A Data-Driven Approach to Diagnostics of Repetitive Processes in the Distribution Domain - Applications to Gearbox Diagnostics in Industrial Robots and Rotating Machines

André Carvalho Bittencourt [a], Kari Saarinen [b], Shiva Sander-Tavallaey [b], Svante Gunnarsson [a], Mikael Norrlöf [c,a]

[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 data-driven approach to diagnostics 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 an available 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 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 approach is demonstrated with successful experimental and simulation applications to wear diagnostics in an industrial robot gearbox and for diagnostics of gear faults in a rotating machine.

Key words: Repetitive processes; Diagnostics; Fault Detection; Data-driven methods; Industrial Robots; Wear; Vibration Analysis; Rotating Machinery; Gearbox

1 Introduction

In the manufacturing industry, preventive scheduled maintenance is a common approach used to improve equipment's safety, reliability, availability and maintainability. 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

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).

URL: http://www.control.isy.liu.se/ andrecb (André Carvalho Bittencourt).

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 the output of a *system model*, see e.g. [25,27]. A system model is a map from input to output data. A system model provides important information about the behavior of the system, facilitating 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.

The use of model-based approaches is common for diagnostics of machines. For robotics, many approaches have been suggested based on the use of nonlinear observers, where the observer stability is typically guaranteed by analyses of the decay rate of a candidate Lyapunov function, see e.g. [20,30,10,22,15,12,9]. Observers can also be designed based only on data, without a description of the system based on first principles. Data-driven design of observers are typically based on subspace identification of linear models and have been suggested for fault detection in [14,42,16,44]. Parameter estimation is also a natural approach to model-based diagnostics because of the physical interpretation of the system parameters, see e.g. [21,29,5].

In cases where the data are ordered with time, signal-driven methods are common for machinery diagnostics. These are typically based on the use of integral transforms, e.g. Fourier, Radon, Karhunen-Loève or Wavelet. Each transform will enhance different properties in the transformed domain and are suitable depending on the characteristics of the signal, e.g. periodic, stationary, etc. The analysis of data in the frequency domain or time-frequency has found particular success in the monitoring of rotating machines, see e.g. [40,13,19,23,38,24,25]. Some approaches have also been proposed for the diagnostics of industrial robots with the use of additional sensory information [32,18].

A common challenge to data-driven methods is that the data characteristics' will vary depending on the operating points, which may complicate determination of fault presence. This is particularly restricting for an industrial robot where the kinematic configuration of the robot may give varying load torques at the joints during motion. This shortcoming can be circumvented by considering data from a specific operation of the system, e.g. under repetition. A *repetitive operation* is found in various applications, e.g. in automated manufacturing. Repetition can also be forced with the execution of specific diagnostic routines but with the drawback of reduced availability. Much attention has been given recently to repetitive processes [36,37]. Study of repetitive processes have mainly focused on control [35,39] and estimation problems [2,1]. A few approaches have been also suggested for model-based diagnostics, e.g. [43].

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 fault indicator proposed here 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. A distribution domain approach does not consider the dynamics of the system generating the data as is the case in, e.g., observer-based approaches. As it will be presented, this leads to very simple diagnostics solutions that can perform well in practical setups.

The proposed framework was initially developed with the interest focused on the diagnostics of wear in industrial robots and a preliminary version of the work can be found in [7]. Here, more aspects are covered, including approaches to detection, isolation and reduction of sensitivity to disturbances. Also, more experimental and simulation results are presented for the robotics application. An additional application is also included for diagnostics of rotating machinery based on vibration data collected from an accelerometer. The paper is organized as follows; a general presentation of data-driven diagnostics and repetitive systems is given in Section 2, followed by the presentation of the proposed approach for diagnostics in the distribution domain in Section 3. The applications are presented in Sections 4 and 5. Conclusions and future work are given in Section 6.

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

2 Data-Driven Diagnostics and Repetitive Systems

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

$$\mathbf{Y}_K = [\mathbf{y}_1, \cdots, \mathbf{y}_k, \cdots, \mathbf{y}_K], \tag{1}$$

where $\mathbf{y}_k = [y_{k,1}, \cdots, y_{k,n}, \cdots, y_{k,N}]^T$ denotes the data vector in $\mathbb{R}^N$ (e.g. measurements or known inputs) with batch index k and element index n. The sequence $\mathbf{y}_k$ could have been generated as the result of deterministic and stochastic inputs, $\mathbf{Z}_K$ and $\mathbf{V}_K$, where $\mathbf{V}_K$ is unknown, and $\mathbf{Z}_K$ may have known and unknown components. For example, the data generation mechanism could be modeled as

$$\mathbf{y}_k = h(\mathbf{z}_k, \mathbf{v}_k), \tag{2}$$

where $h(\cdot)$ is an unknown function. Let the set of deterministic inputs $\mathbf{Z}_K$ be categorized in three distinct groups, $\mathbf{R}_K$, $\mathbf{D}_K$ and $\mathbf{F}_K$. The sequence $\mathbf{f}_k$ is unknown and of interest (a fault), $\mathbf{r}_k$ is known (e.g. references or control inputs) and $\mathbf{d}_k$ is unknown (disturbances).

The objective is to define a data-driven framework for the generation of fault indicators to determine the presence of a fault $\mathbf{f}_k$. Because a data-driven approach depends on availability of data, it is assumed that data generated under no fault is available. Let $\mathcal{Y}^0 = \{\mathbf{y}_k : \mathbf{f}_k = \mathbf{0}\}$ denote the set of data batches that were generated under no fault, the following assumption is made:

A-1 **(Nominal data are available)** A sequence $\mathbf{y}^0 \in \mathcal{Y}^0$ is available.

The rationale is then to generate fault indicators from the comparison of the nominal data $\mathbf{y}^0$ (available from Assumption A-1) against the remaining sequences $\mathbf{y}_k$. In order to generate fault indicators for $\mathbf{y}_k$ using the nominal data $\mathbf{y}^0$, two basic questions arise:

Q-1 How to characterize a sequence $\mathbf{y}_k$?
Q-2 How to compare the sequences $\mathbf{y}^0$, $\mathbf{y}_k$?

The first question targets the issue of finding a data processing mechanism of $\mathbf{y}_k$, written in a general form as $g_k \triangleq g(\mathbf{y}_k) : \mathbb{R}^N \mapsto \mathbb{G}$ with domain $\mathbb{G}$, whose output enhances the ability to discriminate the presence of non-zero $\mathbf{f}_k$. Given the nominal data in the transformed domain $g^0 \triangleq g(\mathbf{y}^0)$, fault indicators can be achieved from the comparison between g^0 and g_k. This is typically, but not necessarily, done with the use of a distance function represented as $d(g^0, g_k) : \mathbb{G} \times \mathbb{G} \mapsto \mathbb{R}_0^+$. Different distances are possible depending on the domain $\mathbb{G}$. For example, for diagnostics of rotating machines g_k could be the spectra of $\mathbf{y}_k$ and $d(\cdot, \cdot)$ a spectral distance, see e.g.[3].

