vector. These process noises are assumed to be independent and identically distributed random variables. Parameters $Q$ with superscripts, such as $Q^\omega$, denote covariances of the continuous-time process noises.

4.1.1 Instant Phase Model with Multiple Phase States

This section discusses a state-space model for frequency tracking, with states consisting of the instant frequency $\omega_k$, instant phases $\phi_k^{(m)}$, and instant amplitudes
The RBPMF for Frequency Tracking

\(B_k^{(m)}\), for \(m = 1, \ldots, M\). The model has been used by, e.g., Ng et al. (2009). The dynamics is given by

\[
\begin{align*}
\phi_{k+1}^{(m)} &= \phi_k^{(m)} + mT\omega_k + mv_k^{\omega,(1)} + v_k^{\phi,(m)}, \\
B_{k+1}^{(m)} &= B_k^{(m)} + v_k^{B,(m)}, \quad m = 1, \ldots, M, \\
\omega_{k+1} &= \omega_k + v_k^{\omega,(2)},
\end{align*}
\]

with parameterized process noise covariances

\[
\text{Cov}
\begin{bmatrix}
v_k^{\omega,(1)} \\
v_k^{\omega,(2)}
\end{bmatrix}
= \begin{bmatrix}
T^3/3 & T^2/2 \\
T^2/2 & T
\end{bmatrix} Q^\omega,
\]

\[
\text{Var}
\begin{bmatrix}
v_k^{\phi,(m)}
\end{bmatrix} = TQ^{\phi}, \quad \text{and} \quad \text{Var}
\begin{bmatrix}
v_k^{B,(m)}
\end{bmatrix} = TQ^A.
\]

The parameterization of the covariances is chosen by discretizing the corresponding continuous-time stochastic process. Then, the measurement equation is given by

\[
y_k = \sum_{m=1}^M B_k^{(m)} \cos \phi_k^{(m)} + e_k.
\]

This model is appealing for its relative simplicity. Since phase and amplitude noise scales can be varied separately, the model is flexible. However, there are two main disadvantages:

- The \(M\) states \(\phi_k^{(m)}\) occur in a nonlinear manner in the measurement equation (4.1e). Linearized filters may not perform well if the state uncertainty is too high (Särkkä, 2013). This also makes Rao-Blackwellized methods scale poorly with \(M\).

- The states \(\phi_k^{(m)}\) which occur nonlinearly are angles, and are equivalent modulo 2\(\pi\). This creates complications in the Gaussian approximations inherent in both linearization-based and Rao-Blackwellized filters, because distributions such as the Gaussian are not suitable for circular estimation; other approaches exist (Agiomyrgiannakis and Stylianou, 2009; Lo, 1977; Tidefelt and Schön, 2009), but they are difficult to apply with Rao-Blackwellization.

### 4.1.2 Instant Phase Model with Single Phase State

This section describes a reformulation of the model in Section 4.1.1, using a single phase state instead of multiple ones. The reformulation is used in e.g., Kim et al. (2008). Only one state, the joint instant phase \(\phi_k\), occurs nonlinearly. To handle varying phase offsets in different harmonics, phasors \(\alpha_k^{(m)} + i\beta_k^{(m)}\) are introduced to represent the phase offset and amplitude of each harmonic. The instant amplitude of harmonic \(m\) is given by \(\left|\alpha_k^{(m)} + i\beta_k^{(m)}\right|\) and the instant phase
is given by $m\omega + \arg(\alpha_k^{(m)} + i\beta_k^{(m)})$. The dynamics is given by

$$
\phi_{k+1} = \phi_k + T\omega_k + v_k^{\omega,(1)} + v_k^\phi,
$$
(4.2a)

$$
\omega_{k+1} = \omega_k + v_k^{\omega,(2)},
$$
(4.2b)

$$
\alpha_{k+1}^{(m)} = \alpha_k^{(m)} + v_k^{\alpha^{(m)}},
$$
(4.2c)

$$
\beta_{k+1}^{(m)} = \beta_k^{(m)} + v_k^{\beta^{(m)}}, \quad m = 1, \ldots, M,
$$
(4.2d)

with noise variances $\text{Var}(v_k^\phi) = TQ^\phi$, $\text{Var}(v_k^{\alpha^{(m)}}) = \text{Var}(v_k^{\beta^{(m)}}) = TQ^{\alpha\beta}$, and

$$
\text{Cov}\left(\begin{bmatrix} v_k^{\omega,(1)} \\ v_k^{\omega,(2)} \end{bmatrix}\right) = \begin{bmatrix} T^3/3 & T^2/2 \\ T^2/2 & T \end{bmatrix} Q^\omega.
$$
(4.2e)

Then, the measurement equation is given by

$$
y_k = \sum_{m=1}^M \left( \alpha_k^{(m)} \cos(m\phi_k) - \beta_k^{(m)} \sin(m\phi_k) \right) + e_k,
$$
(4.2f)

which is the measurement model most similar to (3.1) for frequency estimation. This model is better suited for Rao-Blackwellized filters since only the scalar state $\phi_k$ appears nonlinearly. However, as in Section 4.1.1, the measurement equation is again nonlinear, and $\phi_k$ is an angle, which is unsuitable in conjunction with Gaussians. Furthermore, the model is overparameterized. A phase offset can be represented both by changing $\phi_k$, and by changing $(\alpha_k^{(m)} + i\beta_k^{(m)})_{m=1}^M$; such a reparameterization can be done at any given point in time.

### 4.1.3 Instant Phasor Model

This section shows how to simplify the model in Section 4.1.2 by reparameterization, so that it uses phasors to represent instant phases instead of initial phases. The resulting model was used by, among others, Bittanti and Savaresi (2000) and La Scala and Bitmead (1996). At each time step, reparameterize so that $\phi_k = 0$ by rotating $(\alpha_k^{(m)} + i\beta_k^{(m)})_{m=1}^M$. This means that the instant phase is $\arg(\alpha_k^{(m)} + i\beta_k^{(m)})$. The state $\phi_k$ can then be eliminated from the system description, yielding the dynamics

$$
\omega_{k+1} = \omega_k + v_k^{\omega},
$$
(4.3a)

$$
\begin{bmatrix} \alpha_{k+1}^{(m)} \\ \beta_{k+1}^{(m)} \end{bmatrix} = R(m\omega_k T) \begin{bmatrix} \alpha_k^{(m)} \\ \beta_k^{(m)} \end{bmatrix} + \begin{bmatrix} v_k^{\alpha^{(m)}} \\ v_k^{\beta^{(m)}} \end{bmatrix}, \quad m = 1, \ldots, M,
$$
(4.3b)

with process noise variances $\text{Var}(v_k^{\omega}) = TQ^\omega$, and $\text{Var}(v_k^{\alpha^{(m)}}) = \text{Var}(v_k^{\beta^{(m)}}) = TQ^{\alpha\beta}$. The measurement equation is then given by setting $\phi_k = 0$ in (4.2f), yielding

$$
y_k = \sum_{m=1}^M \alpha_k^{(m)} + e_k.
$$
(4.3c)
The rotation of the phasor \([\alpha_k^{(m)}, \beta_k^{(m)}]^\top\) is done through the rotation matrix

\[
R(\phi) = \begin{bmatrix}
\cos \phi & -\sin \phi \\
\sin \phi & \cos \phi
\end{bmatrix}.
\]

(4.4)

This model is suitable for Rao-Blackwellized filters. The only state which occurs nonlinearly is \(\omega_k\), which is a scalar. If \(m\omega_k T\) is small for all \(m\), a linearized filter may work well. However, note that it is less flexible than the model in Section 4.1.1, since the noise variances in amplitude and phase directions cannot be varied separately. This model will be used for recursive frequency tracking in the remainder of the thesis.

### 4.2 Extensions of The Rao-Blackwellized Point Mass Filter

This section discusses the RBPMF in relation to frequency tracking. The RBPMF is computationally demanding. Two methods for reducing computational cost are proposed, exploiting the structure of the model in Section 4.1.3. Furthermore, a variant of the RBPMF using a Student’s \(t\) measurement noise model is proposed, using variational approximate Rao-Blackwellization. This enables improved modeling of outliers in the measurements.

In Algorithm 2.8, the standard RBPMF (Smidl and Gasperin, 2013) is summarized. This can be compared with Algorithm 4.1, where the variations of the proposed RBPMF are seen.

#### 4.2.1 Modified RBPMF Merge Step

This section describes a variation of the RBPMF which reduces the memory requirements of the algorithm. The variation can be used in the special case that \(F^a(x^n_k) = 0\) when written in the CLSSM form (2.28). This variation can be used in conjunction with the final model in Section 4.1.3 when \(x^n_k = \omega_k\).

First, assume that (2.62) holds, i.e., that

\[
p(x_{k-1}|Y_{k-1}) = p(x^n_{k-1}|x^n_{k-1}, Y_{k-1})p(x^n_{k-1}|Y_{k-1}) \\
\approx \sum_{i=1}^N w_{k-1|k-1}^i \mathcal{N}(x^n_{k-1}|x^n_{k-1}, p^a_{k-1|k-1}) \delta(x_{k-1} - x^n_{k-1}).
\]

(4.5)

Due to the assumption \(F^a(x^n_k) = 0\), one can write

\[
p(x^n_{k}|x^n_{k}, x^n_{k-1}, Y_{k-1}) = p(x^n_{k}|x^n_{k-1}, Y_{k-1}) \\
= \mathcal{N}(\tilde{x}_{k|k-1}^a, \tilde{P}_{k|k-1}^a),
\]

(4.6)

where the parameters \(\tilde{x}_{k|k-1}^a\) and \(\tilde{P}_{k|k-1}^a\) do not depend on \(x^n_{k-1}\).
The variation consists of performing the merge step in the RBPMF before the measurement update, instead of after the measurement update. The time update and merge step are then jointly given by the equations

\[
\begin{align*}
  w_{k|k-1}^j & = \sum_{i=1}^{N_S} p_{k|k-1}^{(j,i)} w_{k-1|k-1}^i \\
  \hat{x}_{k|k-1}^{a,j} & = \frac{1}{w_{k|k-1}^j} \sum_{i=1}^{N_S} p_{k|k-1}^{(j,i)} w_{k-1|k-1}^i \hat{x}_{k|k-1}^{a,i} \\
  p_{k|k-1}^{a,j} & = \frac{1}{w_{k|k-1}^j} \sum_{i=1}^{N_S} p_{k|k-1}^{(j,i)} w_{k-1|k-1}^i \left( \hat{P}_{k|k-1}^{a,i} + \hat{x}_{k|k-1}^{a,i} \hat{x}_{k|k-1}^{a,i \top} \right) - \hat{x}_{k|k-1}^{a,j} \hat{x}_{k|k-1}^{a,j \top}
\end{align*}
\]

(4.7a)  
(4.7b)  
(4.7c)  
(4.7d)  

where the notation \(( \cdot )\) represents the same expression as the preceding parenthesis, and the shorthand for the transition probabilities between states is

\[
p_{k|k-1}^{(j,i)} = p(x_{k}^{n,j} | x_{k-1}^{n,i}).
\]

(4.7e)

The first expression (4.7c) for the covariance matrix update will prove useful later, while the second one (4.7d) is more numerically well behaved (Bierman, 2006; Searle and Khuri, 2017). The measurement update can be performed exactly as in a bootstrap RBPF (see Algorithm 2.7). It is not required to store and update all \(N_S^2\) means and covariance matrices \(\left( \hat{x}_{k|k-1}^{a,i}, P_{k|k-1}^{a,i} \right)_{i,j=1}^{N_S}\) (see Algorithm 2.8). This avoids notable computational and storage cost. The variation of the RBPMF described in this chapter is similar to the interacting multiple model (IMM) method (Blom and Bar-Shalom, 1988; Mazor et al., 1998); in fact, this can thus be seen as an alternate derivation of the IMM approach. The probability \(p_{k|k-1}^{(j,i)}\) then corresponds to the unnormalized transition probability from mode \(i\) to mode \(j\) in the IMM filter.

