

A Data-Driven Method for Monitoring of Repetitive Systems: Applications to Robust Wear Monitoring of a Robot Joint [★]

André Carvalho Bittencourt ^a, Kari Saarinen ^b, Shiva Sander-Tavallaey ^b,
Svante Gunnarsson ^a Mikael Norrlöf ^{a,c}

^a*Department of Electrical Engineering, Linköpings University, Linköping, Sweden*

^b*ABB Corporate Research, Västerås, Sweden*

^c*ABB Robotics, Västerås, Sweden*

Abstract

This paper presents a method for monitoring of systems that operate in a repetitive manner. Considering that data batches collected from a repetitive operation will be similar unless in the presence of an abnormality, a condition change is inferred by comparing the monitored data against a nominal batch. The method proposed considers the comparison of data in the distribution domain, which reveals information of the data amplitude. This is achieved with the use of kernel density estimates and the Kullback-Leibler distance. To decrease sensitivity to unknown disturbances while increasing sensitivity to faults, the use of a weighting vector is suggested which is chosen based on a labeled dataset. The framework is simple to implement and can be used without process interruption, in a batch manner. The method was developed with interests in industrial robotics where a repetitive behavior is commonly found. The problem of wear monitoring in a robot joint is studied based on data collected from a test-cycle. Real data from accelerated wear tests and simulations are considered. Promising results are achieved where the method output shows a clear response to the wear increases.

Key words: Condition monitoring, Data driven methods, Industrial robots, Wear, Condition based maintenance, Automation

1 Introduction

Driven by the severe competition in a global market, stricter legislation and increase of consumer concerns towards environment and health/safety, industrial systems face nowadays higher requirements on safety, reliability, availability and maintainability (SRAM). In the industry, equipment failure is a major factor of accidents and downtime, [13, 18]. While a correct specification and design of the equipments are crucial for increased SRAM, no amount of design effort can prevent deterioration over time and equipments will eventually fail. Nevertheless, its impacts can be considerably reduced if good maintenance practices are performed.

In the manufacturing industry, including industrial robots, preventive scheduled maintenance is a common approach used to improve equipment's SRAM. This setup delivers high availability, reducing operational costs (e.g. small downtimes) with the drawback of high maintenance costs since unnecessary maintenance actions might take place. Condition based maintenance (CBM), "maintenance when required", can deliver a good compromise between maintenance and operational costs, reducing the overall cost of maintenance. The extra challenge of CBM is to define methods to determine the condition of the equipment. This can be done by comparing the observed and expected (known) behaviors of the system through an algorithm. The output of such algorithm is a quantity sensitive to a fault, i.e. a *fault indicator*, which can be monitored to determine the current state of the system (e.g. healthy/broken).

A common approach to generate fault indicators is based on the use of residuals, i.e. fault indicators that are achieved based on deviations between measurements and *model* equation based calculations. A model of the sys-

[★] This work was supported by ABB and the Vinnova Industry Excellence Center LINK-SIC at Linköping University.

Email addresses: andrecb@isy.liu.se (André Carvalho Bittencourt), kari.saarinen@se.abb.com (Kari Saarinen), shiva.sander-tavallaey@se.abb.com (Shiva Sander-Tavallaey), svante@isy.liu.se (Svante Gunnarsson), mikael.norrlof@se.abb.com (Mikael Norrlöf).

tem provides important information about the behavior of the system and facilitates the generation of fault indicators. Different approaches for residual generation are based on, e.g., observers, parity-space and parameter identification. When a model of the system is not available or it is too costly to be developed, alternatives are still possible. These alternatives will typically require extra (redundant) sensory information or expert knowledge about the measured data, e.g., their nominal frequency content or the use of *labeled data*. Essentially, however, any method will attempt to generate quantities that can be used to infer the actual condition of the system given the available knowledge and observations, i.e. data.

In the industrial robotics literature, model-based methods are commonly used. Due to the complex dynamics of an industrial robot, the use of nonlinear observers is a typical approach (see [8] for a review). Since observers are sensitive to model uncertainties and disturbances, some methods attempt to diminish these effects. In [6] and [9], nonlinear observers are used together with adaptive schemes; in [7], the authors mix the use of nonlinear observers with support vector machines and in [21], neural networks are used. Parameter estimation is a natural approach because it can use the physical interpretation of the system, see e.g. [12, 2, 14]. Deriving reliable robot models from physics and identification experiments is however an involving task, see e.g. [15] for identification of flexible manipulators. Alternatives that do not rely on a model have been presented in [17], where relevant features of sound measurements are monitored and in [10], where vibration data are used. No reference was found of condition monitoring methods for industrial robots that make a direct use of the repetitive behavior of the system.

For CBM, it is interesting to study faults* that can be detected before a critical degradation takes place. Faults that follow from a gradual degradation of the robot, e.g. due to aging and wear, are good candidates for CBM because of their typical slowly varying behavior. An example of such fault is studied in [14], where an observer-based method is suggested for health monitoring of the actuators' lubricant. In [2], a method is proposed for robust wear identification in a robot joint under temperature disturbances; the method is based on a custom designed identification experiment, i.e. a test-cycle, and a known friction model which can describe the effects of speed, load, temperature and wear.

In this paper, a data-driven method is proposed for the generation of fault indicators for systems that operate in a repetitive manner. It is considered that in case the condition of the system is nominal, data batches collected

from repetitive executions of the system will be similar to each other and will differ if the condition changes. The comparison of a given data batch against a nominal one can thus be used to infer whether an abnormality is present. The proposed fault indicator relates to changes in the distribution of these batches of data. This is made possible with the use of kernel density estimators and the Kullback-Leibler distance between distributions. The focus of the paper is to present the framework and the ideas for the generation of fault indicators and the topics of alarm generation and fault isolation are not addressed.

The method was developed with the interest focused on condition monitoring of industrial robots, where a repetitive operation is found in many of its applications. A repetitive behavior is also commonly found in automation or can be forced with the execution of a test-cycle, with the drawback of reduced availability. The problem of robust wear monitoring in a robot joint is considered to illustrate the framework based on real and simulated data. The problem description and basic framework are presented in Sections 2 and 3 respectively. The robotics application and results are presented in Section 4. Conclusions and future work are given in Section 5.

2 Monitoring of Repetitive Systems – Problem Description

Consider a general system from which it is possible to extract a sequence of data batches,

$$\mathbf{Y}^M = [\mathbf{y}^0, \dots, \mathbf{y}^k, \dots, \mathbf{y}^{M-1}], \quad (1)$$

where $\mathbf{y}^k = [y_1^k, \dots, y_i^k, \dots, y_N^k]^T$ denotes the N dimensional data vector (e.g. measurements or known inputs) with batch index k .

The sequence \mathbf{y}^k could have been generated as the result of deterministic and stochastic inputs, \mathbf{Z}^M and \mathbf{V}^M , where \mathbf{V}^M is unknown, and \mathbf{Z}^M may have known and unknown components. For example, the data generation mechanism could be modeled as

$$\mathbf{y}^k = h(\mathbf{z}^k, \mathbf{v}^k), \quad (2)$$

where $h(\cdot)$ is a general function. Let the set of deterministic inputs \mathbf{Z}^M be categorized in three distinct groups, \mathbf{R}^M , \mathbf{D}^M and \mathbf{F}^M . The sequences \mathbf{f}^k are unknown and of interest (a fault), while \mathbf{r}^k and \mathbf{d}^k are known and unknown respectively (e.g. references and disturbances). With the purpose of monitoring \mathbf{y}^k to detect changes in \mathbf{f}^k , the following assumptions are made:

- A-1 (**Faults are observable**) Changes on \mathbf{f}^k affect the available data \mathbf{y}^k .
- A-2 (**Regularity of \mathbf{y}^k if no fault**) The monitored data \mathbf{y}^k change only slightly along k , unless a nonzero fault \mathbf{f}^k occurs.