2.1 Detection, Performance and Isolation

Let $\mathcal{D}^{m,n} = \{d(g_i, g_j) : \mathbf{y}_i \in \mathcal{Y}^m, \mathbf{y}_j \in \mathcal{Y}^n\}$, then $\mathcal{D}^{0,0}$ describes the behavior of the fault indicator when no fault is present and $\mathcal{D}^{0,f}$, where $\mathcal{Y}^f = \{\mathbf{y}_k : \mathbf{f}_k \neq \mathbf{0}\}$, describes all possible faulty behaviors. A criterion for **detectability** of an abnormality is that $\mathcal{D}^{0,f}$ is not completely contained in $\mathcal{D}^{0,0}$, i.e. $\mathcal{D}^{0,f} \backslash \mathcal{D}^{0,0} \neq \emptyset$. Since the distance $d(g^0, g_k)$ measures how far g_k is from the nominal g^0, it is expected that it will remain close to zero if $d(g^0, g_k) \in \mathcal{D}^{0,0}$ and to deviate to positive values if $d(g^0, g_k) \in \mathcal{D}^{0,f}$. Suppose that it is possible to find a threshold $\hbar$ such that $d(g^0, g_k) \leq \hbar$ most of the times when $d(g^0, g_k) \in \mathcal{D}^{0,0}$, a simple criterion for **detection** is then to consider a threshold check. Let $\mathcal{H}^0$ denote the hypothesis that no fault is present, i.e. $\mathbf{y}_k \in \mathcal{Y}^0$, and $\mathcal{H}^1$ denote the alternative hypothesis that a fault is present, i.e. $\mathbf{y}_k \in \mathcal{Y}^f$, then the decision mechanism is

$$d(g^0, g_k) \underset{\mathcal{H}^0}{\overset{\mathcal{H}^1}{\gtrless}} \hbar \tag{3}$$

and reads, decide for $\mathcal{H}^1$ if $d(g^0, g_k) \geq \hbar$ otherwise decide for $\mathcal{H}^0$. The probabilities of error for this decision rule can be quantified given the probability distribution of the fault indicator under the different hypotheses, denoted

$p(d|\mathcal{H}^0)$ and $p(d|\mathcal{H}^1)$. The probability of a false detection P_f, i.e. deciding for a fault when none is present, and of correct detection P_d, i.e. deciding for a fault when it is present can be evaluated by

$$P_f = \int_{\hbar}^{\infty} p(x; d|\mathcal{H}^0)\,\mathrm{d}x, \quad P_d = \int_{\hbar}^{\infty} p(x; d|\mathcal{H}^1)\,\mathrm{d}x. \tag{4}$$

Notice that for a fixed P_f there is an associated $\hbar$ (this is known as the Neyman-Pearson criterion for threshold selection [41]) and therefore a P_d. For a satisfactory **performance** of the fault indicator, low P_f and high P_d are typically desirable.

Since this is a data-driven framework, data from the different fault types are needed a priori to address the isolation problem, i.e. determination of the fault type present. Let $\mathcal{Y}^m = \{\mathbf{y}_k : \mathbf{f}_k = \mathbf{f}^m\}$ denote the set of data generated under the mth fault type with M possible fault modes, i.e. $m \in \mathcal{M} = [1, 2, \cdots, M]$. For fault isolation it is assumed:

A-2 **(Data from each fault type are available)** A sequence $\mathbf{y}^m \in \mathcal{Y}^m$ is available for each mth fault type.

Once a fault is detected, the fault type can be determined by choosing $\mathbf{y}^m$ that is closest to $\mathbf{y}_k$ in the sense of the transformation $g(\cdot)$ and distance $d(\cdot,\cdot)$. That is, decide for presence of the m^* fault type according to

$$m^* = \arg\min_{m\in\mathcal{M}} d(g^m, g_k). \tag{5}$$

2.2 Repetitive Systems

The amount of overlaps between the sets $\mathcal{D}^{0,0}$ and $\mathcal{D}^{0,f}$ relates to how difficult it will be to determine the presence of a fault. Ultimately, no overlap is present, i.e. $|\mathcal{D}^{0,0} \cap \mathcal{D}^{0,f}| = 0$, when it is possible to determine the presence of a fault with no errors. This is particularly difficult to achieve since $\mathbf{y}_k$, and thus $\mathcal{D}^{0,0}$ and $\mathcal{D}^{0,f}$, are affected by other inputs than faults. Because $|\mathcal{D}^{0,0} \cap \mathcal{D}^{0,f}| \leq \min|\mathcal{D}^{0,0}|, |\mathcal{D}^{0,f}|$, an attempt to try to reduce the overlaps is to reduce the size of the sets themselves. This can be achieved by restricting the behavior of the data $\mathbf{y}_k$ or, conversely, by *restricting the possible input space*. This paper focuses on monitoring data collected from a **repetitive operation** of the system. A repetitive operation is meant by a system that executes the same trajectory $\mathbf{r}$ over each data batch collected. This type of data is represented by the set

$$\mathcal{Y}_r = \{\mathbf{y}_k : \mathbf{r}_k = \mathbf{r}\}. \tag{6}$$

Monitoring data $\mathbf{y}_k \in \mathcal{Y}_r$ simplifies the problem since the behavior of the data will be more predictable. Examples of systems that behave repetitively are common in automation applications. A repetitive operation can also be forced in case $\mathbf{r}$ can be chosen freely; for instance data can be collected based on the execution of scheduled diagnostics tests. This setup is also commonly found in vibration analyses and in signal-based diagnostics, where data are collected under particular operating conditions, e.g. of speed, load and acceleration.

Considering that faults are detectable for a system operating under repetition, there may still be overlaps between the sets $\mathcal{D}_r^{0,0}$ and $\mathcal{D}_r^{0,f}$ defined by $\mathcal{Y}_r^0 = \mathcal{Y}^0 \cap \mathcal{Y}_r$ and $\mathcal{Y}_r^f = \mathcal{Y}^f \cap \mathcal{Y}_r$. This is due to the effects of disturbances $\mathbf{d}_k$ and noise $\mathbf{v}_k$. By collecting data in a controllable manner, it might be possible to ensure repetition of the disturbance term, i.e. to have $\mathbf{d}_k = \mathbf{d}$ for all batches. This will be however too restrictive for most applications. Even if $\mathbf{d}_k = \mathbf{d}$ is possible, the noise components will always affect the data. To broaden the scope of the framework, a clear determination of the fault presence despite $\mathbf{d}_k$ and $\mathbf{v}_k$ is desirable, leading to the question:

Q-3 How to handle non-repetitive disturbances $\mathbf{d}_k$ and noise $\mathbf{v}_k$?

Questions Q-1 to Q-3 are addressed in the next section which defines the suggested approach for diagnostics of repetitive systems in the distribution domain.

3 A Distribution Domain Approach

3.1 Characterizing the Data – Kernel Density Estimate

The alternative pursued in this work is to consider the distribution of $\mathbf{y}_k$, which contains information about the *amplitude behavior* of the data. Even though information contained in the ordering may be lost, this is a valid approach since the effects of a fault often appear as changes in amplitude. Since data batches from a repetitive operation are considered, i.e. $\mathbf{y}_k \in \mathcal{Y}_r$, it is expected that the data distribution will remain similar in case no fault is present. 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 $p(y)$ based on the data vector $\mathbf{y}$ can be achieved from the empirical characteristic function. For a scalar random variable with probability density function $p(y)$, the characteristic function $\vartheta : \mathbb{R} \to \mathbb{C}$ with argument ν is defined as, [17]:

$$\vartheta(\nu) = \mathbb{E}\left[e^{\mathrm{i}\nu y}\right] = \int_{-\infty}^{\infty} e^{\mathrm{i}\nu y} p(y)\,\mathrm{d}y = \mathbf{F}^{-1}\{p(y)\}2\pi, \tag{7}$$

where $\mathbf{F}^{-1}\{\cdot\}$ is the inverse Fourier transform. So the density function can be found from the characteristic function through its Fourier transform. Given the sample $\mathbf{y} = [y_1, \cdots, y_N]$, the empirical estimate of $\vartheta(\nu)$ is given by

$$\widehat{\vartheta}(\nu) = \frac{1}{N}\sum_{n=1}^{N} e^{\mathrm{i}\nu y_n}, \tag{8}$$

the objective is then to estimate the density function from $\widehat{\vartheta}(\nu)$. This is essentially a spectrum estimation problem. A direct estimation of the density function from the Fourier transform of $\widehat{\vartheta}(\nu)$ will however lead to an estimate with increased variance for large values of ν, [28]. To avoid this problem, the empirical estimate of the characteristic function is multiplied with a weighting function $\psi_h(\nu) = \psi(h\nu)$. The weighting function is typically symmetric, satisfying $\psi(0) = 1$ and tends to zero when ν tends to infinity. The density estimate is then given by