### 4.2.2 Fourier Transform in the Merge Step

This section studies how the computational cost of the RBPMF can be reduced further, under additional requirements of the CLSSM structure. The update equations in the RBPMF variant described in 4.2.1 are written as convolutions, which can be implemented using the fast Fourier transform (FFT) (Kay, 1993).

The structure that must be imposed on the CLSSM is that

\[
p(x_{k}^{n,j} = x_{k}^{n,i} | x_{k-1}^{n,j} = x_{k-1}^{n,i}) = p_{k|k-1}^{(j,i)} = p_{k|k-1}^{(j-i)}
\]

(4.8)

so that the probability of going from state \(x_{k-1}^{n,i}\) at time \(k - 1\) to \(x_{k}^{n,j}\) at time \(k\) only depends on the difference \(j - i\) between the grid indices. This applies, for
The proposed variations of the Rao-Blackwellized point mass filter (RBPMF)

**Algorithm 4.1**

**Require:** $p(x^n_0, x^n_0|Y_0) \approx \sum_{i=1}^{N_S} w^n_{0,i} \mathcal{N}(x^n_0|x^n_0, P^n_{0|0}) \delta(x^n_0 - x^n_0)$

for $k = 1, \ldots, N$ do

Optionally, update grid $x^{n,j}_k$, or set $x^{n,j}_k = x^{n,j}_{k-1}$.

for $j = 1, \ldots, N_S$ do

Linear prediction: Compute $p(x^n_k|x^n_{k-1}, Y_{k-1}) = \mathcal{N}(x^n_k|x^n_{k-1}, P^n_{k|k-1})$ using the Kalman filter prediction in Algorithm 2.3.

Merging step and nonlinear prediction: Compute $w^j_{k|k-1}$, $\hat{x}^a_j$, and $P^a_{k|k-1}$ using the index domain formulas in (4.7) or the index-frequency domain formulas in (4.10).

Linear and nonlinear measurement update: Compute $w^j_{k|k}$, $\hat{x}^a_j$, and $P^a_{k|k}$ using the RBPF measurement update in Algorithm 2.7 or the Student’s $t$ update in Algorithm 4.2.

end for

$p(x^n_k, x^n_k|Y_k) \approx \sum_{j=1}^{N_S} w^j_{k|k} \mathcal{N}(x^n_k|x^n_k, P^n_{k|k}) \delta(x^n_k - x^n_k)$

end for

Example, with the random walk dynamics

$$p^{(j,i)}_{k|k-1} = q_k(x^{n,j}_k - x^{n,i}_{k-1}), \quad (4.9)$$

for a (possibly time-variant) PDF $q_k$. The state grid $(x^{n,i}_k)_{i=1}^{N_S}$ must be equidistant with constant spacing $\Delta$; $x^{n,i}_k = x^{n,0}_k + i\Delta$, where $x^{n,0}_k$ is an offset which may be allowed to vary over the time index $k$.

Let $(:)$ denote vectorization over the given index; thus, the vector $w^j_{k|k}$ contains elements $w^j_{k|k}$ for $j = 1, \ldots, N_S$. Furthermore, let $*$ denote discrete convolution with respect to $(::)$, and let $\circ$ denote element-wise multiplication with respect to $(::)$. Now use (4.8) to rewrite (4.7) as convolutions

$$w^j_{k|k-1} = (p^{(j)}_{k|k-1} * w^{(j)}_{k-1|k-1})_j, \quad (4.10a)$$

$$\hat{x}^a_{k|k-1} = \frac{1}{w^j_{k|k-1}} \left( p^{(j)}_{k|k-1} * (w^{(j)}_{k-1|k-1} \circ \hat{x}^a_{k|k-1}) \right)_j \quad (4.10b)$$

$$P^a_{k|k-1} = \frac{1}{w^j_{k|k-1}} \left( p^{(j)}_{k|k-1} * \left( w^{(j)}_{k-1|k-1} \circ \left( P^a_{k|k-1} + \hat{x}^a_{k|k-1} \hat{x}^a_{k|k-1}^\top \right) \right) \right)_j$$

$$- x^{a,j}_{k|k-1} \hat{x}^{a,j}_{k|k-1}, \quad (4.10c)$$

where (4.7c) had to be used in order to formulate the equation as a convolution. It is well known (Oppenheim, 1999) that convolution can be implemented using the fast Fourier transform with computational complexity $O(N_S \log N_S)$ in the index-frequency domain, instead of the conventional index domain computation which
requires computational complexity $\mathcal{O}(N_S^2)$. This implies that the merging component of the RBPMF, and thus the filter, can be run at computational complexity $\mathcal{O}(N_S \log N_S)$.

**Numerical Issues in the Fourier Merge Step**

Numerical problems have been observed when using the Fourier time update described above. They occur when the scale variations in $w_{k-1|k-1}^{(c)}$ are large. To see this, consider the weight updates in the time domain. The time domain implementation only contains sums and multiplications by positive numbers, which are well-behaved for floating point numbers even if the scales of the various summands and factors are different. For instance, the time domain operations cannot yield a negative value of $w_{k|k-1}^{(j)}$ when carried out in floating point arithmetic. In the index-frequency domain, this is not guaranteed. In order to see this, consider that the Fourier transform is an orthonormal linear transformation (Kreyszig, 1989). Thus, $\mathcal{F} w_{k-1|k-1}^{(c)}$ is $w_{k-1|k-1}^{(c)}$ represented in a different base. The Fourier transform matrix $\mathcal{F}$, which represents the base change, is a matrix where every element has absolute value $|\mathcal{F}_{j,k}| = 1/\sqrt{N_S}$ for all entries $j, k$. Assume that each grid weight $w_{k-1|k-1}^{(j)}$ is numerically represented with floating-point error given by $\varepsilon(w_{k-1|k-1}^{(j)})$. A binary floating-point representation with $B$ bits in the significand will have $\varepsilon(w_{k-1|k-1}^{(j)}) \approx 2^{-B} w_{k-1|k-1}^{(j)}$. Assuming that numerical errors are approximately independent, a loose lower bound on the error can be derived as

$$\varepsilon(\mathcal{F} w_{k-1|k-1}^{(c)})^2 \approx \frac{1}{N_S} \sum_{j=1}^{N_S} \varepsilon(w_{k-1|k-1}^{(j)})^2 \approx \frac{2^{-2B}}{N_S} \sum_{j=1}^{N_S} (w_{k-1|k-1}^{(j)})^2 \geq \frac{2^{-2B}}{N_S^2}. \tag{4.11}$$

This implies that the absolute error in the numerical representation of every component of $\mathcal{F} w_{k-1|k-1}^{(c)}$ is approximately bounded below by the constant $2^{-B}/N_S$. This absolute error will be propagated when the inverse Fourier transform $\mathcal{F}^{-1} = \mathcal{F}^T$ is computed; yielding absolute numerical errors of the same order of magnitude in $w_{k+1|k}^{(c)}$. Since the absolute errors of the largest values of $w_{k-1|k-1}^{(c)}$ are propagated onto all values of $w_{k+1|k}^{(c)}$, the relative error can be large.

Figure 4.1 shows the numerical error caused by applying the forward and inverse FFT to a number sequence $w_{k-1|k-1}^{(c)}$. In this case, the maximum value of $w_{k-1|k-1}^{(c)}$ is 1, and the error $\varepsilon(\max_j w_{k-1|k-1}^{(j)}) \approx 10^{-16}$. The numerical error in the floating-point representation of the maximum value is propagated to all values, causing the errors in the dashed curve.

The same principle applies to the computation of $\hat{x}_{k|k-1}^{a,(c)}$ and $P_{k|k-1}^{a,(c)}$. Absolute numerical errors are propagated in the FFT, and whenever there is a large order of magnitude difference in $w_{k-1|k-1}^{(c)}$, small absolute numerical errors yield large relative numerical errors.
Numerical problems first arise in the computation of $P_{a,k|k-1}$. As previously discussed, (4.7d) is less numerically well behaved, but it must be used in the FFT formulation. Otherwise, the update for $P_{a,k|k-1}$ cannot be written as a convolution. Numerical experiments using double-precision floating-point numbers ($B = 52$) and $N_S = 200$ grid points indicate that problems arise when $\min_j w_{k|k-1}^j \lesssim 10^{-8}$ for the state-space model described in Section 4.1.3.

It is not clear how to preserve the behavior of the filter exactly. However, if a small modification of the filter dynamics is acceptable, one may replace the transition probability $p_{k+1|k}^{(j-i)}$ by

$$
\hat{p}_{k|k-1}^{(j-i)} = \max(p_{k|k-1}^{(j-i)}, \varepsilon)
$$

(4.12)

with a small $\varepsilon > 0$. The value of $\varepsilon$ is empirically selected such that the matrices $P_{k|k-1}$ remain positive-definite when the numerical Fourier computation of the RBPMF iterations are made. Note that the entire probability mass that is changed in the distribution is less than $N\varepsilon$.

To summarize, it has been shown that the update of the RBPMF, described in Section 4.2.1, can be further sped up using the FFT. This yields an improvement in the computational complexity of the filter from $\mathcal{O}(N_S^2)$ to $\mathcal{O}(N_S \log N_S)$. Numerical issues exist, but further approximations such as (4.12) can mitigate them.
4.2 Extensions of The Rao-Blackwellized Point Mass Filter

4.2.3 Student’s $t$ Measurement Noise Models with Variational Approximate Rao-Blackwellization

Rao-Blackwellized methods for Bayesian filtering usually assume a Gaussian additive measurement noise model. This section proposes a method for performing the measurement update which uses a more flexible Student’s $t$ noise model. This enables improved modeling of outliers in the measurements. In order to achieve this, a mean-field variational Bayes approximation is used in the measurement update.

As a start, assume that the predictive distribution

$$p(x^n_k | x^n_{k-1}) = \mathcal{N}(x^n_k | \hat{x}^n_{k|k-1}, P^n_{a|k-1})$$  \hspace{1cm} (4.13)

is available from the previous iteration of the filter. Note that the parameters $\hat{x}^n_{k|k-1}$ and $P^n_{a|k-1}$ are dependent on $x^n_k$, but this dependence is omitted for brevity in this section. The measurement model is given in (2.28c), repeated here as

$$y_k = h(x^n_k) + H(x^n_k) x^a_k + e_k,$$  \hspace{1cm} (4.14)

where the noise $e_k$ is modeled as originating from a Student’s $t$ distribution

$$e_k \sim T(0, \sigma^2, \nu).$$  \hspace{1cm} (4.15)

Assume that $H(x^n_k) \neq 0$, or equivalently, that $p(y_k | x^a_k, x^n_k) \neq p(y_k | x^n_k)$. Then, two key distributions are unavailable as analytic expressions: the marginal likelihood of $y_k$,

$$p(y_k | x^n_{k-1}) = \int p(y_k | x^a_k, x^n_k) p(x^a_k | x^n_k, Y_{k-1}) dx^a_k,$$  \hspace{1cm} (4.16)

and consequently, the filtering distribution of $x^a_k$,

$$p(x^a_k | x^n_k, Y_k) = \frac{p(y_k, x^a_k | x^n_k, Y_{k-1})}{p(y_k | x^n_k, Y_{k-1})}.$$  \hspace{1cm} (4.17)

This section proposes an approach to compute them approximately. The idea of using a variational mean-field approximation to consider Student’s $t$ noise in a filter is similar to Agamennoni et al. (2012), but in this method, all recursions are conditional on $x^n_k$. The primary contribution is to use the evidence lower bound $\mathcal{L}$ as an approximation

$$\log p(y_k | x^n_{k-1}) \approx \mathcal{L},$$  \hspace{1cm} (4.18)

which is then used to compute point mass weights $w^j_{k|k}$ in the point mass filter component as $w^j_{k|k} \propto p(y_k | x^n_{k-1}, Y_{k-1}) w^j_{k|k-1}$.