* A fault is defined as a deviation of at least one characteristic property of the system from the acceptable/usual/nominal condition.

A-3 (**Nominal data are available**) At $k = 0$, $\mathbf{f}^0 = 0$ and the sequence \mathbf{y}^0 is available.

While Assumption A-1 is necessary, Assumption A-2 ensures that two given sequences $\mathbf{y}^m, \mathbf{y}^n$ are comparable as long as there is no change of condition. Nominal data are assumed known to allow for a comparison against nominal.

The rationale is then to generate fault indicators from the comparison of the nominal data \mathbf{y}^0 (available from Assumption A-3) against the remaining sequences \mathbf{y}^k . In order to generate fault indicators using the monitored data \mathbf{y} , two basic questions arise:

Q-1 How to characterize \mathbf{y}^k ?

Q-2 How to compare two sequences $\mathbf{y}^m, \mathbf{y}^n$?

The first question targets the issue of finding a data processing mechanism of \mathbf{y}^k , written in a general form as $g(\mathbf{y}^k)$, that enhances the ability to further discriminate the presence of \mathbf{f}^k . The outputs of the form $g(\mathbf{y}^k)$ can then be compared against nominal with the use of a comparison or distance function, represented e.g. as $d(g(\mathbf{y}^0)||g(\mathbf{y}^k))$. Under Assumptions A-1 to A-3, the output of such comparison can be used as a fault indicator.

Ensuring a regular behavior of \mathbf{y}^k according to Assumption A-2 is however difficult in practice. Uncontrollable inputs often affect the data, leading to an undesired behavior of the fault indicator and confusion of the inference mechanism. When the data are affected by deterministic inputs as in Equation (2), Assumption A-2 can be achieved if \mathbf{r}^k and \mathbf{d}^k are regular over k , leading to the following conditions:

C-1 (**Regularity of \mathbf{r}^k**) The known deterministic inputs \mathbf{r}^k change only slightly along k .

C-2 (**Regularity of \mathbf{d}^k**) The unknown deterministic inputs \mathbf{d}^k change only slightly along k .

Notice that Conditions C-1 and C-2 are deliberately stated in a qualitative manner, favoring the presentation of the ideas in the paper. A more formal treatment is outside the scope of this paper and will depend on the data generation mechanism, e.g. the function $h(\cdot)$ in (2), and on the inference mechanism chosen, i.e. how data are characterized and compared.

Condition C-1 ensures a repetitive operation of the system over k and is natural when \mathbf{r} are references. For example, in case r_i^k can be chosen freely, a test-cycle can be used to guarantee a repetitive behavior by choosing $\mathbf{r}^{k-1} = \mathbf{r}^k$ for all k . A repetitive behavior over k is also required for \mathbf{d} according to Condition C-2, i.e. uncontrollable inputs must have a regular behavior over the batches. Notice though that the sequences \mathbf{r} and \mathbf{d} are

allowed to vary over i , thus allowing for a non-stationary behavior of the system.

Condition C-2 is in many practical cases too restrictive. To broaden the scope of the framework it is desirable that this can be relaxed, leading to the question:

Q-3 How to handle irregular disturbances \mathbf{d}^k ?

Questions Q-1 to Q-3, outlined in this section, are addressed in the next section, where the framework and ideas are described.

3 A Framework for Monitoring of Repetitive Systems

3.1 Characterizing the Data – Kernel Density Estimate

There are several ways to address Question Q-1. A sequence \mathbf{y}^k could be characterized by a single number, such as its mean, peak, range, etc. Summarizing the whole sequence into single quantities might however hide many of the data features. A second alternative would be to simply store the whole sequence and try to monitor the difference $\mathbf{y}^0 - \mathbf{y}^k$ but this requires that the sequences are, or can be, synchronized. Depending on the nature of the data, a synchronization might introduce errors, which can complicate a decision. In cases where the data are ordered and, possibly, collected under stationary conditions, the use of transforms, e.g. Fourier and/or Wavelet, might reveal relevant information about the fault, see e.g. [11].

The alternative pursued in this work is to consider the distribution of \mathbf{y}^k , which contains information about the amplitude behavior of the signal. Even though information contained in the ordering is lost, this is a valid approach since the effects of a fault appear many times as changes in amplitude. Because the mechanisms that generated the data are considered unknown, the use of a nonparametric estimate of the distribution of \mathbf{y}^k is a suitable alternative. A nonparametric estimate of the distribution $p(\cdot)$ of \mathbf{y}^k can be achieved with the use of kernel density estimators,

$$\hat{p}^k(y) = N^{-1} \sum_{i=1}^N k_h(y - y_i^k), \quad (3)$$

where $k_h(\cdot)$ is a kernel function satisfying $k_h(\cdot) \geq 0$ and that integrates to 1 over the real line. The bandwidth $h > 0$ is a smoothing parameter and y includes the domain of \mathbf{Y}^M (see e.g. [5] for more details on kernel density estimators and criteria/methods for choosing h). From the definition, it follows that $\int \hat{p}(y) dy = 1$, that is, the distribution estimate is normalized to 1. The quantity $\hat{p}^k(y)$ is the kernel density estimate (KDE) of \mathbf{y}^k .

Equation (3) can be rewritten as the convolution

$$\hat{p}^k(y) = N^{-1} \int_{-\infty}^{\infty} \sum_{i=1}^N \delta(x - y_i^k) k_h(y - x) dx, \quad (4)$$

where $\delta(\cdot)$ is the Dirac delta. Using the convolution theorem, the kernel density estimator can be seen as a filter in the distribution domain, controlling the smoothness of the estimated distribution. It is typical to choose kernels which are symmetric and with a low pass behavior, where the bandwidth parameter h controls its cutoff frequency. In this work, a Gaussian kernel is considered, with h optimized for Gaussian distributions, see e.g. [5].

3.2 Comparing Sequences – Kullback-Leibler Distance

In statistics and information theory, the Kullback-Leibler divergence (KLD) is commonly used as a measure of difference between two probability distributions. For two continuous distributions on y , $p(y)$ and $q(y)$, it is defined as

$$D_{\text{KL}}(p||q) = - \int_{-\infty}^{\infty} p(y) \log \frac{q(y)}{p(y)} dy \quad (5)$$

The KLD satisfies $D_{\text{KL}}(p||q) \geq 0$ (Gibbs inequality), with equality if and only if $p(y) = q(y)$. The KLD is in general not symmetric, $D_{\text{KL}}(p||q) \neq D_{\text{KL}}(q||p)$. The quantity

$$\text{KL}(p||q) \triangleq D_{\text{KL}}(p||q) + D_{\text{KL}}(q||p), \quad (6)$$

known as the Kullback-Leibler distance, is however symmetric. See [19] for an up to date review of divergences.