$$\begin{aligned}\widehat{p}(y) &= \frac{1}{2\pi}\mathbf{F}\left\{\widehat{\vartheta}(\nu)\psi(h\nu)\right\} = \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{1}{N}\sum_{n=1}^{N} e^{\mathrm{i}\nu(y_n - y)}\psi(h\nu)\,\mathrm{d}\nu \\ &= \frac{1}{Nh}\sum_{n=1}^{N}\kappa\left(\frac{y_n - y}{h}\right) = \frac{1}{N}\sum_{n=1}^{N}\kappa_h(y - y_n),\end{aligned} \tag{9}$$

where $\kappa_h(y)h = \mathbf{F}^{-1}\{\psi_h(\nu)\}$. The function $\kappa_h(y)$ is a *kernel function*, satisfying $\kappa_h(\cdot) \geq 0$ and that integrates to 1. The resulting estimate is known as a *kernel density estimate* (KDE) and can also be generalized to the multidimensional case [11]. The bandwidth parameter h controls the smoothness of the resulting estimate, increasing the smoothness for larger values of h. When $h \to 0$, the kernel function approaches a Dirac delta and the resulting estimate will be a set of impulses located at the data points. Typical kernel functions and their Fourier transforms are shown in Figure 1. In this work, a Gaussian kernel is used with h optimized for Gaussian distributions, see [8]. For a detailed treatment of kernel density estimators and criteria/methods for choosing h see [33,11,8,26].

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, [34]

$$D_{\mathrm{KL}}\left(p||q\right) \triangleq -\int_{-\infty}^{\infty} p(y)\log\frac{q(y)}{p(y)}\,\mathrm{d}y \tag{10}$$

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

$$\mathrm{KL}\left(p||q\right) \triangleq D_{\mathrm{KL}}\left(p||q\right) + D_{\mathrm{KL}}\left(q||p\right), \tag{11}$$

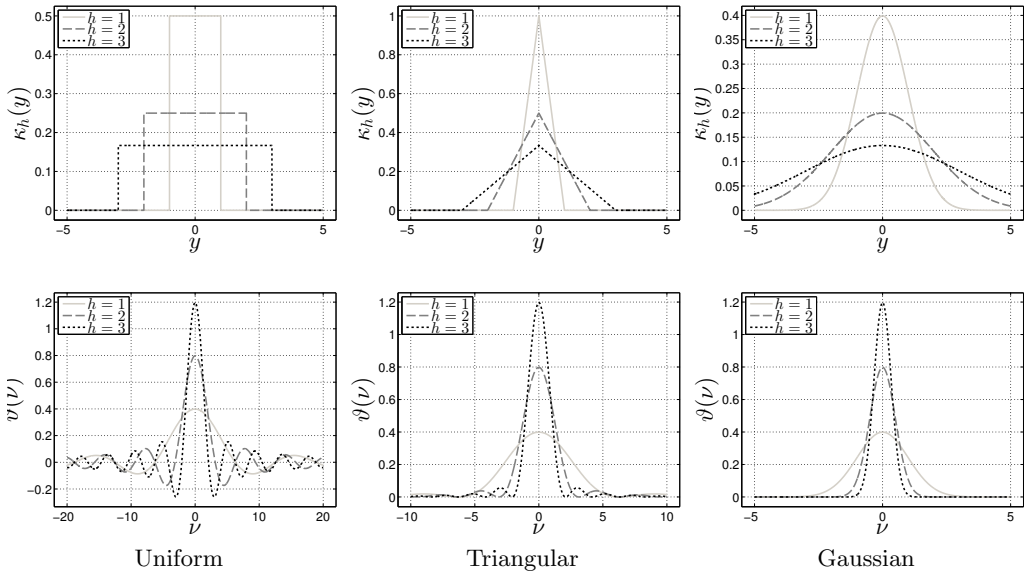


Fig. 1. Kernel functions (upper row) and their respective Fourier transforms. The parameter h controls the smoothness of the density estimate, larger values will give smoother estimates.

known as the Kullback-Leibler distance, is however symmetric. An answer to Question Q-2 can therefore be given with the use of the KL distance defined in (11). From Assumption A-1, fault-free data are always available, so that $\mathbf{y}^0$ is known and $\widehat{p}^0(y)$ can be computed. The quantities $\mathrm{KL}\left(\widehat{p}^0||\widehat{p}_k\right)$ can therefore be used as a fault indicator, remaining close to 0 in case $\widehat{p}^0(y)$ is close to $\widehat{p}_k(y)$ and otherwise deviating to positive values.

3.3 Handling Non-repetitive Disturbances and Noise – Data Weighting

One way to address Question Q-3 is to weight the raw data $\mathbf{y}_k$ according to prior knowledge of the effects of faults, disturbances and noise in order to give more relevance to parts of the data that relate to a fault. The approach considered here will assume availability of a *labeled dataset*, where the fault status (present or not) is known to each component $\mathbf{y}_k$ and is the same to each of its elements $y_{k,n}$. The disturbance vector $\mathbf{d}_k$ and noise $\mathbf{v}_k$ should contain variations that are expected to be found during the system's operation.

The dataset $\mathbf{Y}_K = [\mathbf{Y}^0_{K_0},\, \mathbf{Y}^f_{K_f}]$ contains K_0 fault-free data, $\mathbf{y}^0 \in \mathcal{Y}^0_r$, and $K_f = K - K_0$ faulty data $\mathbf{y}^f \in \mathcal{Y}^f_r$. Each batch $\mathbf{y}_k$ is weighted according to

$$\bar{\mathbf{y}}_k \triangleq \mathbf{w} \circ \mathbf{y}_k, \tag{12}$$

where $\circ$ denotes the Hadamard product (element-wise multiplication). This yields the weighted dataset

$$\overline{\mathbf{Y}}_K \triangleq \left[\overline{\mathbf{Y}}^0_{K_0}, \overline{\mathbf{Y}}^f_{K_f}\right] = \left[\bar{\mathbf{y}}^0_1, \cdots, \bar{\mathbf{y}}^0_{K_0}, \bar{\mathbf{y}}^f_1, \cdots, \bar{\mathbf{y}}^f_{K_f}\right]. \tag{13}$$

The objective is to choose $\mathbf{w}$ to maximize the sensitivity to faults while decreasing sensitivity to disturbances and noise. 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 $\mathrm{KL}(\widehat{\bar{p}}_i||\, \widehat{\bar{p}}_j)$, where $\widehat{\bar{p}}_i$ is the KDE of $\bar{\mathbf{y}}_i$ and therefore dependent on $\mathbf{w}$. When $\mathbf{y}_i \in \mathcal{Y}^0_r$ and $\mathbf{y}_j \in \mathcal{Y}^f_r$, the distance should be maximized and if $\mathbf{y}_i, \mathbf{y}_j \in \mathcal{Y}^0_r$, it should be minimized. A general solution to this problem is however difficult since it depends on how $\widehat{\bar{p}}_k(y)$ is computed (e.g. the kernel function chosen) and optimization over (11).

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. [4]. In LDA, the inner product $\mathbf{w}^T\mathbf{y}$ is used instead of the Hadamard product. 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 $\widehat{\bar{p}}_k$. Furthermore, the objective in LDA is to obtain a classifier. Here, in contrast, $\mathbf{w}$ is chosen as to achieve *average separation* between faulty and fault-free data while giving small variability to disturbances and noise.