Variational approximate Rao-Blackwellization can be performed as follows. First, adjoin an additional latent variable $\lambda_k$ in the measurement update as described in Section 2.1.5, writing the Student’s $t$ distribution as a Gaussian-Gamma mixture. The joint distribution can be written as

$$p(y_k, x^a_k, \lambda_k | x^n_k, Y_{k-1}) = p(y_k | x^a_k, x^n_k, \lambda_k) p(x^a_k | x^n_k, Y_{k-1}) p(\lambda_k),$$  \hspace{1cm} (4.19)
and the desired density to approximate is

\[ p(x_k^a, \lambda_k | Y_k, x_k^n). \]

The approximation is performed using mean-field variational Bayes, as described in Section 2.4.4. In comparison to the notation there,

\[ \theta = \left[ \theta_1 \right] = \left[ x_k^a \right], \quad (4.20) \]

which corresponds to the variational mean-field assumption

\[ q(\theta) = q_{x_k^a}(x_k^a)q_{\lambda_k}(\lambda_k). \quad (4.21) \]

Note that \( q(\theta) \) has an implicit dependence on \( x_k^n \). When computed, \( q_{x_k^a}(x_k^a) \) will be an approximation of \( p(x_k^a | Y_k, x_k^n) \). This approximation is then used to compute the marginal measurement likelihood according to (4.18).

\section*{Approximation of \( p(x_k^a, \lambda_k | Y_k, x_k^n) \)}

Firstly, it will be shown how to compute the approximation

\[ q_{x_k^a}(x_k^a)q_{\lambda_k}(\lambda_k) \approx p(x_k^a, \lambda_k | Y_k, x_k^n) \quad (4.22) \]

using variational Bayes, see Section 2.4.4.

The factors of the joint distribution (4.19) are given by

\[ p(y_k|x_k^a, x_k^n, \lambda_k) = \mathcal{N}(y_k|h(x_k^n) + H(x_k^n)x_k^a, \sigma^2/\lambda_k), \quad (4.23a) \]

\[ p(x_k^n | x_k^a, Y_{k-1}) = \mathcal{N}(x_k^n|\hat{x}_k^{a|k-1}, P_{k|k-1}^{a}), \quad (4.23b) \]

\[ p(\lambda_k) = \mathcal{G}(\nu/2, \nu/2). \quad (4.23c) \]

The distribution \( p(x_k^a | Y_k, x_k^n) \) for the linear state is unavailable when using a Student’s \( t \) noise model. In order to approximate it, the variational mean-field approximation (4.22) is used. As detailed in Section 2.4.4, the approximation can be made accurate by minimizing the Kullback-Leibler divergence

\[ D_{KL}(q_{x_k^a}(x_k^a)q_{\lambda_k}(\lambda_k)||p(x_k^a, \lambda_k | x_k^n, Y_k)) \]

\[ = \int q_{x_k^a}(x_k^a)q_{\lambda_k}(\lambda_k) \log \frac{p(x_k^a, \lambda_k | x_k^n, Y_k)}{q_{x_k^a}(x_k^a)q_{\lambda_k}(\lambda_k)} dx_k^a d\lambda_k, \quad (4.24) \]

where \( q_{x_k^a}(x_k^a) \) and \( q_{\lambda_k}(\lambda_k) \) are arbitrary PDFs. In order to perform the variational updates, described in general in Algorithm 2.2, the logarithm of (4.19) must be used. It is given by

\[ \log p(x_k^a, \lambda_k, y_k | x_k^n, Y_{k-1}) = \]

\[ - \frac{\lambda_k}{2\sigma^2} (y_k - h(x_k^n)) - H(x_k^n)x_k^a)^2 - \frac{1}{2} \log \sigma^2 + \frac{1}{2} \log \lambda_k \]

\[ - \frac{1}{2} (x_k^a - \hat{x}_k^{a|k-1})^T P_{k|k-1}^{a}(x_k^a - \hat{x}_k^{a|k-1}) - \frac{1}{2} \log \text{det} P_{k|k-1}^{a} \]

\[ + \left( \frac{\nu}{2} - 1 \right) \log \lambda_k - \frac{\nu}{2} \lambda_k + \frac{\nu}{2} \log \frac{\nu}{2} - \log \Gamma\left(\frac{\nu}{2}\right), \quad (4.25) \]
where $P_{k|k-1}^{a,-1} = (P_{k|k-1}^a)^{-1}$. Each row in the right hand side of (4.25) corresponds to a density in (4.23).

As detailed in Section 2.4.4, the optimal $q_{x_k}^* (x_k^n)$ and $q_{\lambda_k}^* (\lambda_k)$ minimizing (4.24) satisfy the equations

$$
\log q_{x_k}^* (x_k^n) = E_{q_{x_k}^*} \left[ \log p(x_k^n, \lambda_k, y_k|x_k^n, Y_{k-1}) \right] + \text{const}, \quad (4.26a)
$$

$$
\log q_{\lambda_k}^* (\lambda_k) = E_{q_{x_k}^*} \left[ \log p(x_k^n, \lambda_k, y_k|x_k^n, Y_{k-1}) \right] + \text{const}. \quad (4.26b)
$$

These equations are now analyzed one by one. Expanding (4.26a) yields

$$
\log q_{x_k}^* (x_k^n) = E_{q_{x_k}^*} \left[ \log p(x_k^n, \lambda_k, y_k|x_k^n, Y_{k-1}) \right] + \text{const}
= - \frac{E_{q_{x_k}^*} [\lambda_k]}{2\sigma^2} (y_k - h(x_k^n) - H(x_k^n)x_k^n)^2 - \frac{1}{2} (x_k^n - \hat{x}_k^n) P_{k|k-1}^{a,-1} (x_k^n - \hat{x}_k^n) + \text{const}
= - \frac{1}{2} x_k^n a, x_k \left( \frac{E_{q_{x_k}^*} [\lambda_k]}{\sigma^2} H(x_k^n)^T H(x_k^n) + P_{k|k-1}^{a,-1} \right) x_k^n
+ x_k^n a, x_k \left( \frac{E_{q_{x_k}^*} [\lambda_k]}{\sigma^2} H(x_k^n)^T (y_k - h(x_k^n)) + P_{k|k-1}^{a,-1} \right) + \text{const}. \quad (4.27)
$$

This expression can be compared with the multivariate normal (see (2.15)),

$$
\log \mathcal{N}(x_k^n|\hat{x}_k^n, P_{k|k}^a) = - \frac{1}{2} (x_k^n - \hat{x}_k^n)^T P_{k|k}^{a,-1} (x_k^n - \hat{x}_k^n) + \text{const}
= - \frac{1}{2} x_k^n a, x_k P_{k|k}^{a,-1} x_k^n + x_k^n a, x_k P_{k|k}^{a,-1} \hat{x}_k^n + \text{const}, \quad (4.28)
$$

and this comparison yields that (4.27) can be written as

$$
q_{x_k}^* (x_k^n) = \mathcal{N}(x_k^n|\hat{x}_k^n, P_{k|k}^a), \quad (4.29a)
$$

$$
P_{k|k}^{a,-1} = \frac{E_{q_{x_k}^*} [\lambda_k]}{\sigma^2} H(x_k^n)^T H(x_k^n) + P_{k|k-1}^{a,-1}, \quad (4.29b)
$$

$$
P_{k|k}^{a,-1} \hat{x}_k^n = \frac{E_{q_{x_k}^*} [\lambda_k]}{\sigma^2} H(x_k^n)^T (y_k - h(x_k^n)) + P_{k|k-1}^{a,-1} \hat{x}_k^n. \quad (4.29c)
$$

This expression exactly coincides with the corresponding Gaussian Rao-Blackwellized filtering distribution (Schön et al., 2005) when substituting $E_{q_{x_k}^*} [\lambda_k]$ for 1.

The second equation (4.26b) can be expanded as

$$
\log q_{\lambda_k}^* (\lambda_k) = E_{q_{x_k}^*} \left[ \log p(x_k^n, \lambda_k, y_k|x_k^n, Y_{k-1}) \right] + \text{const}
= - \frac{\lambda_k}{2\sigma^2} E_{q_{x_k}^*} [(y_k - h(x_k^n) - H(x_k^n)x_k^n)^2]
- \frac{\lambda_k}{2} \frac{\nu}{2} \lambda_k + \left( \frac{\nu}{2} - \frac{1}{2} \right) \log \lambda_k + \text{const}. \quad (4.30)
$$
Comparing with a Gamma distribution (2.16),
\[ \log G(\lambda_k|a_k, b_k) = (a_k - 1) \log \lambda_k - b_k \lambda_k + \text{const}, \] (4.31)
it can be seen that (4.30) can be rewritten as
\[ q^*_{\lambda_k}(\lambda_k) = G(\lambda_k|a_k, b_k), \] (4.32a)
\[ a_k = \frac{\nu+1}{2}, \] (4.32b)
\[ b_k = \frac{\nu}{2} + \frac{1}{2\sigma^2} E_{q^*_{\lambda_k}} \left[ (y_k - h(x^n_k) - H(x^n_k)a^n_k)^2 \right]. \] (4.32c)

The updates (4.29) and (4.32) are iterated until convergence, as in Algorithm 2.2. The resulting \( q^*_{\lambda_k}(\lambda_k) \) is taken as an approximation to \( p(x^n_k|Y_k, x^n_k) \). The resulting \( q^*_{\lambda_k}(\lambda_k) \), the approximate posterior distribution of \( \lambda_k \), can be discarded.

**Approximation of \( p(y_k|x^n_k, Y_{k-1}) \) using Evidence Lower Bound**

Secondly, an approximation of \( p(y_k|x^n_k, Y_{k-1}) \) is given. As discussed in Section 2.4.4, one such approximation is the evidence lower bound \( \mathcal{L} \). Write the log-likelihood as
\[ \log p(y_k|x^n_k, Y_{k-1}) = \mathcal{L}(q_{x^n_k}(x^n_k)q^*_{\lambda_k}(\lambda_k)||p(x^n_k, \lambda_k, y_k|x^n_k, Y_{k-1})) + D_{\text{KL}}(q_{x^n_k}(x^n_k)q^*_{\lambda_k}(\lambda_k)||p(x^n_k, \lambda_k|x^n_k, Y_{k})) \]
\[ \approx \mathcal{L}(q_{x^n_k}(x^n_k)q^*_{\lambda_k}(\lambda_k)||p(x^n_k, \lambda_k, y_k|x^n_k, Y_{k-1})). \] (4.33)

The approximation error above is given by the minimal Kullback-Leibler divergence \( D_{\text{KL}}(q_{x^n_k}(x^n_k)q^*_{\lambda_k}(\lambda_k)||p(x^n_k, \lambda_k|x^n_k, Y_{k})) \). The proposed method assumes that seen as a function of \( x^n_k \), the approximation error
\[ \min_{q_{x^n_k}(x^n_k), q^*_{\lambda_k}(\lambda_k)} D_{\text{KL}}(q_{x^n_k}(x^n_k)q^*_{\lambda_k}(\lambda_k)||p(x^n_k, \lambda_k|x^n_k, Y_{k})) \approx \text{const.} \] (4.34)
is approximately constant\(^1\). The criterion (4.34) is difficult to evaluate numerically in order to determine whether it holds. This is both because the input variables are high-dimensional, and because the Kullback-Leibler divergence cannot be evaluated analytically. Empirical results, in the form of testing the filtering algorithm on data, is a method of verifying whether the approximation is acceptable for a given application. See Section 4.3 for a simulation example.