An answer to Question Q-2 can therefore be given with the use of the KL distance defined in (6). From Assumption A-3, fault-free data are always available, so that \mathbf{y}^0 is known and $\hat{p}^0(y)$ can be computed. The quantities $\text{KL}(\hat{p}^0||\hat{p}^k)$ can therefore be used as a fault indicator, remaining close to 0 in case $\hat{p}^0(y)$ is close to $\hat{p}^k(y)$ and otherwise deviating to positive values.

3.3 Handling Irregular Disturbances – Data Weighting

One way to address Question Q-3 is to weight the raw data \mathbf{y}^k according to prior knowledge of the fault and disturbances in order to give more relevance to parts of the data relating to a fault. The approaches considered here will assume availability of a labeled dataset, \mathbf{Y}^M as given in (1), where the fault status (present or not) is known to each component \mathbf{y}^k and is the same to each of its elements y_i^k . The disturbance vector \mathbf{d}^k does not need to satisfy Condition C-2, in fact, the dataset should contain variations in \mathbf{d}^k that are expected to be found during the system's operation.

The fault-free data in the set are said to belong to the class \mathcal{C}_0 , with M_0 observations, while the faulty data belong to class \mathcal{C}_1 , with $M_1 = M - M_0$ observations. Each batch \mathbf{y}^k is weighted according to

$$\bar{\mathbf{y}}^k = \mathbf{w} \circ \mathbf{y}^k, \quad (7)$$

where \circ is the Hadamard product (element-wise multiplication), yielding the weighted dataset

$$\bar{\mathbf{Y}}^M \triangleq [\bar{\mathbf{y}}^0, \dots, \bar{\mathbf{y}}^{M_0}, \bar{\mathbf{y}}^{M_0+1}, \dots, \bar{\mathbf{y}}^{M_1+M_0}]. \quad (8)$$

The objective is to choose \mathbf{w} to maximize the sensitivity to faults while decreasing sensitivity to disturbances.

Considering the basic framework presented in Sections 3.1 and 3.2, a natural criterion would be to choose \mathbf{w} according to its effects to $\text{KL}(\hat{p}^m(\mathbf{w})||\hat{p}^n(\mathbf{w}))$, where $\hat{p}(\cdot)$ is the KDE of $\bar{\mathbf{y}}$ and therefore dependent on \mathbf{w} . When \mathbf{y}^m is fault-free and \mathbf{y}^n is faulty (and vice-versa), the distance should be maximized otherwise it should be minimized. A general solution to this problem is however difficult since it depends on how $\hat{p}^k(\mathbf{w})$ is computed (e.g. the kernel function chosen) and optimization over (6).

In this work, simpler criteria are considered in a compromise to explicit solutions. As it will be shown, the results are related to linear discriminant analysis (LDA) used in classification problems, see e.g. [1]. In LDA, instead of the Hadamard product used in (7), data are weighted using the inner product, yielding $\mathbf{w}^T \mathbf{y}$. While the data are reduced to a scalar quantity in LDA, the use of the Hadamard product keeps the data dimensionality and therefore the KDE can still be computed, yielding the estimates $\hat{p}^k(\mathbf{w})$. Furthermore, the objective in LDA is to obtain a classifier; here, \mathbf{w} is chosen as to achieve *average separation* between faulty and fault-free data while giving small variability to disturbances.

Notice that once the weights are chosen, the same vector \mathbf{w} is used for new data batches. For consistency, it is thus required that the data sequences are synchronized. This can however be overcome in case the weights are strongly correlated to measured data. In such case, an approximate function can be used to describe the weights relation to the data, e.g. described as a static function $h(\cdot)$ such that $w_i = h(y_i^k)$. The use of such representation of the weights is illustrated in Section 4.2.

3.3.1 Choosing \mathbf{w} – Linear Discriminant Analysis

A simple criterion is to maximize the difference between the classes means in average. The average of the c th class

mean over all M_c observations is

$$\begin{aligned}\bar{\mu}^c &\triangleq N^{-1} \sum_{i=0}^{N-1} \left[M_c^{-1} \sum_{k \in \mathcal{C}_c} w_i y_i^k \right] \\ &= N^{-1} \sum_{i=0}^{N-1} w_i \underbrace{\left[M_c^{-1} \sum_{k \in \mathcal{C}_c} y_i^k \right]}_{\triangleq \mu_i^c} = N^{-1} \mathbf{w}^T \boldsymbol{\mu}^c.\end{aligned}$$

The distance between the means of classes \mathcal{C}_0 and \mathcal{C}_1 is proportional to

$$\bar{\mu}^1 - \bar{\mu}^0 \propto \mathbf{w}^T (\boldsymbol{\mu}^1 - \boldsymbol{\mu}^0)$$

and the objective is to choose \mathbf{w} which maximizes the expression. This problem is equivalently found in LDA. Constraining \mathbf{w} to unit length $\mathbf{w}^T \mathbf{w} = 1$ (otherwise the criterion can be made arbitrarily large), it is possible to find that the optimal choice is according to (see e.g. [1, Exercise 4.4]),

$$\mathbf{w}^* \propto (\boldsymbol{\mu}^1 - \boldsymbol{\mu}^0). \quad (9)$$

A criterion based only on the distance between the classes means does not consider the variability found within each class, e.g. caused by disturbances. An alternative is to consider maximum separation between the classes means while giving small variability within each class. The average value of the weighted variance vector over k for class c is given by

$$\begin{aligned}\bar{s}^c &\triangleq N^{-1} \sum_{i=0}^{N-1} \left[M_c^{-1} \sum_{k \in \mathcal{C}_c} (w_i y_i^k - w_i \mu_i^c)^2 \right] \\ &= N^{-1} \sum_{i=0}^{N-1} w_i^2 \underbrace{\left[M_c^{-1} \sum_{k \in \mathcal{C}_c} (y_i^k - \mu_i^c)^2 \right]}_{\triangleq s_i^c} \\ &= N^{-1} \mathbf{w}^T S^c \mathbf{w},\end{aligned}$$

where S^c is a diagonal matrix with diagonal elements given by s_i^c . Defining the total within class variation as $\sum_c \bar{s}^c$, the following criterion can be used when two classes are considered

$$\frac{(\bar{\mu}^1 - \bar{\mu}^0)^2}{\bar{s}^1 + \bar{s}^0} \propto \frac{\mathbf{w}^T (\boldsymbol{\mu}^1 - \boldsymbol{\mu}^0) (\boldsymbol{\mu}^1 - \boldsymbol{\mu}^0)^T \mathbf{w}}{\mathbf{w}^T (S^1 + S^0) \mathbf{w}},$$

which is a special case of the Fisher criterion. It can be shown, see e.g. [1], that solutions for this problem satisfy

$$\mathbf{w}^* \propto (S^1 + S^0)^{-1} (\boldsymbol{\mu}^1 - \boldsymbol{\mu}^0). \quad (10)$$

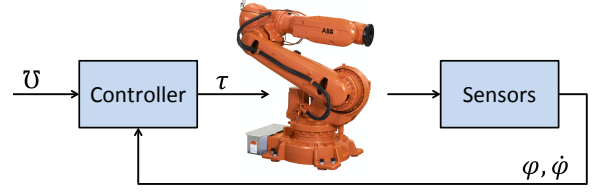


Fig. 1. A simplified scheme of a robot motion control. The trajectory $\bar{\mathbf{U}}$, determined from a robot program, is used as a reference to the controller. The applied torques τ are the control inputs to the system. The controller is based on feed-forward actions (not shown) computed based on $\bar{\mathbf{U}}$ and feedback from measured motor positions and velocities $(\varphi, \dot{\varphi})$.