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 $w(\cdot)$ such that $w_n = w(y_{k,n})$. 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 average difference between the means of the datasets $\mathbf{Y}^0_{K_0}$ and $\mathbf{Y}^f_{K_f}$. For $\mathbf{Y}^0_{K_0}$ (and similarly for $\mathbf{Y}^f_{K_f}$), the average of the weighted data is given by

$$\bar{\mu}^0 \triangleq \frac{1}{N}\sum_{n=1}^{N}\left[\frac{1}{K_0}\sum_{k=1}^{K_0} w_n y^0_{k,n}\right] = \frac{1}{N}\sum_{n=1}^{N} w_n \underbrace{\left[\frac{1}{K_0}\sum_{k=1}^{K_0} y^0_{k,n}\right]}_{\triangleq \mu^0_n} = \frac{1}{N}\mathbf{w}^T\boldsymbol{\mu}^0.$$

The average distance $\bar{\mu}^f - \bar{\mu}^0$ is thus proportional to $\mathbf{w}^T(\boldsymbol{\mu}^f - \boldsymbol{\mu}^0)$ 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. [4, Exercise 4.4]),

$$\mathbf{w}^* \propto (\boldsymbol{\mu}^f - \boldsymbol{\mu}^0). \tag{14}$$

A criterion based only on the distance between the datasets means does not consider the variability found within each dataset, e.g. caused by disturbances and noise. An alternative is to consider maximum average separation while giving small variability within each dataset. The average value of the weighted variance vector over k for $\mathbf{Y}^0_{K_0}$ (and similarly for $\mathbf{Y}^f_{K_f}$) is given by

$$\bar{s}^0 \triangleq \frac{1}{N}\sum_{n=1}^{N}\left[\frac{1}{K_0}\sum_{k=1}^{K_0}(w_n y_{k,n} - w_n\mu^0_n)^2\right] = \frac{1}{N}\sum_{n=1}^{N} w_n^2 \underbrace{\left[\frac{1}{K_0}\sum_{k=1}^{K_0}(y_{k,n} - \mu^0_n)^2\right]}_{\triangleq s^0_n} = \frac{1}{N}\mathbf{w}^T S^0 \mathbf{w},$$

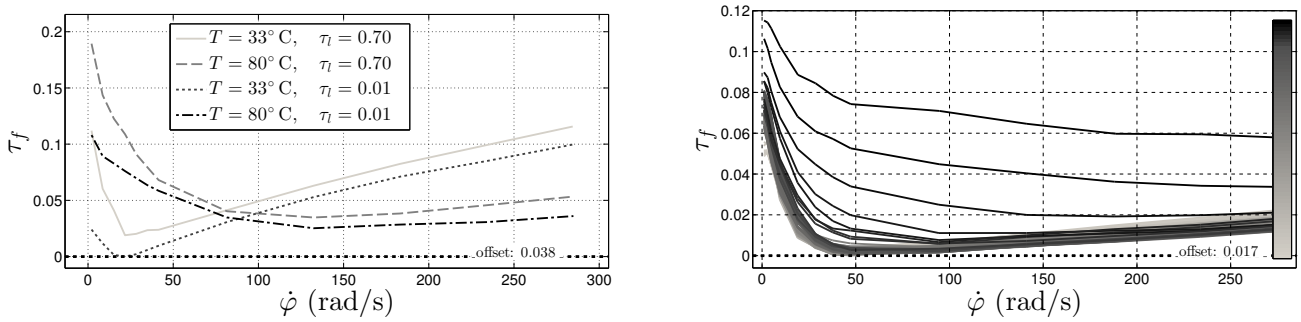
where S^0 is a diagonal matrix with diagonal elements given by s^0_n. Defining the total within class variation as $\bar{s}^f + \bar{s}^0$, the following criterion can be used

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

which is a special case of the Fisher criterion in LDA. It can be shown, see e.g. [4], that solutions for this problem satisfy

$$\mathbf{w}^* \propto (S^f + S^0)^{-1}(\boldsymbol{\mu}^f - \boldsymbol{\mu}^0). \tag{15}$$

That is, each weight w^*_n is proportional to the ratio between the average changes caused by faults, $\mu^f_n - \mu^0_n$, and the total variability found in the data, $s^f_n + s^0_n$, caused by disturbances and noise.



(a) Effects of load and temperature.

(b) Effects of wear.

Fig. 2. Friction levels τ_f in a robot joint as a function of motor speed $\dot{\varphi}$ under different conditions of wear, temperature and load. 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.

4 Wear Monitoring in an Industrial Robot Joint

In this first application, the objective is to determine the presence of excessive levels of wear, w, in the gearbox of a robot joint. Because increased wear levels may lead to increased friction in the gearbox, it is possible to monitor friction to infer about wear. Since the friction torques must be overcome by the applied motor torques, τ, during its operation, it is possible to extract information about friction from available data. Friction is however not only affected by wear, but also by unknown disturbances, such as variations of load torques, τ_l, and the lubricant temperature, T, see Figure 2 [*]. While it may be simpler to ensure constant load conditions, 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 friction are in fact comparable to those caused by w (recall Figure 2) and the problem is therefore challenging. In [5], a model-based approach was suggested for this problem based on the identification of a wear parameter from friction data collected under dedicated experiments. A shortcoming with this approach lies in the need of a detailed friction model, which requires a large amount of experiments to be found. The data-driven approach suggested here is considered as an alternative which requires little design effort since no model development is needed.

Since τ is affected by friction, and thus by wear, torque (current [†]) data are considered for the generation of fault indicators. The monitored data are collected from repetitive executions of a trajectory $\mho$. Relating to the notation introduced in Section 2, the deterministic unknown input of interest, $\mathbf{f}$, is the wear level w and the monitored data, $\mathbf{y}$, is τ which is affected by disturbances, $\mathbf{d}$, caused by load τ_l and temperature T and by measurement noise $\mathbf{v}$. A trajectory, $\mho$, is a known deterministic sequence used as a reference to the motion control, i.e. it relates to $\mathbf{r}$. In many applications, the same trajectory is executed over and over again, ensuring a repetitive behavior of the robot. For the results presented here, data collected from the execution of a trajectory $\mho$ based on a test-cycle are used. Torque data collected form this trajectory can be seen in Figure 3(a). Nominal torque data, τ^0, are achieved from the execution of $\mho$ when the gearbox is new and no significant wear is present.

Section 4.1 presents experimental results for the wear monitoring problem when the changes in disturbances are kept small. In Section 4.2, temperature disturbances are introduced in simulation studies and the use of weights described in Section 3.3 are used to illustrate how to reduce sensitivity to disturbances.

4.1 Experimental Wear Monitoring under Constant Disturbances

Accelerated wear tests were performed with ABB IRB 6620 industrial robots 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, the trajectory $\mho$

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

[†] In the application, a torque estimate based on a constant relationship between current measurements is used. This simplification is commonly used for control purposes since the current controller has much faster dynamics compared to the dynamics of the robot arm and the such estimate is therefore perceived as the control input signal.

was executed regularly a total of K times yielding a dataset $[\boldsymbol{\tau}_0, \cdots, \boldsymbol{\tau}^{K-1}]$. The data were collected from axis two of the robot which is equipped with a rotary vector gearbox type. The experiments were performed in a lab, in a setup to avoid temperature variations * and no load variations were present. It is thus considered that the disturbances had a repetitive behavior $\mathbf{d}_k = \mathbf{d}$ over all batches. The data batch taken from the start of the operation of a gearbox is considered to be fault-free and is labeled as nominal, $\boldsymbol{\tau}^0$. The quantities KL $\left(\widehat{p}^0 || \widehat{p}_k\right)$ are computed for $k = 1, \ldots, K-1$ and are used as fault indicators. Data collected from two accelerated wear tests are considered here. For an illustration of the wear behavior during the experiments, the friction levels in the joint were estimated using a dedicated experiment (see [6] for a description of such experiment) at each kth execution of $\mho$ and are shown as function of motor speed $\dot{\varphi}$.