The evidence lower bound can be expanded as
\[ \mathcal{L}(q_{x^n_k}(x^n_k)q^*_{\lambda_k}(\lambda_k)||p(x^n_k, \lambda_k, y_k|x^n_k, Y_{k-1})) = E_{q^*_{\lambda_k}} E_{q_{x^n_k}} \left[ \log p(x^n_k, \lambda_k, y_k|x^n_k, Y_{k-1}) \right] - E_{q^*_{\lambda_k}} \left[ \log q^*_{\lambda_k}(\lambda_k) \right] - E_{q_{x^n_k}} \left[ \log q^*_{x^n_k}(x^n_k) \right]. \] (4.35)

\(^1\)The reason that an approximately constant error is sufficient, instead of the stricter requirement that the error is approximately 0, is that a constant additive error in the log-likelihood corresponds to a constant multiplicative error in the likelihood. This multiplicative error will be normalized away later in the measurement update.
The first term in the right hand side of (4.35) is given by

\[
\mathbb{E}_{q^*_{\lambda_k}} \mathbb{E}_{q^*_{x|a}} \left[ \log p(x^a_k, \lambda_k, y_k|x^n_k, Y_{k-1}) \right] = 
\frac{-E_{q^*_{\lambda_k}}[\lambda_k/2]}{2\sigma^2} E_{q^*_{x|a}}[(y_k - h(x^n_k) - H(x^n_k)x^n_k)^2] - \frac{1}{2} \log \sigma^2 + \frac{1}{2} E_{q^*_{\lambda_k}}[\log \lambda_k] 
- \frac{1}{2} \text{tr} \left( P^a_{k|k-1} E_{q^*_{x|a}}[(x^a_k - \hat{x}^a_{k|k-1})(\cdot)^T] \right) - \frac{1}{2} \log \det P^a_{k|k-1} 
+ (\nu - 1) E_{q_{\lambda_k}}[\log \lambda_k] - \frac{\nu}{2} E_{q_{\lambda_k}}[\lambda_k] + \frac{\nu}{2} \log \frac{\nu}{2} - \log \Gamma(\frac{\nu}{2}).
\] (4.36)

Furthermore, the last two terms in the right hand side of (4.33) are given by

\[
- E_{q^*_{\lambda_k}}[\log q^*_{\lambda_k}(\lambda_k)] = -(a_k - 1) E_{q^*_{\lambda_k}}[\log \lambda_k] + b_k \left[ E_{q^*_{\lambda_k}}[\lambda_k] \right] 
- b_k \log a_k + \log \Gamma(a_k),
\] (4.37a)

\[
- E_{q^*_{x|a}}[q^*_{x|a}(x^n_k)] = \frac{1}{2} \left( 1 + \log \det P^a_{k|k} \right).
\] (4.37b)

Certain expectations required above are given by

\[
E_{q^*_{x|a}}[(y_k - h(x^n_k) - H(x^n_k)x^n_k)^2] 
= \left( y_k - h(x^n_k) - H(x^n_k)\hat{x}^a_{k|k} \right)^2 + H(x^n_k)P^a_{k|k}H(x^n_k)^T,
\] (4.38a)

\[
E_{q^*_{x|a}}[(x^a_k - \hat{x}^a_{k|k})(\cdot)^T] 
= P^a_{k|k} + (\hat{x}^a_{k|k} - \hat{x}^a_{k|k-1})(\cdot)^T,
\] (4.38b)

\[
E_{q^*_{\lambda_k}}[\lambda_k] = a_k / b_k,
\] (4.38c)

\[
E_{q^*_{\lambda_k}}[\log \lambda_k] = \psi(a_k) - \log b_k.
\] (4.38d)

where \( \psi(a_k) = \frac{d}{da_k} \log \Gamma(a_k) \), and \( \Gamma(a_k) \) is the Gamma function (Abramowitz and Stegun, 1964).

To conclude, the approximation for the weight update is given in (4.35)–(4.37). In comparison to the Gaussian Rao-Blackwellized weight update, similar quantities such as 
\( (y_k - h(x^n_k) - H(x^n_k)\hat{x}^a_{k|k})^2 \) occur in both expressions, however, expressions such as the trace in (4.36) are related to the variational approximation, and do not have a straightforward counterpart in the exact Gaussian expression.

The full measurement update is described in Algorithm 4.2. This approximate Rao-Blackwellized point mass filter can then be applied on a CLSSM with a Student’s \( t \) measurement noise model, as described in Algorithm 4.1.
Algorithm 4.2 Approximate Rao-Blackwellized measurement update of one point mass value $x^{n,j}_k$ within a point mass filter with Student’s $t$ noise (c.f. Algorithm 2.2)

Require: Measurement $y_k$ and predictive distribution $p(x^a_k|x^{n,j}_k, Y_{k-1})$.

Set $i = 0$. Initialize $q^{(i)}_{\lambda_k}(\lambda_k) = G(a_k, b_k)$ with $a_k = b_k = \nu+1$.

while not converged do

Set $i = i + 1$.
Update $q^{(i)}_{x^a_k}(x^a_k)$ using (4.29) given $q^{(i-1)}_{\lambda_k}(\lambda_k)$.
Update $q^{(i)}_{\lambda_k}(\lambda_k)$ using (4.32) given $q^{(i)}_{x^a_k}(x^a_k)$.

end while

Compute $\mathcal{L}(q^{(i)}_{x^a_k}(x^a_k)q^{(i)}_{\lambda_k}(\lambda_k)||p(x^a_k, \lambda_k, y_k|x^{n,j}_k, Y_{k-1}))$ using (4.33)–(4.37).

Set $p(x^a_k|x^{n,j}_k, Y_k) = q^{(i)}_{x^a_k}(x^a_k)$.
Set $w_{j,k} \propto w_{j,k-1} \exp(\mathcal{L}(q^{(i)}_{x^a_k}(x^a_k)q^{(i)}_{\lambda_k}(\lambda_k)||p(x^a_k, \lambda_k, y_k|x^{n,j}_k, Y_{k-1})))$.

4.3 Simulated Example of RBPMF Frequency Tracking

In order to exemplify the performance of the different RBPMF variations, the filters are tested on a simple frequency tracking example using the instant phasor model described in Section 4.1.3.

4.3.1 Simulation Setup

The data is simulated according to the instant phasor model (4.3) for $k = 1, \ldots, N$, with $N = 500$. Each state trajectory is simulated from fixed initial values of the angular velocity and phasors with with differing random values of the process noises. The measurement noise $e_k$ is simulated as a Gaussian mixture

$$p(e_k) = (1 - d)\mathcal{N}(e_k|0, \sigma^2_{\text{gen}}) + d\mathcal{N}(e_k|0, K^2\sigma^2_{\text{gen}}),$$

(4.39)

where $d$ is the probability of an outlier, $K$ is the factor the noise is increased in the outlier, and $\sigma^2_{\text{gen}}$ is the variance of the outlier-free noise. The noise is exactly Gaussian when $K = 1$, but higher values of $K$ yield outliers which are larger, yielding a noise distribution which is less similar to a Gaussian one.

The tested algorithms are:

- Standard RBPMF (Smidl and Gasperin, 2013, and Section 2.4.5)
- RBPMF with modified merge step (Section 4.2.1)
- RBPMF with Fourier merge step (Section 4.2.2)
- $t$-RBPMF with modified merge step (Section 4.2.1 and 4.2.3)
- $t$-RBPMF with Fourier merge step (Section 4.2.2 and 4.2.3)
Table 4.1: Parameters used for simulation and filtering in Section 4.3.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M$</td>
<td>1</td>
</tr>
<tr>
<td>$\sigma^2_{\text{gen}}$</td>
<td>$10^{-2}$</td>
</tr>
<tr>
<td>$d$</td>
<td>0.05</td>
</tr>
<tr>
<td>$Q^{\alpha,\beta}$</td>
<td>$10^{-2}$</td>
</tr>
<tr>
<td>$Q^\omega$</td>
<td>$10^{-2}$</td>
</tr>
<tr>
<td>$T$</td>
<td>$10^{-2}$</td>
</tr>
<tr>
<td>$N$</td>
<td>500</td>
</tr>
<tr>
<td>$N_S$</td>
<td>300</td>
</tr>
</tbody>
</table>

The algorithms are ordered, increasingly, by how many of the proposed RBPMF variations in this chapter they contain.

The filters are applied with number of grid points $N_S = 500$ with the model (4.3). For both the Gaussian and Student’s $t$ filters, the used noise model parameters are estimated by simulating $10^4$ noise variates from the true noise model, and computing the maximum-likelihood estimate of the noise model parameters. All other model parameters are the true ones. The simulation parameters are chosen to be similar to the corresponding parameters in Chapter 6. Table 4.1 for values of the simulation and modeling parameters.

4.3.2 Simulation Results and Discussion

Figure 4.2 shows the rmse of the RBPMF with modified merge step and the $t$-RBPMF with modified merge step, as a function of the outlier variance scale $K$, with the other parameters given in Table 4.1. The different variants of the Gaussian filters have similar rmse values, and the same applies for the two variants of the $t$-RBPMF, so only one curve for each of them is shown. In the case that $K = 1$, so that the true noise is Gaussian, and it can be seen that the RBPMF has somewhat lower estimation variance. This may be because it is better to use the true noise model than an unnecessarily complex one, in particular when the more complex one requires additional filter approximations. However, when $K$ is increased, it can be seen that the Student’s $t$ filter has better performance. The additional effort required to implement the Student’s $t$ filter is seen to reduce frequency tracking error.

In Figure 4.2, the rmse is reduced when increasing the outlier scale $K$ from 7 to 8, and so on. An intuitive explanation for this is that when $K$ is increased from this stage onwards, the noise model can “more easily” distinguish whether it is a non-outlier or an outlier. The Student’s $t$ noise model cannot do this as easily for a smaller value of $K$. This can be related to the discussion of robust estimation in Section 2.5. The influence function for a mean estimator is not the same as the influence function of a frequency tracker, but analogies can be drawn. The influence function for the Gaussian distribution is monotonically increasing, but the Student’s $t$ influence function redescends, see Figure 2.2. This effect corresponds
The RBPMF for Frequency Tracking

Figure 4.2: Frequency tracking RMSE of the RBPMF and t-RBPMF as a function of the outlier scale \( K \). The ground truth frequency is between 0 and 50 rad/s. The three variants of the RBPMF have almost identical RMSE, and the same applies for the two variants of the t-RBPMF. Note the scale of the ordinate. When \( K \) is larger than the values illustrated, a simple outlier rejection scheme would also work well.

Table 4.2: Computational time of execution of the filter on one data sequence \((N = 500)\), for variations of the RBPMF in Section 4.3.

<table>
<thead>
<tr>
<th>Filter</th>
<th>Comp. time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard RBPMF (Smidl and Gasperin, 2013)</td>
<td>34.4 s</td>
</tr>
<tr>
<td>RBPMF with modified merge step (Section 4.2.1)</td>
<td>6.1 s</td>
</tr>
<tr>
<td>RBPMF with Fourier merge step (Section 4.2.2)</td>
<td>1.9 s</td>
</tr>
<tr>
<td>t-RBPMF with modified merge step (Section 4.2.1 and 4.2.3)</td>
<td>7.2 s</td>
</tr>
<tr>
<td>t-RBPMF with Fourier merge step (Section 4.2.2 and 4.2.3)</td>
<td>3.3 s</td>
</tr>
</tbody>
</table>

to Figure 4.2.