That is, each weight w_i is proportional to the ratio between the average changes, $\mu_i^1 - \mu_i^0$, and the total variability found in the data, $s_i^1 + s_i^0$.

4 Wear Monitoring in an Industrial Robot Joint

In a typical setup, industrial robots are equipped only with sensors that relate to the robot's states. With little/no information of the surroundings, the behavior of the robot is determined from a robot program, which contains user defined instructions specified in task (or joint) space. Based on the robot program, the control system generates a trajectory, $\bar{\mathbf{U}}$, describing the time dependence of the robot motion required to behave according to the robot program. A trajectory is a known deterministic sequence used as a *reference* to the motion control (see Figure 1), i.e. it relates to \mathbf{r} in the previously introduced framework. In many applications, the same robot program is executed over and over again, in a repetitive manner. Let $\bar{\mathbf{U}}^k$ denote the trajectory to be executed at instance k , a repetitive operation is ensured with $\bar{\mathbf{U}}^{k-1} = \bar{\mathbf{U}}^k$ for all k , thus satisfying Condition C-1.

The controller ensures real-time motion performance and high repeatability. Both feedforward and feedback control actions are used. Typical measured quantities are angular position at the motor side and motor current. Angular position measurements φ are achieved with high resolution resolvers (or encoders) and can be differentiated to achieve motor angular speed $\dot{\varphi}$. A current controller is used to provide a desired torque τ on the motor output. Since the current controller has much faster dynamics compared to the arm, it is common to accept a constant relationship between current and torque, and to consider τ as the control input to the system. The relation between applied torque and motion at a given joint can be described from a multi-body dynamic mechanism by

$$\tau = M(\varphi) \ddot{\varphi} + C(\varphi, \dot{\varphi}) + \tau_g(\varphi) + \tau_s(\varphi) \quad (11a)$$

$$+ \tau_f(\dot{\varphi}, \tau, T, \mathbf{w}), \quad (11b)$$

where τ is the applied torque at the joint. The terms given by $M(\varphi)$, $C(\varphi, \dot{\varphi})$, $\tau_g(\varphi)$, $\tau_s(\varphi)$ and $\tau_f(\cdot)$ relate to

the inertia, speed dependent torques (e.g. Coriolis and centrifugal), gravity-induced torque, nonlinear stiffness and friction at that joint. The friction term in (11b) considers its relation to joint speed $\dot{\varphi}$, the manipulated load τ_l , the temperature inside the joint T , and the wear levels w . These dependencies of friction are motivated from the experimental studies presented in [3, 2] and are illustrated in Figure 2*.

The deterministic unknown input of interest, i.e. a fault sequence \mathbf{f} , is the wear level w . The available data are the quantities $(\varphi, \dot{\varphi})$ and the control input τ . The measurements are corrupted by random noise, i.e. \mathbf{v} , and so is τ due to feedback. Among the available data, it is clear from (11) that τ is affected by wear through friction, satisfying Assumption A-1. The applied torque, τ , is therefore included as the monitored sequence \mathbf{y} . The behavior of τ is mainly dependent on φ and its derivatives as given in (11) which are function of the trajectory \mathcal{U} . Notice that, for the same reference trajectory, the required friction torques to drive the joint will differ in case there are friction changes present. This is due to the presence of feedback in the controller.

The variables τ_l and T are deterministic and unknown and thus relate to disturbances \mathbf{d} . Given that the robot is operating in a repetitive manner ($\mathcal{U}^{k-1} = \mathcal{U}^k$), Assumption A-2 is achieved in case τ_l and T satisfy Condition C-2. The manipulated load is dependent on the arm configuration through time (described by the trajectory \mathcal{U}) and on external forces/torques acting on the robot (present e.g. in contact applications). Joint temperature is the result of complicated losses mechanisms in the joint and heat exchanges with the environment which are difficult to control. The effects of τ_l and T to τ are in fact comparable to those caused by w (recall Figure 2). The problem of robust wear monitoring is therefore challenging. Finally, fault-free data, and thus Assumption A-3, are made possible if, e.g., data are available from the beginning of the robot operation, when no significant wear is yet present.

The next subsection presents experimental results for the wear monitoring problem when the changes in disturbances are kept small. In this simplified setting, Condition C-2 is considered valid and the basic framework described in Sections 3.1 and 3.2 is used. In Section 4.2, temperature disturbances are introduced in simulation studies and the approaches described in Section 3.3 are used to illustrate how robustness can be achieved.

4.1 Experimental Wear Monitoring under Constant Disturbances

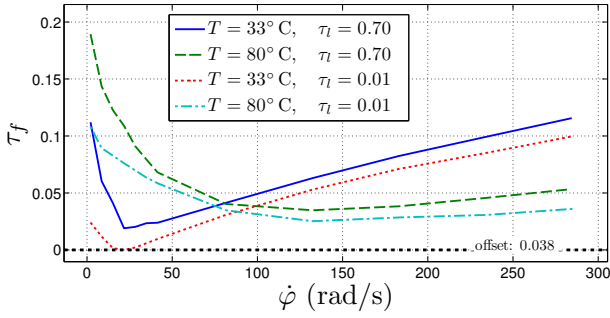
Accelerated wear tests were performed in a robot joint with the objective of studying the wear effects. In an accelerated wear test, the robot is run under high load and stress levels for several months or years until the wear levels become significant and maintenance is required. Throughout the tests, a trajectory \mathcal{U} from a *test-cycle* was executed regularly a total of M times yielding a dataset $[\boldsymbol{\tau}^0, \dots, \boldsymbol{\tau}^{M-1}]$. The experiments were performed in a lab, in a setup to avoid temperature variations and effects of load caused by external forces/torques. It is thus considered that the disturbances satisfy Condition C-2. Considering $\boldsymbol{\tau}^0$ to be fault-free, the quantities $\text{KL}(\hat{p}^0 || \hat{p}^k)$ are computed for $k=1, \dots, M-1$.

Data collected from two accelerated wear tests are considered here to illustrate the usage of $\text{KL}(\hat{p}^0 || \hat{p}^k)$ as a fault indicator. For an illustration of the wear behavior during the experiments, the friction curves in the joint were estimated using a dedicated experiment (see [3] for a description of such experiment) at each k th execution of \mathcal{U} . The results are shown in Figures 3 and 4 where relevant quantities are shown.