For the first case, displayed in Figure 3, $K = 36$ batches of data are considered. From analyses of the friction levels in Figure 3(c), it is possible to note that wear only starts to considerably affect friction after $k = 25$. 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 distribution peaks. Notice further that the distributions are similar 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 of the fault indicator are estimated as $[\mu_0, \sigma_0] = [1.19\,10^{-2}, 5.09\,10^{-3}]$. The dashed line in Figure 3(d) shows the value of $\hbar = \mu_0 + 3\sigma_0$.

The second case, shown in Figure 4, illustrates the situation where a gearbox is replaced after a wear related failure takes place. A total of $K = 111$ data batches are collected during accelerated wear tests using the same test-cycle. A gearbox failure occurs at $k = 73$ when it is replaced by a new one. 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 distribution estimates 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 data densities 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, $\hbar = \mu_0 + 3\sigma_0$. The filled circle highlights the moment when the gearbox was replaced. As it can be seen in these studies, 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

Simulation studies were carried out to illustrate the ideas to reduce sensitivity to disturbances and noise presented in Section 3.3. 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. The simulation model is based on the two link manipulator with elastic gear transmission presented in the benchmark problem of [31]. With the objective of studying friction changes related to wear in a robot joint, the static friction model described in [5] is included in the simulation model. The friction model included 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.

4.2.1 Finding the weights $\mathbf{w}$

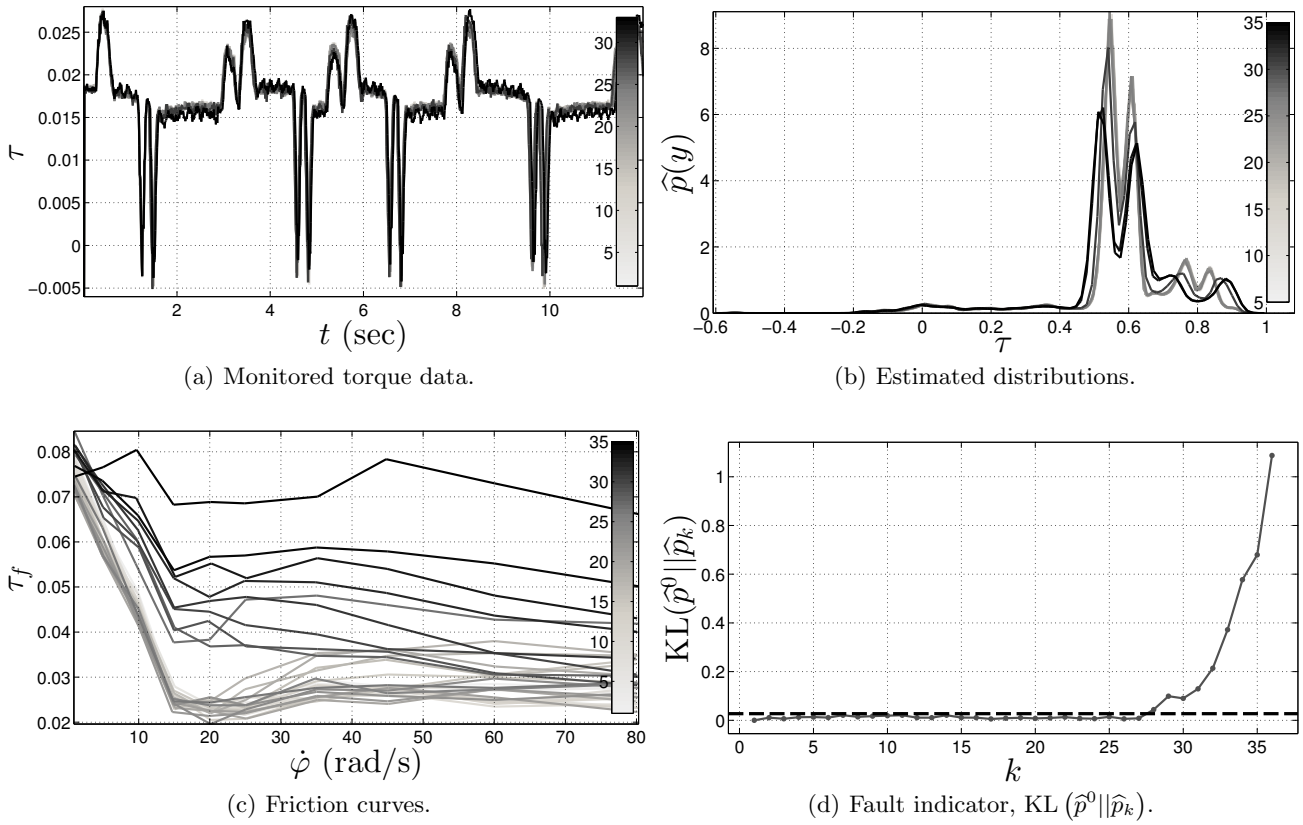
According to the procedures described in Section 3.3, a labeled dataset is needed in order to find the optimal weights. The dataset is achieved here based on simulations of the same test-cycle trajectory $\mho$ used in Section 4.1. Each labeled dataset $\mathbf{Y}^0_{K_0}$ and $\mathbf{Y}^f_{K_f}$ contain $K_0 = K_f = 100$ batches with torque data generated from the sets

$$\mathcal{Y}^0_r = \{\boldsymbol{\tau}_k : \mathsf{w}_k = 0,\, T_k \sim \mathrm{U}\left[\underline{\mathrm{T}}, \underline{\mathrm{T}} + \Delta_T\right],\, \mho_k = \mho\} \tag{16a}$$

$$\mathcal{Y}^f_r = \{\boldsymbol{\tau}_k : \mathsf{w}_k = \mathsf{w}_\mathrm{c},\, T_k \sim \mathrm{U}\left[\underline{\mathrm{T}}, \underline{\mathrm{T}} + \Delta_T\right],\, \mho_k = \mho\} \tag{16b}$$

respectively, where $\mathsf{w}_\mathrm{c} = 35$ is a wear level considered critical to generate an alarm (see [5] for details of the wear model). Here, T is considered random, with uniform distribution given by $\underline{\mathrm{T}} = 30°\mathrm{C}$ and $\Delta_T = 40°\mathrm{C}$. This assumption is carried out for analysis purposes and allows for great variations of temperature disturbances.

* The environment temperature was controlled and the experiments were only performed after the robot temperature was expected to be in equilibrium with the environment.



(a) Monitored torque data.

(b) Estimated distributions.

(c) Friction curves.

(d) Fault indicator, KL $\left(\widehat{p}^0||\widehat{p}_k\right)$.

Fig. 3. Monitoring of a wear fault in an industrial robot joint under accelerated wear tests and controlled load and temperature disturbances. The friction changes caused by wear were estimated during the experiments and are shown in (c) for a comparison. In the figures, the colormap relates to the friction values found. The monitored torque data from the execution of the trajectory $\mho$ 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 KL $\left(\widehat{p}^0||\widehat{p}_k\right)$ 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.