The computational time is given in Table 4.2. It can be seen that the modified merge step greatly reduces computational cost, and the Fourier time update reduces it further. The addition of the variational Student’s \( t \) measurement update into the filter increases computational cost.

In conclusion, we have seen that the proposed modifications to the RBPMF can reduce computational cost. The t-RBPMF can improve frequency tracking performance when the noise model deviates from a Gaussian.
In this chapter, Application 1, frequency tracking of engine sound, is investigated. The objective of this chapter is to validate the methods described in the previous chapters on a practical application. The experimental data used in this chapter consists of microphone measurements of the engine sound from a vehicle traversing an area (see Figure 5.1). These measurements are recorded in different Doppler-shifted versions. The original motivation of the experiments is described by Lindgren et al. (2015). In short, the engine sound frequency estimates can be used to compute the speed and position of the vehicle. By improving the frequency tracking, this estimation can be improved.

The scenario is first studied in a controlled environment using simulations, see
Section 5.1. After this, the method is evaluated on an experimental data sequence, see Section 5.2, in order to confirm that the simulations are realistic.

5.1 Simulation Evaluation in an Acoustic Doppler Scenario

A simulation study was carried out in order to evaluate the algorithm in a controlled environment similar to the experimental scenario described later. In the simulation study, 50 realizations were generated. A vehicle emitting a sound with constant source frequency of 50 Hz has been simulated as if it was driving along the trajectory shown in Figure 5.4. Measurements from each microphone location in the figure were generated, yielding nine measurement time series in total.

5.1.1 Simulation Model

This section describes the simulation model used to evaluate the frequency tracking methods. In order to perform the simulations, the Doppler effect is first studied. For an illustration of the Doppler effect, see Figure 5.2.

Let \( r^{(i)}(t) \) be the distance between the transmitter and the stationary receiver \( i \) at time \( t \). If the acoustic signal arrives at time \( t \), then it was transmitted at time \( t - \Delta^{(i)}(t) \), where \( \Delta^{(i)}(t) \) is the propagation delay. Given the acoustic transmission speed \( c \), the transmission distance is \( c\Delta^{(i)}(t) \), so that the propagation delay can be computed through solving

\[
 r^{(i)}(t - \Delta^{(i)}(t)) = c\Delta^{(i)}(t). \tag{5.1}
\]

Differentiate this equation with respect to time, and isolate \( \dot{\Delta}^{(i)}(t) \), yielding

\[
 \dot{\Delta}^{(i)}(t) = \frac{v^{(i)}_r(t - \Delta^{(i)}(t))}{c_s + v^{(i)}_r(t - \Delta^{(i)}(t))}, \tag{5.2}
\]

where \( v^{(i)}_r(t) = \frac{d}{dt}r^{(i)}(t) \) is the radial velocity of the transmitter to microphone \( i \).

Assume that the transmitter emits an acoustic signal with frequency \( f^{(0)}(t) \). The \( i \)th receiver will measure the Doppler-shifted signal with frequency

\[
 f^{(i)}(t) = (1 - \dot{\Delta}^{(i)}(t))f^{(0)}(t - \Delta^{(i)}(t))
 = \frac{c_s}{c_s + v^{(i)}_r(t - \Delta^{(i)}(t))}f^{(0)}(t - \Delta^{(i)}(t)) \tag{5.3}
\]

With known vehicle and microphone positions, these quantities can be computed by first solving (5.1) for the transmission delay \( \Delta^{(i)}(t) \) and then computing the Doppler shift \( \dot{\Delta}^{(i)}(t) \) using (5.2).

In the simulations, the sound emitted from the vehicle at time \( t_k \) is proportional to

\[
 \sum_{m=1}^{M} B_k^{m,0} \cos(m2\pi f^{(0)}t_k + \phi_k^{(m,0)}), \tag{5.4}
\]
5.1 Simulation Evaluation in an Acoustic Doppler Scenario

The Doppler effect. The center dot and arrow correspond to a stationary and moving transmitter, respectively. The transmitted harmonic signal can be, for instance, engine sound from a vehicle. The black lines correspond to sound waves. Whenever the vehicle is moving, the measured frequency of the sound depends on the location of the observer.

Figure 5.2: The Doppler effect. The center dot and arrow correspond to a stationary and moving transmitter, respectively. The transmitted harmonic signal can be, for instance, engine sound from a vehicle. The black lines correspond to sound waves. Whenever the vehicle is moving, the measured frequency of the sound depends on the location of the observer.

where $B_{k}^{m,0}$ are amplitudes, and $\phi_{k}^{(m,0)}$ are phase offsets.

The measured sound at microphone $i$ is simulated as follows. Let $\Delta_{k}^{(i)} = \Delta^{(i)}(t_{k})$ be the transmission delay at time $t_{k}$, and let $t_{k} = \Delta_{k}^{(i)} - \Delta_{k}^{(i)}$ be the time of transmission of the signal received at time $t_{k}$. Furthermore, let $t_{k} = i_{k+1} - i_{k}$ be the difference in transmission time. Then, the measurements $y_{k}^{(i)}$ of microphone $k$ are simulated according to

$$y_{k}^{(i)} = \sum_{m=1}^{M} \frac{B_{k}^{m,i}}{r^{(i)}(t_{k})} \cos(m2\pi f^{(0)} t_{k} + \phi_{k}^{(m,i)}) + e_{k}^{(i)},$$

where $e_{k}^{(i)}$ is Gaussian and i.i.d. The factor $\frac{1}{r^{(i)}}$ models the decay of the sound level due to increasing distance from the source. With a perfect source, $B_{k}^{m,i}$ and $\phi_{k}^{(m,i)}$ may be taken as constant initial amplitudes and phase offsets. However, in order to take imperfections in the generating process into account, the amplitudes and phase offsets are allowed to change according to

$$B_{k+1}^{m,i} = (1 - L)B_{k}^{m,i} + L \left( B_{0}^{m,i} + \sqrt{i_{k}v_{k}^{\phi^{(m,i)}}} \right),$$

$$\phi_{k+1}^{m,i} = \phi_{k}^{m,i} + \sqrt{i_{k}v_{k}^{\phi^{(m,i)}}},$$

where $v_{k}^{\phi^{(m,i)}}$ and $v_{k}^{A^{(m,i)}}$ are Gaussian and i.i.d., and $L$ is a low-pass constant ensuring that the average amplitude stays close to $A_{0}^{(m,i)}$. The parameters used are shown in Table 5.1. The sample frequency of each simulated microphone is 1200 Hz.
Table 5.1: Parameter values used for simulated data generation.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f^{(0)}$</td>
<td>50</td>
<td>Hz</td>
</tr>
<tr>
<td>$M$</td>
<td>6</td>
<td>–</td>
</tr>
<tr>
<td>$\text{Cov}(v^A_{k,m,i})$</td>
<td>$10^{-4}$</td>
<td>Pa$^2$/s</td>
</tr>
<tr>
<td>$\text{Cov}(v^\phi_{k,m,i})$</td>
<td>$10^{-4}$</td>
<td>–</td>
</tr>
<tr>
<td>$\text{Cov}(e^{(i)}_k)$</td>
<td>$10^{-3}$</td>
<td>Pa$^2$</td>
</tr>
<tr>
<td>$L$</td>
<td>$10^{-3}$</td>
<td>–</td>
</tr>
<tr>
<td>$B^0_{0,m,i}$</td>
<td>0.2</td>
<td>Pa</td>
</tr>
<tr>
<td>$\phi^0_{0,m,i}$</td>
<td>0</td>
<td>rad</td>
</tr>
<tr>
<td>$f_s$</td>
<td>1200</td>
<td>Hz</td>
</tr>
</tbody>
</table>

Table 5.2: Parameter values used for the different algorithms. For more detail about discrete periodogram maximization (DPM), see Lindgren et al. (2015).

<table>
<thead>
<tr>
<th>RBPMF Par.</th>
<th>Value</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M$</td>
<td>6</td>
<td>–</td>
</tr>
<tr>
<td>$N_P$</td>
<td>250</td>
<td>–</td>
</tr>
<tr>
<td>$R$</td>
<td>1</td>
<td>Pa$^2$</td>
</tr>
<tr>
<td>$Q^\omega$</td>
<td>$10^4$</td>
<td>rad$^2$/s$^3$</td>
</tr>
<tr>
<td>$Q^{\alpha\beta}$</td>
<td>0.1</td>
<td>Pa$^2$/s</td>
</tr>
<tr>
<td>$x^{n,j}$</td>
<td>$2\pi(20 + \frac{40j}{N})$</td>
<td>rad/s</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DPM Par.</th>
<th>Value</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M$</td>
<td>6</td>
<td>–</td>
</tr>
<tr>
<td>$N_b$</td>
<td>80</td>
<td>–</td>
</tr>
<tr>
<td>$c_p$</td>
<td>$1 -</td>
<td>p</td>
</tr>
<tr>
<td>$\omega_{\text{min}}$</td>
<td>$2\pi \cdot 40$</td>
<td>rad/s</td>
</tr>
<tr>
<td>$\omega_{\text{max}}$</td>
<td>$2\pi \cdot 60$</td>
<td>rad/s</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SHS Par.</th>
<th>Value</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M$</td>
<td>6</td>
<td>–</td>
</tr>
<tr>
<td>$N_b$</td>
<td>80</td>
<td>–</td>
</tr>
<tr>
<td>$c_m$</td>
<td>0.99$^m$</td>
<td>–</td>
</tr>
</tbody>
</table>

5.1.2 Studied Methods and Evaluation

The studied algorithms are:

- RBPMF with modified merge step, Algorithm 4.1.
- Discrete periodogram maximization (DPM), (Lindgren et al., 2015). Variation of periodogram frequency estimation.
- Subharmonic summation (SHS), Algorithm 3.1.

The parameters used in the algorithms are shown in Table 5.2.

For the simulations, the ground truth frequency at microphone $i$ is

$$f_k^{*(i)} = f^{(0)}(i) \cdot (1 - \hat{A}^{(i)}_k). \quad (5.6)$$

The batch-oriented algorithms are run every $N_b$ samples on a batch of length $N_b$, yielding an estimate $\hat{f}_k^{(i)}$. The RBPMF is run continuously, and every $N_b$ samples, an estimate $\hat{f}_k^{(i)}$ is extracted by computing the mean of the $N_b$ estimates. Given
Table 5.3: Simulation RMSE results (Hz) over 50 noise realizations. It can be seen that the RBPMF achieves the lowest RMSE.

<table>
<thead>
<tr>
<th></th>
<th>M1</th>
<th>M2</th>
<th>M3</th>
<th>M4</th>
</tr>
</thead>
<tbody>
<tr>
<td>RBPMF</td>
<td>0.057</td>
<td>0.058</td>
<td>0.068</td>
<td>0.058</td>
</tr>
<tr>
<td>DPM</td>
<td>0.169</td>
<td>0.171</td>
<td>0.172</td>
<td>0.177</td>
</tr>
<tr>
<td>SHS</td>
<td>0.139</td>
<td>0.176</td>
<td>0.150</td>
<td>0.161</td>
</tr>
</tbody>
</table>

the estimated frequency $\hat{f}_k(i)$ and the ground truth frequency $f_k^*(i)$, the RMSE is then computed as

$$\text{RMSE}(\{\hat{f}_k(i)\}_{k=1}^{N/N_b}) = \sqrt{\frac{N}{N_b} \sum_{k=1}^{N/N_b} (\hat{f}_k(i) - f_k^*(i))^2}.$$  \hspace{1cm} (5.7)

5.1.3 Simulation Results and Discussion

The RMSE values from the simulation are given in Table 5.3. The RBPMF has the lowest RMSE. This may be explained by the fact that the RBPMF considers phase and amplitude coherence over time, and inherently allows for variability in the source frequency. The batch length is comparatively short, $N_b = 80$. In this case, phase and amplitude coherence appears to be useful in order to improve the accuracy of the estimate. Note that the RBPMF implementation in MATLAB requires roughly 100 times more computational effort than the corresponding SHS or DPM implementations, although it also provides an estimate 80 times more often.