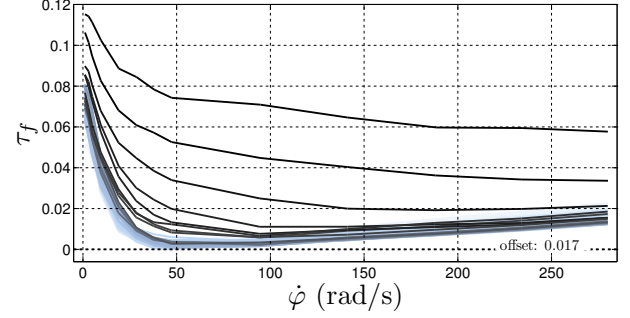
For the first case, displayed in Figure 3, $M=36$ batches of data are considered. From analyses of the friction curves in Figure 3(c), it is possible to note that wear only starts to considerably affect friction after $k \approx 25$. The effects of wear to the torque sequences, shown in Figure 3(a), appear as small variations in amplitude due to increased friction. The variations in the torque sequences are more easily distinguishable in the distribution domain. As seen in Figure 3(b), wear affects the location and size of the KDEs peaks. Notice further that the KDEs are similar during the first part of the tests, i.e. for $k \leq 25$, when the robot condition has not significantly changed. The resulting fault indicator, shown in Figure 3(d), shows a clear response to the changes in friction, remaining close to 0 for $k \leq 25$ and increasing thereafter. To allow for CBM, it is considered that, in this test, a fault should be detected before $k = 30$. Using data for $k \leq 25$, the mean and standard deviation for the (considered) nominal behavior fault indicator are estimated as $[\mu_0, \sigma_0] = [1.19 \cdot 10^{-2}, 5.09 \cdot 10^{-3}]$. The dashed line in Figure 3(d) shows the value of $\mu_0 + 3\sigma_0$, making it clear that such early detection is made possible with the proposed fault indicator.

The second case illustrates the situation where a gearbox is replaced after a wear related failure takes place. A total of $M=111$ data batches are collected during accelerated wear tests using the same test-cycle. At the beginning, the nominal data are assigned as the one collected from the start of the experiments. A gearbox failure occurs at $k=73$ when it was replaced by a new one and the

* Throughout the paper, all torque quantities are normalized to the maximum allowed torque and are therefore dimensionless.

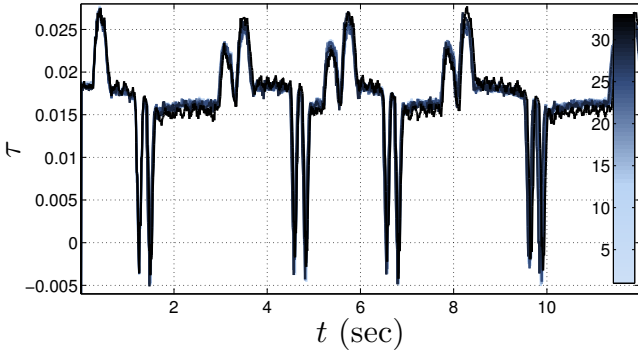


(a) Disturbances effects.

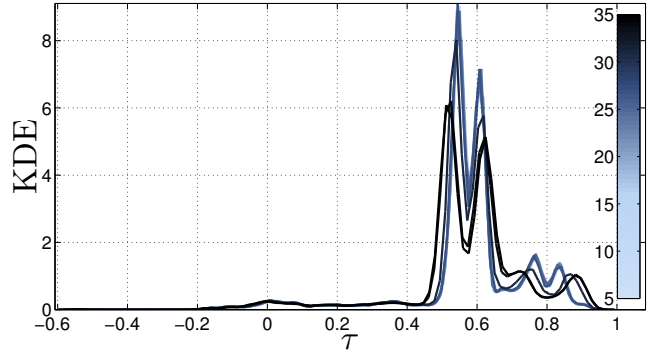


(b) Wear effects.

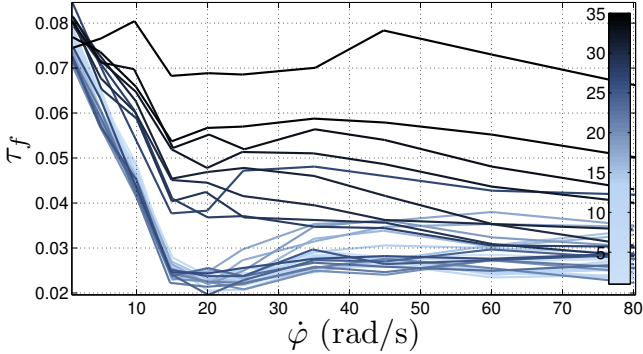
Fig. 2. Friction dependencies in a robot joint based on experimental studies. The offset values were removed for a comparison, their values are shown by the dotted lines. The data were collected from similar gearboxes and are directly comparable. Notice the different scales used and the larger amplitude of effects caused by temperature and load compared to those caused by wear. In (b), the colormap relates to the length of accelerated wear tests during which the curves were registered.



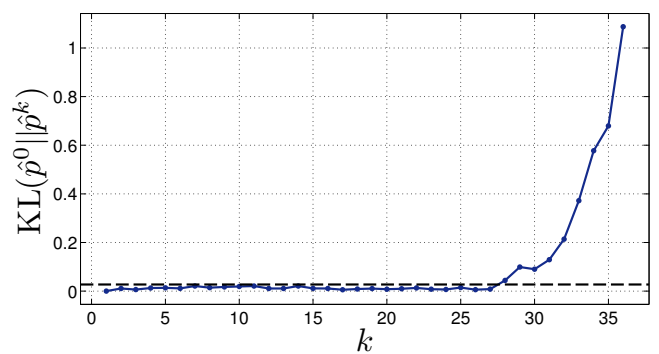
(a) Monitored torque data.



(b) Estimated KDEs.



(c) Friction curves.



(d) Fault indicator, $\text{KL}(\hat{p}^0 || \hat{p}^k)$.

Fig. 3. Monitoring of a wear fault in an industrial robot joint under accelerated wear tests and controlled load and temperature disturbances. A trajectory \mathcal{U} was executed repetitively through the experiments, and the colormaps relate to its k th execution and is chosen as to highlight increased friction values. The friction changes caused by wear were estimated during the experiments and are shown in (c) for a comparison. The monitored torque data are shown in (a), their respective KDEs were computed using a Gaussian kernel and are shown in (b). At $k=0$, it is considered that the robot is fault-free and the fault indicator given by $\text{KL}(\hat{p}^0 || \hat{p}^k)$ is shown in (d) where the dashed line represents an upper limit for the nominal behavior of the fault indicator. Notice the clear response of the fault indicator to the wear changes.

nominal data were thus reset for the new gearbox. The friction curves related to the faulty gearbox are shown in Figure 4(c), where it can be noticed that the changes due to wear start to appear around $k = 64$. The related KDEs for this gearbox are shown in Figure 4(a), where a similar behavior as in the previous case can be noticed, with changes in the size and position of the distributions' peaks. The KDEs for the replaced gearbox can be seen in Figure 4(b) where it is possible to notice that no significant variations are present. The fault indicator is shown in Figure 4(d), where, as in the previous case, the dashed lines represent the sum of mean and three times the standard deviation of the fault indicator when the gearboxes are considered healthy. The filled circle highlights the moment when the gearbox was replaced. As it can be seen, an early detection of the increased wear is made possible with the use of the proposed fault indicator, allowing for CBM.

4.2 Simulated Wear Monitoring under Temperature Disturbances

The experimental studies presented previously were based on experiments performed in a lab where only small variations of load and temperature were present. To illustrate the ideas to improve robustness presented in Section 3.3, simulation studies were carried out. The use of simulations allow for more detailed studies of the effects of the disturbances compared to what could be achieved based on experiments in a feasible manner. A realistic friction model is used in the simulation that can represent, amongst others, the effects of wear w and joint temperature T . See Appendix A for details of the simulation environment used.