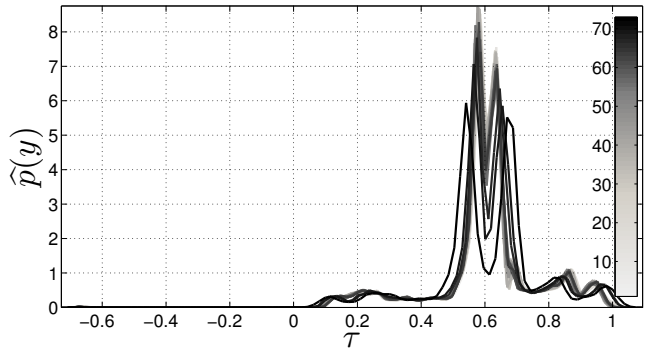
The optimal weights given in (14) and (15) depend on the average changes found in the data, $\mu_n^f - \mu_n^0$, and the total variability, $s_n^f + s_n^0$. These quantities are computed based on the labeled dataset and are displayed in Figure 5(a) as a function of the motor speed $\dot{\varphi}$. As 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. $\mu_n^f - \mu_n^0$ closest to zero and largest $s_n^f + s_n^0$. Figure 5(b) presents the ratio for such worst case estimates, which are considered as the optimal weights according to (15). The solid line in Figure 5(b) is a function approximation of the optimal weights given by

$$w(\dot{\varphi}) = \operatorname{sech}(\beta\dot{\varphi}) \tanh(\alpha\dot{\varphi}) \tag{17}$$

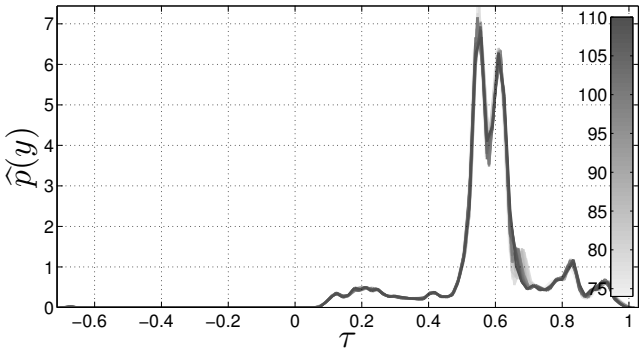
with $\alpha = 1.45\,10^{-2}$ and $\beta = 4.55\,10^{-2}$. The parametrization of the weight vector as a function of $\dot{\varphi}$ allows for a more general use of the optimal weights and the same weighting function can be used for other trajectories. Effectively, the optimal weighting function selects a speed region that is more relevant for wear monitoring, giving more emphasis to data in low to intermediate speed regions. A similar behavior was also found in [5] for the achievable quality of a wear estimate for different speeds under temperature disturbances.

4.2.2 Improvements in detection performance

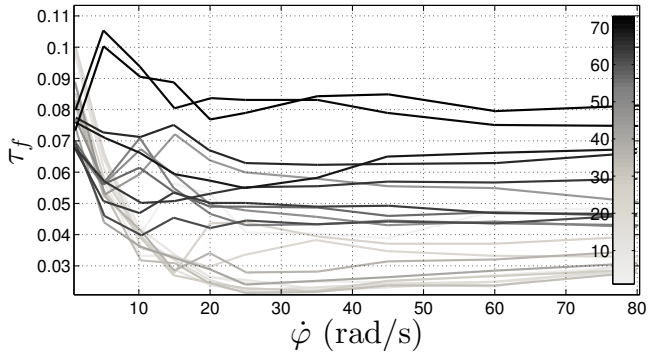
To illustrate the possible improvements achieved with the use of the weighting function, an abrupt change detection is considered. Given a nominal data batch $\boldsymbol{\tau}^0 \in \mathcal{Y}_r^0$, the detection problem is to decide whether a test batch $\boldsymbol{\tau}_k$ belongs to $\mathcal{Y}_r^0$ or $\mathcal{Y}_r^f$ based on the fault indicator KL $\left(\widehat{p}^0||\widehat{p}_k\right)$ and where the sets $\mathcal{Y}_r^0$ and $\mathcal{Y}_r^f$ are given by (16). This



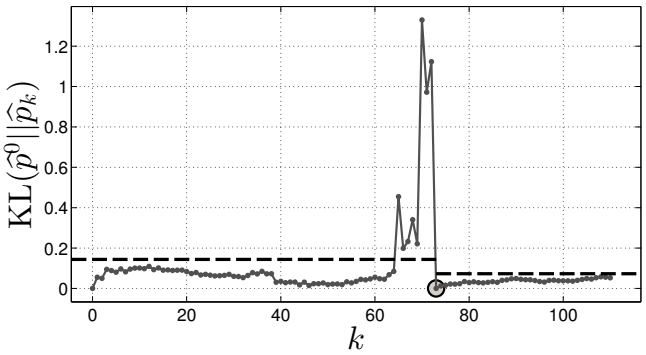
(a) Estimated distributions for $0 \leq k \leq 72$.



(b) Estimated distributions for $73 \leq k \leq 110$.

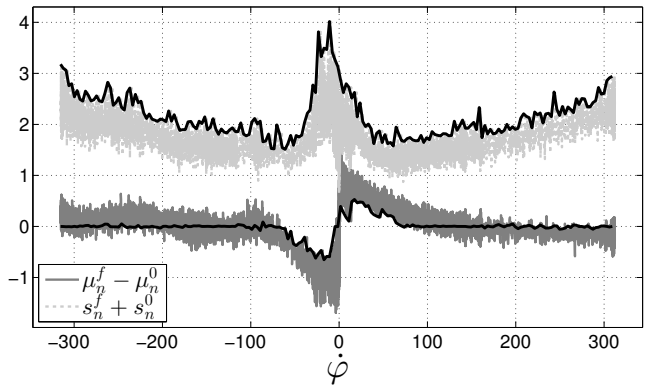


(c) Friction curves for $0 \leq k \leq 72$.

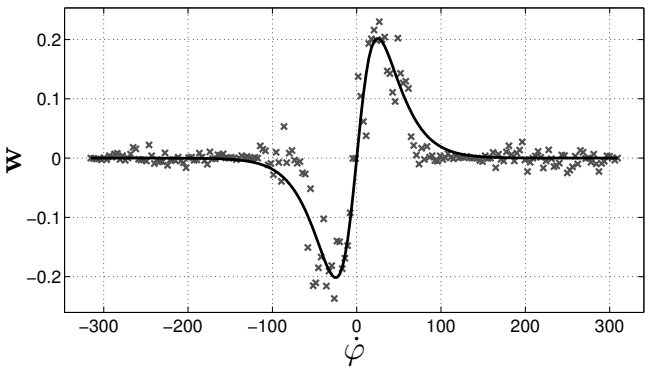


(d) Fault indicator, KL $\left(\widehat{p}^0||\widehat{p}_k\right)$.

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 $\mho$ 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 data distribution estimates 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 data distributions for the new gearbox are shown in Figure 4(b), where only small deviations are visible. The nominal data are assigned at the start of operation of the gearboxes at $k=0$ and at $k=73$. The resulting fault indicators are shown in Figure 4(d), with a clear response to the friction changes and a regular behavior when no fault is present; the circle in the figure highlights when the gearbox replacement took place and the dashed lines represent an upper limit for the nominal behavior of the fault indicator.

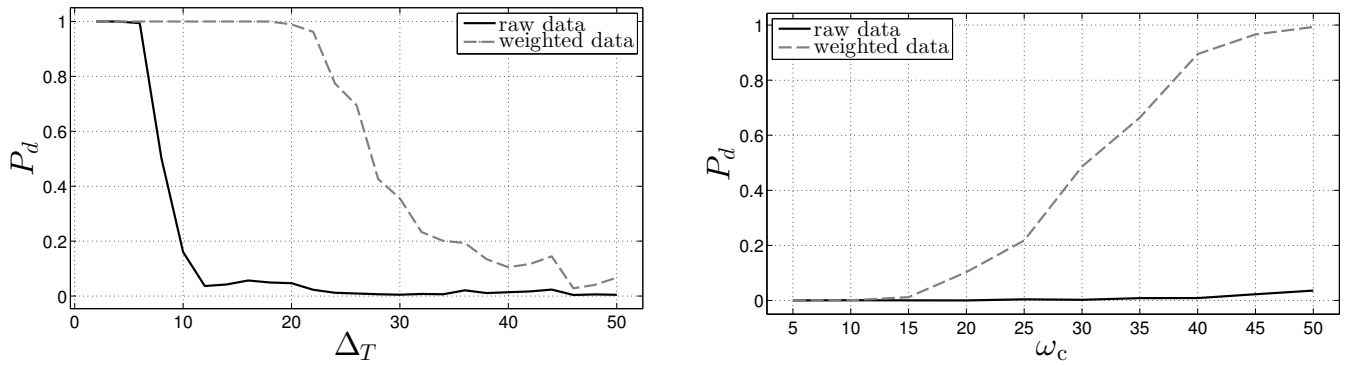


(a) Average effects.