In Figure 5.3, the average RMSE over all sensors and all simulation runs is plotted against time. It can be seen that the RMSE is smallest in the middle of the trajectory. This is because the engine is closest to the microphones at this point in time, so that the SNR is at its highest. This effect is most pronounced for the SHS method, and one explanation for this could be that its frequency interpolation is better than the DPM method. The RBPMF can again be seen to have the lowest average RMSE for each time instant as compared to its competitors.

5.2 Experimental Evaluation in an Acoustic Doppler Scenario

For experimental evaluation, a data set described in Deleskog et al. (2014) is used. The experimental data was collected in the scenario illustrated in Figure 5.4. An all-terrain vehicle (ATV) traversed the circular segment at constant speed. Microphones with known positions in the area measured the sound emitted by the ATV. The recorded sound contains clearly audible harmonics from the ATV’s engine. Unfortunately, no ground truth frequency is available. Instead a reference frequency is estimated using the fact that the different microphones measure differ-
ent, Doppler-shifted, delayed versions of the same engine sound with continuously varying frequency.

5.2.1 Experimental Setup

The experimental data used for evaluation of the methods was collected using a setup of eight single microphones (M1–M8), and one acoustic array (A) with four microphones (A1–A4), as described in Deleskog et al. (2014). Acoustic data from an ATV was collected while the vehicle passed through the area at approximately constant vehicle speed and engine speed. The acoustic data from the microphones was recorded at a sample rate of 1200 Hz. It contains multiple harmonics generated by the vehicle’s internal combustion engine, frequency shifted by the Doppler effect due to the movement of the ATV. The area used for testing is shown in Figure 5.4.

The positions of the microphones are known, and in order to validate the method, a differential global positioning system (DGPS) receiver measures the position of the vehicle with high accuracy.

5.2.2 Reference Estimation

A fundamental problem with evaluation of the performance of different frequency tracking methods using the experimental data detailed above is that the ground truth, i.e., the source frequency $f^{(0)}(t)$ emitted by the ATV engine, is unavailable.
One approach for circumventing this problem is to estimate a reference frequency which can be used for comparison. This can be done using the fact that all measurements originate from the same acoustic sound signal with continuously varying source frequency. However, the measurements are frequency shifted by the Doppler effect, and delayed due to the transmission delay from the ATV to the microphone in question.

The values of the source frequency \( f^{(0)} \) at particular times \( f_k^{(0)} = f^{(0)}(\bar{t}_k) \) are of interest, where \( \Lambda^{(i)}_k = \Lambda^{(i)}(t_k) \). The notation is the same as in Section 5.1. Assuming a “constant velocity” model for the source frequency, its discrete time state-space formulation with state \( z_k^{(0)} = [f_k^{(0)} \, \dot{f}_k^{(0)}]^T \) is given by

\[
\begin{align*}
z_{k+1}^{(0)} &= \begin{bmatrix} 1 & \bar{t}_k \\ 0 & 1 \end{bmatrix} z_k^{(0)} + v_k^{(0)}, \\
\hat{f}_k^{(i)} &= (1 - \dot{\Lambda}_k^{(i)}) \begin{bmatrix} 1 & 0 \end{bmatrix} z_k^{(0)} + e_k^{(i)}, \quad 1 \leq i \leq S,
\end{align*}
\] (5.8a, 5.8b)

where the outputs \( \hat{f}_k^{(i)} \) of the frequency tracking algorithm from microphone \( i \) are considered as measurements of the Doppler shifted and delayed source frequency, and \( S \) is the total number of microphones. The frequency tracker errors \( e_k^{(i)} \) are approximated as independent Gaussian white noise signals for \( i = 1, \ldots, S \). The
process noise $v_k^{(0)}$ is modeled as i.i.d. and Gaussian,

$$v_k^{(0)} \sim \mathcal{N}(0, Q_k),$$

$$Q_k = \sigma_{cv}^2 \begin{bmatrix} \tilde{t}_k^3/3 & \tilde{t}_k^2/2 \\ \tilde{t}_k^2/2 & \tilde{t}_k \end{bmatrix},$$

where $\sigma_{cv}^2$ is a tuning parameter controlling the variability of the source frequency. This covariance structure is derived from the corresponding continuous time state-space model. The values $\Delta_k^{(i)}$ and $\hat{\Delta}_k^{(i)}$ are computed using logged microphone positions and DGPS measurements for the vehicle position, as described in Section 5.1. For evaluation purposes, it is assumed that they are exactly known. A Rauch-Tung-Striebel smoother as described by Särkkä (2013) and in Section 2.4.5 is applied to this model, and the posterior mean estimate

$$\hat{f}_k^{(0)} = \mathbb{E}[f_k^{(0)} | F^{(1)}, \ldots, F^{(S)}]$$

of the source frequency given the set of all frequency tracker estimate time series $F^{(1)}, \ldots, F^{(S)}$, where $F^{(i)} = \{\hat{f}_k^{(i)}\}_{k=1}^N$, is obtained. Then, for evaluation, the expected frequency is compared to the sensor value for microphone $i$ using the root mean square error (RMSE) value

$$\text{RMSE}(\{\hat{f}_k^{(i)}\}_{k=1}^N) = \sqrt{\frac{1}{N} \sum_{k=1}^N \left( \hat{f}_k^{(i)} - (1 - \hat{\Delta}_k^{(i)}) \hat{f}_k^{(0)} \right)^2},$$

where the $\hat{\ }$ notation is used since the comparison is not made to a measured ground truth value.

This comparison is performed for each frequency tracker individually. That is, each frequency tracker outputs a set of time series $F^{(1)}, \ldots, F^{(S)}$, and a separate posterior mean $\hat{f}_k^{(0)}$ and corresponding RMSE is computed for each frequency tracker from those values. This is done in order to make the comparisons equitable for different methods.

### 5.2.3 Experimental Results

The three tracking methods (RBPMF, DPM, and SHS) described above were compared and tested on experimental data using the parameters shown in Table 5.2. All results use a batch length of $N_b = 80$ unless otherwise noted. The number of harmonics used is $M = 6$, and for reference frequency estimation, $\sigma_{cv}^2 = 10^3$.

#### Evaluation using Microphone Array

The frequency tracking performance is first evaluated using the microphone array A in Figure 5.4. Since its four microphones are co-located, they experience the same Doppler shift, so the estimates are directly comparable. RMSE values in this case are shown in Table 5.4. For all sensors in this data sequence, the RBPMF has the lowest RMSE. In this case, with the given batch length $N_b = 80$, tracking performance is improved with the RBPMF.
Table 5.4: Experimental rmse results (Hz) for one trajectory and a microphone array of four microphones located in the same sensor array A. In this case the estimation of source frequency is not required, since the Doppler effect is the same for all of them.

<table>
<thead>
<tr>
<th></th>
<th>A1</th>
<th>A2</th>
<th>A3</th>
<th>A4</th>
</tr>
</thead>
<tbody>
<tr>
<td>RBPMF</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
</tr>
<tr>
<td>DPM</td>
<td>0.12</td>
<td>0.11</td>
<td>0.11</td>
<td>0.12</td>
</tr>
<tr>
<td>SHS</td>
<td>0.11</td>
<td>0.11</td>
<td>0.12</td>
<td>0.10</td>
</tr>
</tbody>
</table>

Evaluation using Estimated Reference Frequency

An example of the estimated reference frequency is given in Figure 5.5, where the spectrogram of the experimental microphone data is shown. Superimposed is the reference frequency computed as described in Section 5.2.2. The estimated source frequency is shown in Figure 5.6, and it can be seen that the variation of the source frequency is comparable in magnitude to the Doppler frequency shift. Without a variable source frequency, the red curves in Figure 5.5 would be monotonically decreasing, and this does not agree with the visual shape of the harmonics in the periodogram. This motivates the relevance of estimating the source frequency.

In Figure 5.7, the experimental error of the three algorithms compared to the estimated reference frequency is shown for one data sequence and one sensor. It can be observed that the estimation error for the RBPMF has the lowest variation, although all methods systematically deviate from the true frequency around \( t = 0 \) s. For the same data sequence, Table 5.5 shows estimated rmse values for the different sensors as determined in Section 5.2.2; the values are presented for four sensors for brevity, but the results are representative for all nine sensors. For all sensors in this data sequence, the RBPMF has the lowest rmse. This result is similar to the microphone array comparison in Table 5.4. The magnitude of the errors is slightly higher in Table 5.5. This may be explained by an imperfect reference frequency estimation inflating rmse in Table 5.5 or certain correlated errors (for instance, correlated wind noises) reducing rmse values in Table 5.4.

In the case that frequency estimates are not required as often, or that the variability of the source frequency is small, the batch length may be set to, e.g., \( N_b = 600 \) (batch duration of \( \frac{1}{2} \) s) instead of \( N_b = 80 \) (batch duration of \( \frac{1}{15} \) s) which was used above. In this case, the methods are more comparable, as seen in Table 5.6, with the SHS method having the best rmse. One explanation for this is that a large amount of the performance degradation for the periodogram-based methods in Tables 5.5–5.4 is due to them ignoring phase and amplitude coherence between batches. This is not a problem when the batch size \( N_b \) is large enough, since then, there is more information identifying the phase and amplitude of each harmonic in the data. This trade-off must be made with respect to the variability of the frequency, since a higher variability can be expected to require a smaller batch size \( N_b \). It is straightforward to realize that a time resolution required to capture Figure 5.6, cannot be achieved with \( N_b = 600 \).
Figure 5.5: An example of the reference frequency (red) computed as detailed in Section 5.2.2, superimposed on the spectrogram of the acoustic measurements of one microphone. Since the frequency tracking methods consider six harmonics, six multiples of the curves are illustrated for easy comparison with the underlying spectrogram, where darker shades indicate more energy at that time and frequency band. At time 0, the vehicle is approximately in the middle of the trajectory. This is also, approximately, the time when the vehicle is closest to the microphone.

Figure 5.6: Estimated source frequency from the same data sequence as Figure 5.5. It can be seen that the source frequency is clearly varying.
Figure 5.7: Estimation errors for frequency for the three compared methods over an experimental data set. Note that the RBPMF does not provide frequency estimates for a few fractions of a second until its frequency estimate has sufficiently low variance. The estimation error is smaller than the variation of the source frequency in Figure 5.6.
Table 5.5: Experimental RMSE results (Hz) for one trajectory and four different microphones M1–M4. The source frequency was estimated using all 12 microphones, however, only four RMSE values are presented for brevity.

<table>
<thead>
<tr>
<th></th>
<th>M1</th>
<th>M2</th>
<th>M3</th>
<th>M4</th>
</tr>
</thead>
<tbody>
<tr>
<td>RBPMF</td>
<td>0.11</td>
<td>0.11</td>
<td>0.11</td>
<td>0.11</td>
</tr>
<tr>
<td>DPM</td>
<td>0.15</td>
<td>0.14</td>
<td>0.13</td>
<td>0.15</td>
</tr>
<tr>
<td>SHS</td>
<td>0.14</td>
<td>0.14</td>
<td>0.14</td>
<td>0.14</td>
</tr>
</tbody>
</table>

Table 5.6: Experimental RMSE results (Hz) for one trajectory and four different microphones M1–M4. The source frequency was estimated using all 12 microphones, however, only four RMSE values are presented for brevity. For these values, $N_b = 600$ instead of $N_b = 80$.