4.2.1 Finding the weights \mathbf{w}

First, the weight vector \mathbf{w} must be found. According to the procedures described in Section 3.3, this requires the use of a labeled dataset \mathbf{Y}^M . This dataset is achieved here based on $M = M_0 + M_1$ simulations of a trajectory \mathcal{U} based on the same test-cycle used in Section 4.1. The first $M_0 = 100$ batches of data are generated for class \mathcal{C}_0 , under no presence of wear but with variations of temperature. The remaining $M_1 = 100$ batches contain the same characteristic of temperature variations and an increased wear level. The simulation setup for each class is according to

$$\mathcal{C}_0 : \quad \mathbf{w} = 0, \quad T \sim \mathcal{U}[\underline{T}, \underline{T} + \Delta_T] \quad (12a)$$

$$\mathcal{C}_1 : \quad \mathbf{w} = \mathbf{w}_c, \quad T \sim \mathcal{U}[\underline{T}, \underline{T} + \Delta_T] \quad (12b)$$

where $\mathbf{w}_c = 35$ is a wear level considered critical to generate an alarm (see [2] for details of the wear model). Here, T is considered random, with uniform distribution given by $\underline{T} = 30^\circ\text{C}$ and $\Delta_T = 40^\circ\text{C}$. This assumption is carried out for analysis purposes and allows for great variations of temperature disturbances.

The solution for the optimal weights given in (9) is proportional to the average changes found in the data, $\mu_i^1 - \mu_i^0$, while the solution given by (10) relates to the ratio between these quantities and the total variability, $s_i^1 + s_i^0$. These quantities are computed based on the labeled dataset and are displayed in Figure 5(a) as a function of the joint speed $\dot{\varphi}$. As it can be seen, the optimal weights present a strong correlation with $\dot{\varphi}$. This is not a surprise since the effects of w and T depend on $\dot{\varphi}$, recall Figure 2. In the same figure, worst case estimates along speed are also shown (solid lines), i.e. $\mu_i^1 - \mu_i^0$ closest to zero and largest $s_i^1 + s_i^0$. Figure 5(b) presents the ratio for such worst case estimates, which is considered as the optimal weights according to (10).

The solid line in Figure 5(b) is a function approximation of the optimal weights given by

$$w(\dot{\varphi}) = \text{sech}(\beta\dot{\varphi}) \tanh(\alpha\dot{\varphi}) \quad (13)$$

with $\alpha = 1.45 \cdot 10^{-2}$ and $\beta = 4.55 \cdot 10^{-2}$. The parametrization of the weight vector as a function of $\dot{\varphi}$ allows for a more general use of the optimal weights, e.g. the same weighting function can be used for other trajectories. Effectively, the optimal weighting function selects a speed region that is more relevant for robust wear monitoring, giving less relevance for data collected at speeds close to zero or higher than 100 rad/s. A similar behavior was found in [2] for the quality (variance) of a wear estimate for different speeds under temperature disturbances.

4.2.2 Robustness improvements

The improvements in robustness achieved using the weighting function can be illustrated by considering the detection of an abrupt change of w from 0 to w_c . Considering a dataset generated according to (12), a pair $(\boldsymbol{\tau}^m, \boldsymbol{\tau}^n)$ is given and the objective is to decide whether the pair is from the same class or not, that is, the two hypotheses are considered

$$\mathcal{H}_0 : \quad m, n \in \mathcal{C}_0 \quad \text{or} \quad m, n \in \mathcal{C}_1 \quad (14a)$$

$$\mathcal{H}_1 : \quad m \in \mathcal{C}_0, n \in \mathcal{C}_1 \quad \text{or} \quad m \in \mathcal{C}_1, n \in \mathcal{C}_0. \quad (14b)$$

In view of the framework presented in Section 3, this problem is analyzed by computing the distribution of $\text{KL}(\hat{p}^m || \hat{p}^n)$ for each hypothesis, i.e. $p(\text{KL} | \mathcal{H}_0)$ and $p(\text{KL} | \mathcal{H}_1)$. The density $p(\text{KL} | \mathcal{H}_0)$ should concentrate values close to zero, indicating that no change is present while $p(\text{KL} | \mathcal{H}_1)$ should contain large positive values, clearly indicating the change. Applying the weighting function to the pair $(\boldsymbol{\tau}^m, \boldsymbol{\tau}^n)$ will hopefully provide more separation between the resulting densities.

The overlap of these distributions relates to how difficult it is to make a correct decision, i.e. whether a change is present or not. Given the value of the fault indicator,