(b) Optimal weights.

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) Temperature variations. (b) Fault size.

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

corresponds to the following hypotheses

$$\mathcal{H}^0 : \boldsymbol{\tau}_k \in \mathcal{Y}_r^0, \quad \mathcal{H}^1 : \boldsymbol{\tau}_k \in \mathcal{Y}_r^f \tag{18a}$$

where $\mathcal{H}^0$ indicates that no wear fault is present, with $\mathsf{w} = 0$, and $\mathcal{H}^1$ indicates presence of a wear fault of size w_c. For given values of w_c, $\underline{T}$ and Δ_T , the test-cycle trajectory $\mho$ is simulated to generate data according to (16). The probability densities of the fault indicator under each hypothesis is estimated based on 5000 Monte Carlo runs with and without the use of the weighting function. Based on the hypotheses densities', the probability of detection P_d is computed according to (4) for a threshold check when the probability of false alarm is $P_f = 0.01$.

Figure 6(a) presents the achieved P_d as a function of Δ_T for the fixed $\mathsf{w}_c = 35$ and $\underline{T} = 30°C$ with and without the use of the weighting function. Notice that the use of the weighting function considerably improves P_d under temperature variations, but for too large Δ_T it becomes difficult to distinguish the effects. A similar study is performed to illustrate how w_c affects the performance. Figure 6(b) presents P_d as a function of w_c for the fixed $\Delta_T = 25°C$ and $\underline{T} = 30°C$. The improvements achieved using the weighted data are clear.

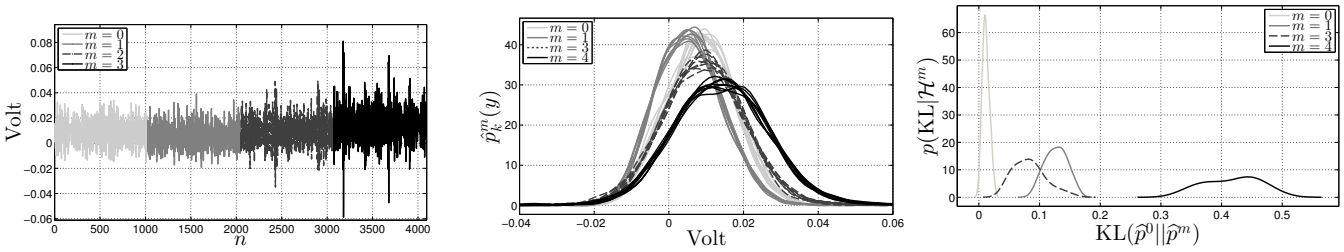
5 Gearbox Monitoring based on Vibration Data

In this application, vibration data collected from the gearbox test rig described in [19,23] are considered. The test rig is composed of a motor coupled to a gearbox with three shafts and four spur gears. It is possible to study the effects of different types of gear faults in the rig by replacing healthy gears with damaged ones. Four different fault types (modes) are considered:

- $m = 0$: healthy gears are used,
- $m = 1$: a gear at the input shaft is damaged,
- $m = 2$: a gear at the output shaft is damaged,
- $m = 3$: a gear at the input and a gear at the output shafts are damaged.

Fault detection approaches for this problem have been proposed in [19] with the use of Hilbert and wavelet packet transform and in [23] with a combination of Wavelet transform and time domain averaging. In these approaches, the data are transformed to a time-frequency domain where different faults can be distinguished based on inspection of the transformed data. The distribution domain approach proposed here is considered as an alternative for fault diagnosis.

Data collected from an accelerometer placed close to the output shaft are considered for the analyses. All data collection is performed under constant load and speed conditions in all settings, ensuring a repetitive behavior. For each different condition, 8×1024 samples are available with data sampled at 2.56KHz. The datasets are divided in $K = 8$ batches with $N = 1024$ samples to form $\mathbf{Y}_K^m$ for each mode m. Figure 7(a) shows the fist data batch $\mathbf{y}_1^m$ for each mode m, notice that it is difficult to distinguish differences in the data sequences. In Figure 7(b), the



(a) Data sequences. (b) Density estimates. (c) Fault indicator behavior.

Fig. 7. Diagnostics of a gearbox based on vibration data. A total of four different modes m are possible. An example of data sequence from each mode is shown in (a). The data behavior in the distribution domain is shown in (b). The statistical behavior of the fault indicator given by $\mathrm{KL}\left(\widehat{p}^0||\widehat{p}_k\right)$ is shown for $\mathbf{y}_k \in \mathcal{H}^m$ in (c). Notice that, despite the densities in (b) being similar, the fault indicator clearly indicates the presence of a change in (c).

density estimates $\widehat{p}_k^m(y)$ for every batch in every mode are displayed. Notice the smaller variability of the distribution estimates within each mode compared to the variability found between modes.

To evaluate the fault indicator performance to change detection, the distance $\mathrm{KL}\left(\widehat{p}^0||\widehat{p}_k\right)$ is computed for every possible pair such that $\mathbf{y}^0 \in \mathbf{Y}_K^0$ and $\mathbf{y}_k \in \mathbf{Y}_K^m$ for $m \in [0,1,2,3]$. Since the distances are computed in pairs, this gives 28 samples from the nominal set $\mathcal{D}^{0,0}$ ($K=8$ combined two by two) and $K^2=64$ samples from the sets $\mathcal{D}^{0,m}$ for $m \in [1,2,3]$. The hypotheses densities $p(\mathrm{KL}|\mathcal{H}^m)$ are estimated based on these samples and are shown in Figure 7(c). Notice the clear separation between the null hypothesis density, $p(\mathrm{KL}|\mathcal{H}^0)$, from the alternatives. For $P_f=0.01$, P_d is computed based in (4) for a threshold check when deciding between $\mathcal{H}^0$ and $\mathcal{H}^m$ for $m \in [1,2,3]$; the achieved values are $P_d=[1,\ 0.991,\ 1]$. To illustrate the approach to isolation as given in (5), the first data batch in each mode, $\mathbf{y}_1^m$, is considered available. The isolation criterion given by $m^*=\arg\min_m \mathrm{KL}\left(\widehat{p}_1^m||\widehat{p}_k\right)$ chooses the correct mode for all $(K-1)\times 4$ remaining data batches.

6 Conclusions and Future Work

The suggested framework considers the monitoring of changes in the distribution of data batches. Because no prior knowledge is assumed about the data behavior, nonparametric kernel density estimates are used, which give great flexibility, are simple to implement and have an inherent smoothing behavior. The validity of the framework and methods were illustrated with promising results on real case studies and simulations for gearbox monitoring in robotics and rotating machines. An important advantage of the framework presented is that no model or expert knowledge of the system are required. Furthermore, it gives an alternative for systems where faults affect the data amplitude but where stationary, periodic or linear behaviors are difficult or not possible, as in the robotics application presented.

As future work, it would be interesting to further investigate the methods applicability to other domains and faults. In this direction, the derivation of conditions for diagnosability are relevant. The determination of a change in data collected under repetitive conditions simplifies the diagnostics problem considerably. In general however, it might not possible to ensure the same repetitive behavior of the system. This is the case, for example, in the industrial robotics application where trajectories are normally only repeated trough a certain period, depending on the manufacturing plan. Thus, it would relevant to study approaches to handle systems with a varying repetitive behavior. The effects of different kernel functions for the KDE, choice of bandwidth parameter and use of different distances between densities are also important.

Acknowledgments

This work was supported by ABB and the Vinnova Industry Excellence Center LINK-SIC at Linköping University. The authors would also like to thank Professor Sirish L. Shah from the University of Alberta, Canada, for sharing the data used in the studies of Section 5.