<table>
<thead>
<tr>
<th></th>
<th>M1</th>
<th>M2</th>
<th>M3</th>
<th>M4</th>
</tr>
</thead>
<tbody>
<tr>
<td>RBPMF</td>
<td>0.053</td>
<td>0.053</td>
<td>0.049</td>
<td>0.051</td>
</tr>
<tr>
<td>DPM</td>
<td>0.063</td>
<td>0.049</td>
<td>0.058</td>
<td>0.061</td>
</tr>
<tr>
<td>SHS</td>
<td>0.045</td>
<td>0.045</td>
<td>0.045</td>
<td>0.042</td>
</tr>
</tbody>
</table>

5.3 Summary

The problem of acoustic frequency tracking of a harmonic signal has been considered. Simulation and experimental results indicate that the RBPMF has lower RMSE in comparison to periodogram-based methods in RMSE when accurate tracking of fast variations in the frequency is required.
This chapter studies Application 2, frequency tracking of wheel vibrations. The objective is to determine the wheel speed of a vehicle using vibration measurements from accelerometers mounted in the vehicle chassis. It has been observed that the accelerometer signal contains a dominant frequency component proportional to the rotational speed of the wheels of the vehicle. A concept illustration can be seen in Figure 6.1. By tracking this frequency component, an angular velocity estimate can be attained. This enables access to wheel angular velocity when direct speed measurements from wheel speed sensors (WSS) or GPS sensors are unavailable, and without relying on suitable light conditions like camera systems typically do. In contrast to dead-reckoning by integrating a longitudinal accelerometer, the error of the proposed wheel speed estimate does not accumulate with time, and it is insensitive to accelerometer offsets caused by sensor bias and road inclination.

Figure 6.1: Illustration of a vehicle equipped with an accelerometer (Acc). Vibrations (grey) arise from the wheel axles, and are transmitted to the vehicle chassis, where the accelerometer is mounted and can measure them. The idea is to estimate wheel speed by tracking the frequency of the vibrations, so that access to the WSS is not required.
6.1 Experiments and Methods

This section describes the experimental setup, the studied methods, and the evaluation metrics used for the experiments.

6.1.1 Data Collection and Pre-Processing

The experimental data are in the form of lateral accelerometer data collected from an in-vehicle sensor cluster in an Audi A6 vehicle. A standard vehicular logging device (Vector Informatik GmbH, 2014) samples the vehicle’s FlexRay bus collecting accelerometer data $y_k$ from the built-in sensor cluster, as well as ground truth data in form of angular velocity $\omega_k^*$ from a wheel speed sensor (WSS) with a sampling rate of $\omega^S = 2\pi \cdot 100 \text{rad/s}$, thus with $T = 0.01 \text{s}$. This sample rate is chosen to enable estimation of vehicle wheel speed up to typical legal speed limits (120 km/h), without unnecessarily increasing sample rate requirements and computational effort. The WSS measurement error is in the order of $0.1 \text{rad/s}$, which will be seen to be small in comparison to the frequency estimation errors.

The lateral accelerometer signal was chosen because it contains clear harmonics related to the wheel speed. The longitudinal accelerometer signal also does, but it often contains harmonics proportional to the engine speed as well. These harmonics depend on the gearing as well as on the engine speed, and this influence is undesired for the application. Engine vibrations do not influence the lateral accelerometer signal as much. The vertical accelerometer signal has too low an $\text{SNR}$.

The raw lateral accelerometer signal is high-pass filtered with a Butterworth filter of order 2 and cut-off frequency $20 \text{rad/s}$ in order to remove the static level of the accelerometer signal as well as undesired resonance modes unrelated to the tracked harmonics. This yields the measurement signal $y_k$ illustrated as a spectrogram in Figure 6.2.

There are 90 test drives in total. The test drives are carried out on varying road surfaces. These varying conditions lead to different signal and noise characteristics, thus varying $\text{signal-to-noise ratios (SNRs)}$. The duration of each test drive is between 100 and 1500 s.

6.1.2 Data Segmentation

The test drives were split into subsequences in order to be able to compare them equitably. Every test drive was split up into test sequences of duration $L_{\text{min}}$ to $L_{\text{max}}$, where angular velocity, or fundamental frequency, was within the range

$$\omega_{\text{min}} < \omega_k^* < \omega_{\text{max}}, \quad (6.1)$$

and average initial absolute angular acceleration within the first $L_{\text{acc}}$ seconds was sufficiently small,

$$\frac{1}{L_{\text{acc}}} \sum_{k=1}^{L_{\text{acc}}/T} |\omega_k^* - \omega_{k-1}^*| < \dot{\omega}_{\text{max}}. \quad (6.2)$$
The latter criterion is imposed in order to avoid a large initial acceleration during the start of the test sequence. This is to study the algorithms in normal situations rather than in the initial transient part. The criterion is straightforward to implement, and it is a substitute for more complex initialization procedures. The parameters used are shown in Table 6.1, and statistics about the segmentation are shown in Figure 6.3.

### 6.1.3 Accelerometer Input

The basic filter does not integrate the accelerometer measurements in order to estimate the wheel angular velocity from vibrations. However, an accelerometer integration component can be added as an input straightforwardly in order to predict changes in the fundamental frequency, reducing the error of the filter. This is done by changing the dynamics of the state $\omega_k$ in the model (4.3) into

$$\omega_{k+1} = \omega_k + T a_{k}^{\text{long}} / r_{\text{nom}} + v_{k}^{\omega},$$

where $a_{k}^{\text{long}}$ is the longitudinal acceleration and $r_{\text{nom}}$ is the nominal wheel radius, which must be known or externally estimated for the purpose at hand.
Table 6.1: Parameters used for data segmentation.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega_{\text{min}}$</td>
<td>30</td>
<td>rad/s</td>
</tr>
<tr>
<td>$\omega_{\text{max}}$</td>
<td>125</td>
<td>rad/s</td>
</tr>
<tr>
<td>$L_{\text{acc}}$</td>
<td>5</td>
<td>s</td>
</tr>
<tr>
<td>$\dot{\omega}_{\text{max}}$</td>
<td>0.03</td>
<td>rad/s²</td>
</tr>
<tr>
<td>$L_{\text{min}}$</td>
<td>30</td>
<td>s</td>
</tr>
<tr>
<td>$L_{\text{max}}$</td>
<td>60</td>
<td>s</td>
</tr>
</tbody>
</table>

6.1.4 Studied Methods and Parameter Selection

The studied methods include batched frequency tracking methods, Algorithm 3.4, based on the frequency estimation methods below:

- Periodogram, Algorithm 3.1 (periodogram).
- Generalized Katkovnik, Algorithm 3.2, with a Student’s $t$ noise assumption (Katkovnik).

They also include the recursive frequency tracking methods below:

- RBPMF, Algorithm 4.1, using the state-space model in Section 4.1.3, with a Gaussian noise assumption (RBPMF).
- Variational approximate RBPMF, Algorithm 4.1, using the state-space model in Section 4.1.3, with a Student’s $t$ noise assumption ($t$–RBPMF).
- RBPF, Algorithm 2.7, using the state-space model in Section 4.1.3, with a Gaussian noise assumption (RBPF).
- Variational approximate RBPF, Algorithm 2.7, using the state-space model in Section 4.1.3, with a Student’s $t$ noise assumption ($t$–RBPF).

The recursive frequency tracking methods may or may not include a longitudinal accelerometer component as an input in the filters, as described in Section 6.1.3. This is shown by writing (input) after the method name. The RBPMF-based methods are used with a grid $x_{n,j}^k$ linearly spaced from 5 to 105 rad/s, with 250 grid points. The same grid was used for the generalized Katkovnik method. The RBPF-based methods use 250 particles.

In order to tune the parameters for every method, a 100 second calibration sequence of the data was selected, which is the data sequence shown in Figure 6.2. This calibration sequence was chosen for its representativeness of the data at large, and the length is due to considerations of computational cost, since the computations need to be rerun thousands of times in order to consider the many parameter combinations. Every method was applied to the calibration data set with a multitude of parameters on a grid, and the parameter set which yielded the lowest RMSE on the data set was chosen. The resulting parameters, for all the studied methods, are shown in Table 6.2.
### Table 6.2: Parameters used for the methods presented in the evaluation.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Parameter</th>
<th>Value</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t$–RBPMF,</td>
<td>$\nu$</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>$t$–RBPF</td>
<td>$\sigma^2$</td>
<td>$10^{-11/4}$</td>
<td>m$^2$/s$^4$</td>
</tr>
<tr>
<td></td>
<td>$Q^\omega$</td>
<td>3</td>
<td>rad/s$^2$</td>
</tr>
<tr>
<td></td>
<td>$Q^{\alpha,\beta}$</td>
<td>$10^{-3/2}$</td>
<td>m$^2$/s$^5$</td>
</tr>
<tr>
<td>RBPMF,</td>
<td>$R$</td>
<td>$10^{-3/2}$</td>
<td>m$^2$/s$^4$</td>
</tr>
<tr>
<td>RBPF</td>
<td>$Q^\omega$</td>
<td>3</td>
<td>rad$^2$/s$^3$</td>
</tr>
<tr>
<td></td>
<td>$Q^{\alpha,\beta}$</td>
<td>$10^{-5/2}$</td>
<td>m$^2$/s$^5$</td>
</tr>
<tr>
<td>$t$–RBPMF,</td>
<td>$\nu$</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>(input)</td>
<td>$\sigma^2$</td>
<td>$10^{-2}$</td>
<td>m$^2$/s$^4$</td>
</tr>
<tr>
<td></td>
<td>$Q^\omega$</td>
<td>1</td>
<td>rad$^2$/s$^2$</td>
</tr>
<tr>
<td></td>
<td>$Q^{\alpha,\beta}$</td>
<td>$10^{-3/2}$</td>
<td>m$^2$/s$^5$</td>
</tr>
<tr>
<td>RBPMF,</td>
<td>$R$</td>
<td>$10^{-7/4}$</td>
<td>m$^2$/s$^4$</td>
</tr>
<tr>
<td>(input)</td>
<td>$Q^\omega$</td>
<td>1</td>
<td>rad$^2$/s$^3$</td>
</tr>
<tr>
<td></td>
<td>$Q^{\alpha,\beta}$</td>
<td>$10^{-11/4}$</td>
<td>m$^2$/s$^5$</td>
</tr>
<tr>
<td>Periodogram</td>
<td>$T_{\text{batch}}$</td>
<td>2</td>
<td>s</td>
</tr>
<tr>
<td></td>
<td>$\Delta\omega$</td>
<td>5</td>
<td>rad/s</td>
</tr>
<tr>
<td>Katkovnik</td>
<td>$T_{\text{batch}}$</td>
<td>2</td>
<td>s</td>
</tr>
<tr>
<td></td>
<td>$\nu$</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>$\sigma^2$</td>
<td>$10^{-7/4}$</td>
<td>m$^2$/s$^4$</td>
</tr>
<tr>
<td></td>
<td>$\Delta\omega$</td>
<td>5</td>
<td>rad/s</td>
</tr>
</tbody>
</table>
Figure 6.3: Statistics of the test sequences illustrated in histograms. Top: the distribution of the speed of the tests; the speeds are approximately equally represented between 30 and 110 rad/s. Middle: Distribution of the length of the test sequences after data segmentation. Bottom: the distribution of test sequences in the 90 test drives.