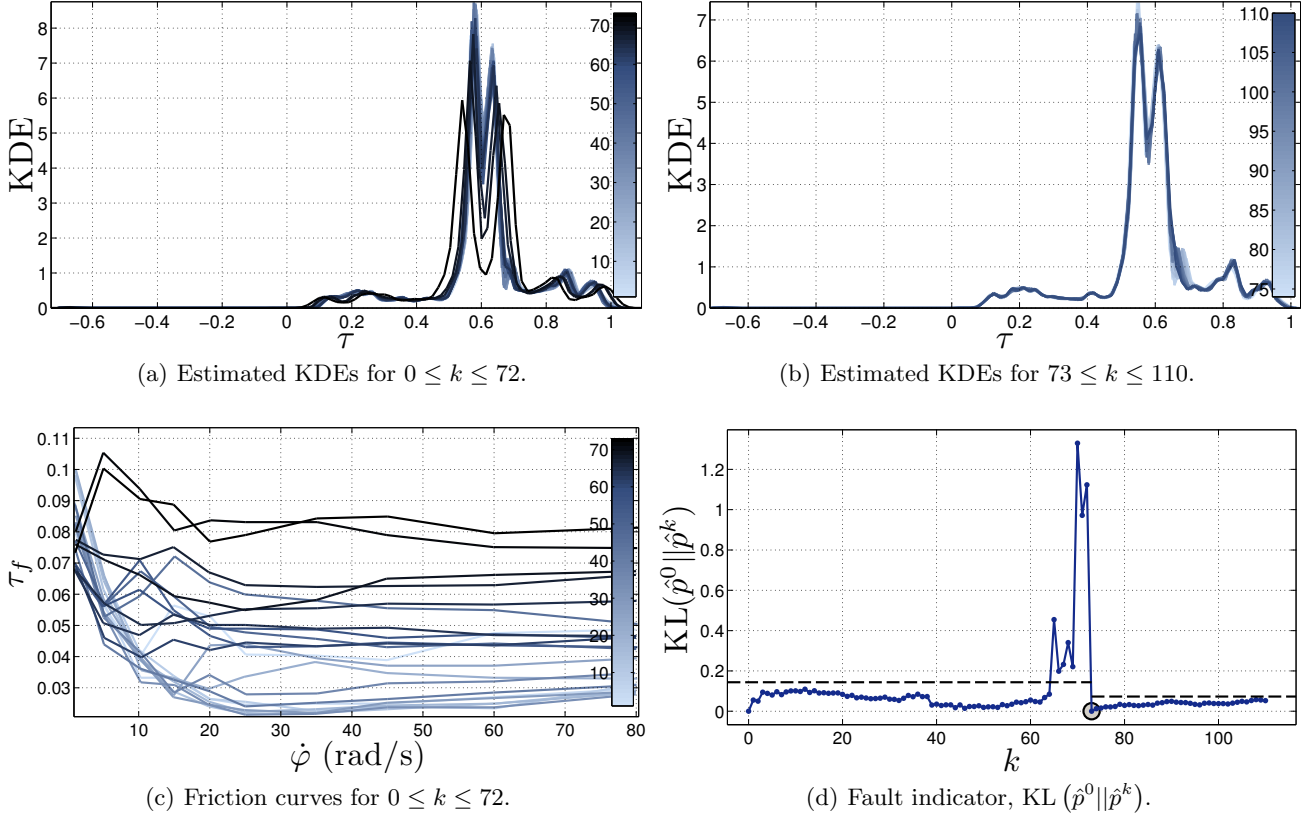


Fig. 4. Monitoring of a wear fault in an industrial robot joint under accelerated wear tests and controlled load and temperature disturbances. Data collected from the same trajectory \bar{U} used in Figure 3 are considered. A wear fault develops in the gearbox from $k=0$ to $k=72$, whereafter the faulty gearbox is replaced by a new one. The KDEs for the faulty gearbox are shown in Figure 4(a), which presents a similar behavior as for the previous case, recall Figure 3(b); the respective friction curves are shown in Figure 4(c). The KDEs for the new gearbox are shown in Figure 4(b), where only small deviations are visible. The nominal data are assigned at $k=0$ and at $k=73$ before and after the replacement respectively. The resulting fault indicators are shown in Figure 4(d), with a clear response to the friction changes and regular behavior when no fault is present; the circle in the figure highlights when the replacement took place and the dashed lines represent an upper limit for the nominal behavior of the fault indicator.

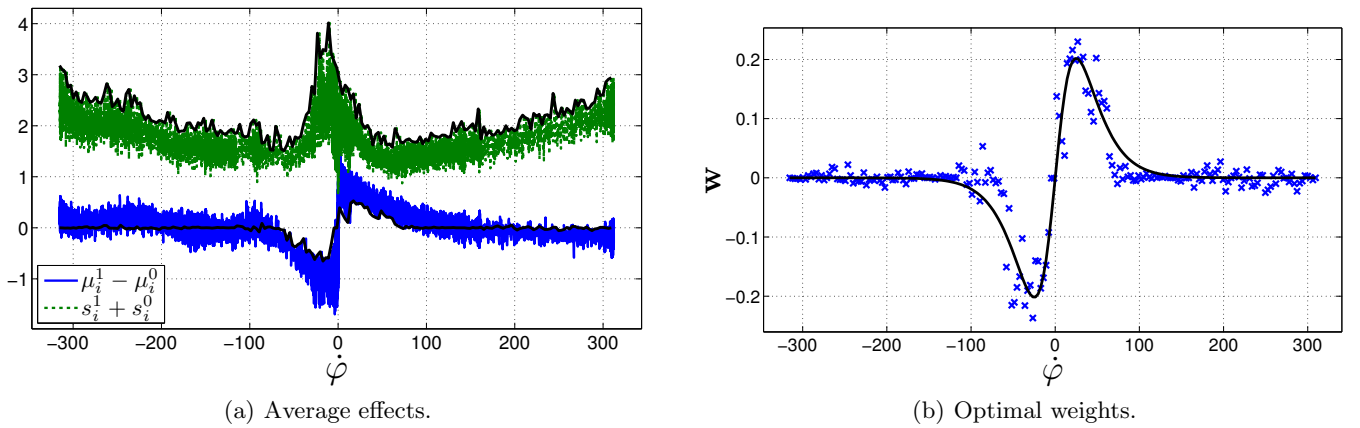


Fig. 5. Choice of optimal weights \mathbf{w} . The effects of disturbances by temperature and faults are shown in (a), together with a worst case estimate (solid lines). The optimal weights for the worst case estimate are shown in (b) together with a function approximation (solid). Notice how the optimal region for wear monitoring is concentrated in a narrow speed range.

a decision regarding which hypothesis is present can be made with a threshold check

$$\text{KL}(\hat{p}^m || \hat{p}^n) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \bar{h} \quad (15)$$

and reads, decide for \mathcal{H}_1 if $\text{KL}(\hat{p}^m || \hat{p}^n) \geq \bar{h}$ otherwise choose \mathcal{H}_0 . The decision mechanism and densities $p(\text{KL}|\mathcal{H}_0)$ and $p(\text{KL}|\mathcal{H}_1)$ define a *binary hypothesis test*^{*}. It is possible to evaluate the probabilities of a false detection P_f , i.e. accepting \mathcal{H}_1 when \mathcal{H}_0 is true, and of correct detection P_d , i.e. accepting \mathcal{H}_1 when \mathcal{H}_1 is true as

$$P_f = \int_{\bar{h}}^{\infty} p(x; \text{KL}|\mathcal{H}_0) dx, \quad P_d = \int_{\bar{h}}^{\infty} p(x; \text{KL}|\mathcal{H}_1) dx. \quad (16)$$

Notice that for a fixed P_f there is an associated \bar{h} (this is known as the Neyman-Pearson criterion for threshold selection) and therefore a P_d . For a satisfactory performance of the fault indicator, low P_f and high P_d are typically desirable.

For given values of w_c , \underline{T} and Δ_T , the trajectory based on the test-cycle is simulated using Monte Carlo simulation for the classes described in (12) and the hypotheses densities are estimated. The threshold is found using the Neyman-Pearson criterion for the fixed $P_f = 0.01$ and the related P_d is computed. For $w_c = 35$ and $\underline{T} = 30^\circ\text{C}$, Figure 6(a) presents the achieved P_d as a function of Δ_T with and without the use of the weighting function. Notice that the use of the weighting function considerably improves the robustness to temperature variations, but for too large Δ_T it becomes difficult to distinguish the effects.

A similar study can be performed to illustrate how w_c affects the performance. For the fixed $\Delta_T = 25^\circ\text{C}$ and $\underline{T} = 30^\circ\text{C}$, data are generated according to (12) for different values of w_c . The hypotheses' densities are estimated using Monte Carlo simulation. Figure 6(b) presents P_d as a function of w_c for the fixed $P_f = 0.01$. The improvements achieved using the weighted data are clear.

5 Conclusions and Future Work

The suggested framework considers the monitoring of changes in the distribution of the data batches. Because no prior knowledge is assumed about the data distribution, nonparametric kernel density estimates are used,

^{*} The presentation of the topic was put into the context of this paper. The topic is however common and is found, e.g., in detection theory, related to receiver operating characteristics, (see e.g. [20]) and classification problems (see e.g. [1]).

which give great flexibility, are simple to implement and have an inherent smoothing behavior. More studies are however needed regarding the selection of kernel functions and the bandwidth parameter. The effects of the use of different distances than the symmetrized Kullback-Leibler is also relevant.

The validity of the framework and methods were illustrated with promising results on real case studies and simulations for the wear monitoring problem in a robot joint. In the application, the execution of the same trajectory, based on a test-cycle, ensured a repetitive behavior of the robot. In general however, there might not be a trajectory that is repeated through all of the robot's lifetime. Nevertheless, trajectories are quite often repeated through a certain period. The study of approaches to relax Condition C-1 and extend the framework to applications where the repetitive behavior of the system varies are therefore important.