References

[1] R. Aguilar-Lopez and R. Martinez-Guerra. Robust state estimation for repetitive operating mode process: Application to sequencing batch reactors. *Chemical Engineering Journal*, 126(2):155–161, 2007.

[2] I. A. Alvarado, R. Findeisen, P. Kuhl, F. Allgower, and D. Limón. State estimation for repetitive processes using iteratively improving moving horizon observers. In *Decision and Control, 2005 and 2005 European Control Conference. CDC-ECC'05. 44th IEEE Conference on*, pages 7756–7761. IEEE, 2005.

[3] M. Basseville. Distance measures for signal processing and pattern recognition. *Signal Process.*, 18(4):349–369, Dec. 1989.

[4] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, New York, USA, 1st edition, 2007.

[5] A. C. Bittencourt and P. Axelsson. Modeling and experiment design for identification of wear in a robot joint under load and temperature uncertainties based on friction data. *IEEE/ASME Transactions on Mechatronics*, PP(99):1–13, 2013.

[6] A. C. Bittencourt and S. 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.

[7] A. C. Bittencourt, K. Saarinen, and S. Sander-Tavallaey. A data-driven method for monitoring systems that operate repetitively - applications to wear monitoring in an industrial robot joint. In *Proc. of the 8th IFAC SAFEPROCESS*, volume 8, Mexico City, Mexico, 2012.

[8] A. W. Bowman and A. Azzalini. *Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations (Oxford Statistical Science Series)*. Oxford University Press, USA, Nov. 1997.

[9] D. Brambilla, L. 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.

[10] 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.

[11] T. Cacoullos. Estimation of a multivariate density. *Annals of the Institute of Statistical Mathematics*, 18:179–189, 1966.

[12] A. De Luca and R. Mattone. Actuator failure detection and isolation using generalized momenta. In *Proc. of the 2003 IEEE International Conference on Robotics and Automation (ICRA)*, volume 1, pages 634 – 639 vol.1, Taipei, Taiwan, sept. 2003.

[13] C. W. de Silva. *Vibration Monitoring, Testing and Instrumentation*. CRC Press, Apr. 2007.

[14] S. X. Ding, S. Yin, Y. Wang, Y. Wang, Y. Yang, and B. Ni. Data-driven design of observers and its applications. In *Proceedings of the 18th IFAC World Congress*, 2011.

[15] W. E. Dixon, I. D. Walker, D. M. Dawson, and J. P. Hartranft. Fault detection for robot manipulators with parametric uncertainty: A prediction-error-based approach. *IEEE Transactions on Robotics and Automation*, 16(6):3628–3634, 2000.

[16] J. Dong, M. Verhaegen, and F. Gustafsson. Robust fault detection with statistical uncertainty in identified parameters. *IEEE Transactions on Signal Processing*, 60(10):5064–5076, 2012.

[17] R. Durrett. *Probability: Theory and examples*. Cambridge University Press, August 2010.

[18] I. Eski, S. Erkaya, S. Savas, and S. Yildirim. Fault detection on robot manipulators using artificial neural networks. *Robotics and Computer-Integrated Manufacturing*, 27(1):115 – 123, Jul 2011.

[19] X. Fan and M. J. Zuo. Gearbox fault detection using hilbert and wavelet packet transform. *Mechanical Systems and Signal Processing*, 20(4):966 – 982, 2006.

[20] V. Filaretov, M. Vukobratovic, and A. Zhirabok. Observer-based fault diagnosis in manipulation robots. *Mechatronics*, 9(8):929 – 939, 1999.

[21] 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.

[22] S. C. Guo, M. H. Yang, Z. R. Xing, Y. Li, and J. Q. Qiu. Actuator fault detection and isolation for robot manipulators with the adaptive observer. *Advanced Materials Research*, 482 - 484(8):529–532, 2012.

[23] E. B. Halim, M. S. Choudhury, S. L. Shah, and M. J. Zuo. Time domain averaging across all scales: A novel method for detection of gearbox faults. *Mechanical Systems and Signal Processing*, 22(2):261 – 278, 2008.

[24] R. Isermann. *Fault-Diagnosis Systems - An Introduction from Fault Detection to Fault Tolerance*. Springer, 2006.

[25] R. Isermann. *Fault-diagnosis applications - model-based condition monitoring: actuators, drives, machinery, plants, sensors, and fault-tolerant systems*. Springer, New York, 2011.

[26] M. Jones and D. Henderson. Maximum likelihood kernel density estimation: On the potential of convolution sieves. *Computational Statistics & Data Analysis*, 53(10):3726 – 3733, 2009.

[27] X. Li and K. Zhou. A time domain approach to robust fault detection of linear time-varying systems. *Automatica*, 45(1):94 – 102, 2009.

[28] L. Ljung. *System Identification: Theory for the User (2nd Edition)*. Prentice Hall PTR, December 1998.

[29] L. Marton and F. van der Linden. Temperature dependent friction estimation: Application to lubricant health monitoring. *Mechatronics*, 22(8):1078 – 1084, 2012.

[30] M. McIntyre, W. Dixon, D. Dawson, and I. Walker. Fault identification for robot manipulators. *IEEE Transactions on Robotics*, 21(5):1028–1034, Oct. 2005.

[31] S. Moberg, J. Öhr, and S. Gunnarsson. A benchmark problem for robust control of a multivariable nonlinear flexible manipulator. In *Proc. of the 17th IFAC World Congress*, Mar 2008.

[32] E. Olsson, P. Funk, and N. Xiong. Fault diagnosis in industry using sensor readings and case-based reasoning. *Journal of Intelligent & Fuzzy Systems*, Vol. 15:10, December 2004.

[33] E. Parzen. On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33(3):pp. 1065–1076, 1962.

[34] M. D. Reid and R. C. Williamson. Information, divergence and risk for binary experiments. *Journal of Machine Learning Research*, 12:731 – 817, 2011.

[35] E. Rogers, K. Galkowski, and D. H. Owens. *Control systems theory and applications for linear repetitive processes*, volume 349. Springer, 2007.

[36] E. Rogers, K. Galkowski, and D. H. Owens. Two decades of research on linear repetitive processes part i: Theory. In *Multidimensional Systems (nDS), 2013. Proceedings of the 8th International Workshop on*, pages 1–6, 2013.

[37] E. Rogers, K. Galkowski, W. Paszke, and D. H. Owens. Two decades of research on linear repetitive processes part ii: Applications. In *Multidimensional Systems (nDS), 2013. Proceedings of the 8th International Workshop on*, pages 1–6, 2013.

[38] S. Sander-Tavallaey and K. Saarinen. Backlash identification in transmission unit. In *Proc. of the 2009 IEEE Control Applications & Intelligent Control*, pages 1325 –1331, jul 2009.

[39] B. Sulikowski, K. Gałkowski, E. Rogers, and D. H. Owens. Output feedback control of discrete linear repetitive processes. *Automatica*, 40(12):2167–2173, 2004.

[40] J. I. Taylor. *The Vibration Analysis Handbook*. Vibration Consultants, Feb. 1994.

[41] H. L. Van Trees and K. L. Bell. *Detection, Estimation and Modulation Theory, Part I*. Wiley, New York, 2nd edition, 2013.

[42] Y. Wang, G. Ma, S. X. Ding, and C. Li. Subspace aided data-driven design of robust fault detection and isolation systems. *Automatica*, 47(11):2474–2480, 2011.

[43] L. Wu, X. Su, and P. Shi. Mixed H_2/H_∞ approach to fault detection of discrete linear repetitive processes. *Journal of the Franklin Institute*, 348(2):393–414, 2011.

[44] S. Yin, S. X. Ding, A. H. Abandan Sari, and H. Hao. Data-driven monitoring for stochastic systems and its application on batch process. *International Journal of Systems Science*, 44(7):1366–1376, 2013.