6.1.5 Evaluation

Ground truth data \( \{\omega^*_k\}_{k=1}^N \) from the WSS is available. For each test sequence, RMSE values are computed by

\[
\text{RMSE}(\{\hat{\omega}_k\}_{k=1}^N) = \sqrt{\frac{1}{N} \sum_{k=1}^{N} (\hat{\omega}_k - \omega^*_k)^2}.
\]  

(6.4)

One property of the experimental data is the ratio of the energy in the periodic signal to the energy in the noise; the SNR. This quantity is unknown, but may be estimated using the spectrogram. The signal energy is expected to be within the main lobe width of the Hamming window used to compute the spectrogram; i.e., within \( m\omega^* \pm \omega^S/N_b \) for \( m = 1, \ldots, M \). The noise energy is spread out over the entire spectrum, but the energy within the frequency band of the signal is neglected. The estimated SNR is thus

\[
\hat{\text{SNR}} = \frac{E_{\text{signal}}}{E_{\text{noise}}},
\]  

(6.5)
with the estimated signal and noise energies given by

\[
E_{\text{signal}} = 2 \sum_{\ell=1}^{N/N_b} \sum_{m=1}^{M} \int_{m\omega^s - \omega^s / N_b}^{m\omega^s + \omega^s / N_b} |\mathcal{Y}_{\ell}(\omega)|^2 d\omega,
\]

\[
E_{\text{noise}} = 2 \sum_{\ell=1}^{N/N_b} \int_0^{\omega^s} |\mathcal{Y}_{\ell}(\omega)|^2 d\omega - E_{\text{signal}},
\]

where \(\mathcal{Y}_{\ell}(\omega)\) is the FFT of the \(\ell\):th batch of the data, with batch lengths \(N_b = 100\). The integrals are approximated by summations of the FFT values.
6.2 Results and Discussion

In this section, the results from the experimental evaluation are described and discussed. The figures below contain illustrations of the performance of subsets of the algorithms listed in Section 6.1.4.

6.2.1 Experimental Results

Figure 6.4 shows the RMSEs of each test sequence, averaged over time, as an empirical cumulative distribution function (CDF). At lower quantiles, the periodogram has the lowest RMSE, while the RBPMF, t-RBPMF, RBPF and t-RBPF have somewhat higher RMSEs. At higher quantiles (for example, at quantile 0.9), the t-RBPF has the lowest error. From practical perspectives, it has been found that an RMSE below 1 rad/s is sufficiently low. This frequency RMSE limit corresponds to a velocity RMSE of approximately 0.3 m/s ≈ 1 km/h, since the wheel radius \( r_{\text{nom}} \approx 0.3 \) m. Thus, above quantiles 0.5–0.6, no algorithms have a sufficiently low RMSE.

![Figure 6.4: Comparison of RMSE of frequency estimation and tracking algorithms on experimental data. Every point in this CDF plot corresponds to one test sequence. An error of 1 rad/s corresponds to a speed error of 0.3 m/s, or approx. 1 km/h, if the wheel radius is 0.3 m, a typical value for a car. An error below 1 rad/s is deemed sufficiently good for typical use cases. The methods have sufficiently good performance only for somewhat more than half the test data.](image-url)
The estimated SNR was computed for each test sequence. In Figure 6.5, the RMSE as a function of the $\hat{SNR}$ is illustrated. All algorithms have higher RMSE when the $\hat{SNR}$ is reduced, but the RBPMF is the most adversely affected when $\hat{SNR}$ is low. At high values of SNR, periodogram and Katkovnik methods perform best.

Figure 6.5: RMSE for six studied algorithms as a function of the $\hat{SNR}$, averaged over intervals of width 0.3. The performance of all considered frequency tracking methods is dependent on the $\hat{SNR}$. 
Figure 6.6 shows the RMSE when only test sequences with $\hat{SNR}$ greater than 0.5 have been included. For these cases, the RMSE of all algorithms is lower. The batch-based algorithms perform well at lower quantiles, while the $t$-RBPMF performs well at higher ones.

![Figure 6.6: Comparison of RMSE of frequency estimation and tracking algorithms on experimental data. Every point in this CDF plot corresponds to one test sequence. Only test sequences with $\hat{SNR}$ more than 0.5 have been included. Note the difference in comparison with Figure 6.4. In most of these test sequences, the methods have an RMSE below 1 rad/s.](image)
In Figure 6.7, the RMSE of algorithms using, and not using, an added accelerometer input signal (see Section 6.1.3) is illustrated. It can be seen that the included input improves the performance of the recursive algorithms.

![Figure 6.7: Comparison of RMSE of Rao-Blackwellized frequency tracking algorithms on experimental data. The first four curves in the caption are the same as in Figure 6.4, without an accelerometer input, while the last four have an accelerometer input. The inclusion of the input is seen to reduce RMSE.](image)

Figures 6.8 and 6.9 show frequency tracking results for two example test sequences. In the first figure, the methods accurately follow the true frequency. In the second example, including an accelerometer input signal reduces frequency tracking error.

6.2.2 Discussion

In Figure 6.4, the batch-based methods perform well, in particular in the lower quantiles. One reason for this could be that the test sequences which compose the lower quantiles contain a lot of driving at approximately constant speed. The batch length is relatively long, and in these cases, batch-based methods perform well, as exemplified in Figure 6.8. For the higher quantiles, performance significantly degrades. In the higher quantiles, the $t$-RBPMF and $t$-RBPF perform best. One possible explanation for the difference between the RBPF and RBPMF at higher quantiles can be seen by looking at the time update. The analytic states of the RBPF do not interact between different particles, while the analytic states of the RBPMF do interact in the merge step. Assume that the analytic state of one
particle or grid point in the RBPF or RBPMF acquires a poor state distribution. In the RBPF, the particle will be down-weighed and eventually resampled away. On the other hand, in the RBPMF the analytic state will be propagated to neighboring grid points due to the merge step; this may be undesirable.

To analyze the instances of poor performance further, the estimated SNR was computed, as shown in Figure 6.5. The reason for such a disparity in frequency tracking performance can be attributed to variations in experimental data collection, such as different road surfaces. Selecting only test sequences with sufficiently high $\tilde{\text{SNR}}$, the performance of the algorithms is notably better, as seen in Figure 6.6 contrasted to Figure 6.4. In lower quantiles, the periodogram algorithm has the lowest RMSE. At higher quantiles, the $t$-RBPMF and Katkovnik methods perform best. This may be because the noise can be well approximated as Gaussian at lower quantiles, but that such an approximation is poorer at higher quantiles.

In order to improve performance, it is studied how the addition of an accelerometer integration component into the recursive methods influences the performance. In Figure 6.7 and in the example of Figure 6.9, it can be seen that this improves performance in comparison to methods which do not include an input signal. With the accelerometer input, the RBPMF perform best at lower quantiles, while the $t$-RBPF performs best at higher quantiles. Again, one possible explanation is that the noise is approximately Gaussian in lower quantiles while it may be more heavy-tailed at higher quantiles. There are multiple reasons why this improvement occurs. One is that there is a reduction in bias, since systematic changes in the source frequency are partially taken into account by the accelerometer signal. Another is that the additional signal makes it possible to reduce the process variance for the state $\omega_k$, reducing the variance of the estimate. However, note that there are disadvantages to including the accelerometer input component. Accelerometers often tend to be biased due to sensor imperfections. Bias also arises when the vehicle is weighed down by a load in the trunk or when it is driving on slopes. Disregarding the frequency information fully and using only accelerometer dead-reckoning will yield an estimate that drifts off and becomes inaccurate. Depending on sensor bias and road inclination, the error typically accumulates by roughly 0.1 to 0.5 rad/s every second.

### 6.3 Summary

This chapter studied Application 2, frequency tracking of wheel vibrations. It was seen that the problem is challenging, but feasible for subsets of the experimental data. Thanks to the flexible state-space model framework for the recursive methods, it was straightforward to include an accelerometer input to predict changes in the fundamental frequency. This improved performance for low SNR test sequences.
Figure 6.8: Top: Illustration of true and estimated fundamental frequency over time for a subset of the studied algorithms. Bottom: Spectrogram of the input data for the algorithms. In this example test sequence, the true frequency is approximately constant. In the end, the signal appears to be slightly weaker. The tracks of the batch-based methods are the least oscillatory ones.
Figure 6.9: Top: Illustration of true and estimated fundamental frequency over time for a subset of the studied algorithms. Bottom: Spectrogram of the input data for the algorithms. In this example test sequence, the true frequency changes relatively quickly, and the SNR is low. Without accelerometer input information, the methods lose track of the frequency.
Summary, Conclusions, and Further Work

This thesis has studied frequency estimation and tracking methods. For frequency estimation, methods which consider measurements that contain outliers were studied in detail. For recursive frequency tracking methods, methods for Rao-Blackwellized filtering including variational approximate Rao-Blackwellization were proposed. The methods were evaluated using simulated data and experimental data from two applications relating to estimation of vehicle speed.

7.1 Summary and Conclusions

This section summarizes the thesis and discusses selected conclusions to be drawn from the theory and results.

Maximum Likelihood Frequency Estimation

Chapter 3 presented theory and methods for maximum-likelihood frequency estimation with heavy-tailed noise models to handle measurements with outliers. In a simulation evaluation, it was seen that methods for frequency estimation using heavy-tailed noise models improved performance if the noise contains outliers. Usage of the generalized real Katkovnik method is recommended in general for the lowest estimation error, with the EM frequency estimation method with a Student’s $t$ model providing a computationally cheaper alternative. As expected, the periodogram works well when the noise is Gaussian.

The RBPMF for Frequency Tracking

Chapter 4 discussed models and Rao-Blackwellized filtering methods for frequency tracking. One state-space model was proposed due to its suitability for Rao-Blackwellized filtering methods. Variants of the Rao-Blackwellized point mass
filter (RBPMF) were proposed and evaluated in a simulation study. The proposed RBPMF variants could save computational effort while maintaining accuracy for the example. The $t$-RBPMF was proposed in order to enable a Student's $t$ measurement noise model using variational approximate Rao-Blackwellization. The method was seen to perform better when the measurement noise contains outliers.

### Frequency Tracking of Engine Sound

The acoustic frequency tracking application was discussed and evaluated in Chapter 5. Batch-oriented methods work well when the batch length was high. Meanwhile, when the batch length was smaller, they performed comparatively poorer. In these contexts, both in simulation and in experiments, recursive frequency tracking methods in the form of the RBPMF with computational speed-ups proposed in Chapter 4 are recommended for their lower error.

### Frequency Tracking of Wheel Vibrations

In Chapter 6, the application on frequency tracking of wheel vibrations measured using an accelerometer was discussed and evaluated. It was seen that algorithm performance on the experimental data was dependent on estimated SNR. On the high SNR test sequences where most methods performed well, standard batch-oriented methods were superior, while on more difficult test sequences, recursive methods, and methods which handle heavy-tailed noise can have lower error than batch-oriented methods. A general recommendation for the frequency tracking problem at hand is to use batch-oriented methods. In low SNR conditions, including a longitudinal accelerometer input was shown to be useful in order to improve performance for the recursive methods.

### 7.2 Further work

There is a multitude of ways to extend this work. It may be of interest to study the theoretical limits of frequency estimation algorithms with heavy-tailed noise models. Another theoretical area of interest is to generalize the concept of approximate variational Rao-Blackwellization introduced in Section 4.2.3. This idea does not appear to have been proposed previously in the literature. In what other model structures and inference algorithms may variational approximate Rao-Blackwellization be of use?

The applications also have relevant areas of future study. For the acoustic frequency tracking application in Chapter 5, the improvements in localization performance gained from frequency tracking accuracy could be studied. Both improved time and frequency fidelity may improve localization performance. Lindgren et al. (2015) assumes the source frequency to be constant. Allowing the source frequency to be variable may yield better performance. In the wheel vibration tracking application, in Chapter 6, a logical next step is to combine this method into a positioning system as discussed by Vilhelmsson (2017). In such a system, the
accelerometer input component may be combined with a pose estimation system in order to model biases in the accelerometer.
Bibliography