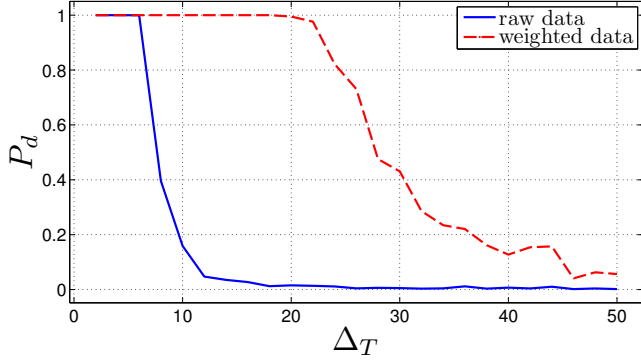
The fast increase of friction due to wear found in the experimental studies in Section 4.1 is common to other applications, as presented in [4]. Such transient behavior is important when determining the scheduling of the data collection. The transient behavior of wear is also related to the equipment's remaining lifetime, a quantity important for decision support of maintenance actions.

In the future, the framework should be considered to other types of mechanisms and failures. An important advantage of the framework presented is that no model of the system is required and modeling efforts are therefore not needed. Furthermore, it opens up for use in systems where a stationary behavior is difficult or not possible.

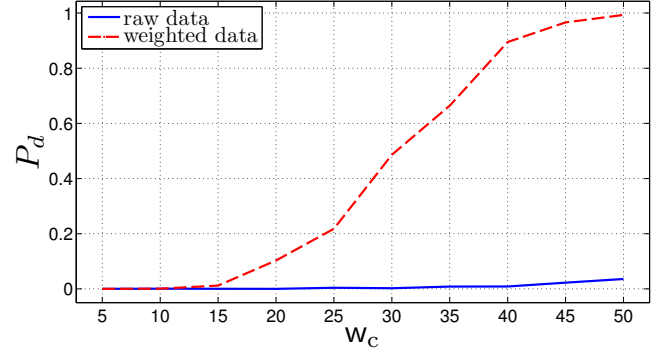
Finally, to achieve a diagnosis based on the suggested fault indicator, methods for alarm generation and fault isolation should be addressed in the future.

References

- [1] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, New York, USA, 1st edition, 2007.
- [2] André Carvalho Bittencourt, Patrik Axelsson, Ylva Jung, and Torgny Brogårdh. Modeling and identification of wear in a robot joint under temperature disturbances. In *Proc. of the 18th IFAC World Congress*, Milan, Italy, Aug 2011.
- [3] André Carvalho Bittencourt and Svante Gunnarsson. Static friction in a robot joint—modeling and identification of load and temperature effects. *Journal of Dynamic Systems, Measurement, and Control*, 134(5), 2012.
- [4] Peter J. Blau. Embedding wear models into friction models. *Tribology Letters*, 34(1), Apr. 2009.
- [5] Adrian W. Bowman and Adelchi Azzalini. *Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations* (Oxford Statistical



(a) Temperature variations.



(b) Fault size.

Fig. 6. Probability of detection P_d when $P_f=0.01$ for an abrupt fault with $w_c=35$ as a function of temperature variations Δ_T (a) and as function of the fault size w_c for $\Delta_T=25^\circ\text{C}$ (b). Notice the considerable improvements when using the weighted data.

- Science Series*). Oxford University Press, USA, November 1997.
- [6] D. Brambilla, L.M. Capisani, A. Ferrara, and P. Pisu. Fault detection for robot manipulators via second-order sliding modes. *IEEE Transactions on Industrial Electronics*, 55(11):3954–3963, Nov. 2008.
 - [7] F. Caccavale, P. Cilibrizzi, F. Pierri, and L. Villani. Actuators fault diagnosis for robot manipulators with uncertain model. *Control Engineering Practice*, 17(1):146–157, 2009.
 - [8] Fabrizio Caccavale and Luigi Villani, editors. *Fault Diagnosis and Fault Tolerance for Mechatronic Systems: Recent Advances*. Springer Tracts in Advanced Robotics, Vol. 1. Springer-Verlag, New York, 2003.
 - [9] A. De Luca and R. Mattone. An adapt-and-detect actuator FDI scheme for robot manipulators. In *Proc. of the 2004 IEEE International Conference on Robotics and Automation (ICRA)*, volume 5, pages 4975–4980 Vol.5, Barcelona, Spain, apr. 2004.
 - [10] Ikbāl Eski, Selcuk Erkaya, Sertaç Savas, and Sahin Yildirim. Fault detection on robot manipulators using artificial neural networks. *Robotics and Computer-Integrated Manufacturing*, 27(1):115–123, Jul 2011.
 - [11] Xianfeng Fan and Ming J. Zuo. Gearbox fault detection using hilbert and wavelet packet transform. *Mechanical Systems and Signal Processing*, 20(4):966–982, 2006.
 - [12] B. Freyermuth. An approach to model based fault diagnosis of industrial robots. In *Proc. of the 1991 IEEE International Conference on Robotics and Automation (ICRA)*, volume 2, pages 1350–1356, Sacramento, USA, Apr 1991.
 - [13] Faisal I. Khan and S. A. Abbasi. Major accidents in process industries and an analysis of causes and consequences. *Journal of Loss Prevention in the Process Industries*, 12(5):361–378, 1999.
 - [14] L. Marton. On-line lubricant health monitoring in robot actuators. In *Proc. of the 2011 Australian Control Conference (AUCC)*, pages 167–172, Melbourne, Australia, nov. 2011.
 - [15] Stig Moberg. *On Modeling and Control of Flexible Manipulators*. PhD thesis, Linköpings Universitet, 2007.
 - [16] Stig Moberg, J. Öhr, and Svante Gunnarsson. A benchmark problem for robust control of a multivariable nonlinear flexible manipulator. In *Proc. of the 17th IFAC*

World Congress, Mar 2008.

- [17] Ella Olsson, Peter Funk, and Ning Xiong. Fault diagnosis in industry using sensor readings and case-based reasoning. *Journal of Intelligent & Fuzzy Systems*, Vol. 15:10, December 2004.
- [18] B. K. N. Rao. Condition monitoring and the integrity of industrial systems. In A. Davies, editor, *Part 1: Introduction to Condition Monitoring*, Handbook of Condition Monitoring – Techniques and Methodology, chapter 1, pages 3–34. Chapman & Hall, London, UK, 1998.
- [19] Mark D. Reid and Robert C. Williamson. Information, divergence and risk for binary experiments. *Journal of Machine Learning Research*, 12:731–817, 2011.
- [20] Harry L. Van Trees. *Detection, Estimation and Modulation Theory, Part I*. Wiley, New York, 2001.
- [21] Arun T. Vemuri and Marios M. Polycarpou. A methodology for fault diagnosis in robotic systems using neural networks. *Robotica*, 22(04):419–438, 2004.

A Simulation Model

The simulation model considered is the 2 link manipulator with elastic gear transmission presented in the benchmark problem of [16]. The simulation model is representative of many of the phenomena present in a real manipulator, such as,

- measurement noise,
- coupled inertia,
- torque ripple,
- torque disturbances,
- nonlinear stiffness,
- closed loop operation.

With the objective of studying friction changes related to wear in a robot joint, the static friction model described in [2] is included in the simulation model. The static friction model was developed from empirical studies in a robot joint and describes the effects of angular speed $\dot{\varphi}$, manipulated load torque τ_l , temperature T , and wear w .

In the simulation setup, a trajectory \mathcal{U} is described by a set of reference joint positions through time to the robot, which is controlled with feedforward and feedback control actions, guaranteeing the motion performance. If no variations of w

and T are allowed, the torque sequence required for the execution of a task \mathcal{U} varies only slightly due to the stochastic components and feedback.