A Data-driven Method for Monitoring Systems that Operate Repetitively – Applications to Robust Wear Monitoring in an Industrial Robot Joint [1]

André Carvalho Bittencourt * **Kari Saarinen** **
Shiva Sander-Tavallaey **

* *Division of Automatic Control, Department of Electrical Engineering, Linköping University, Linköping, Sweden,* andrecb@isy.liu.se
** *ABB Corporate Research, Västerås, Sweden,*
{kari.saarinen,shiva.sander-tavallaey}@se.abb.com

Abstract: This paper presents a method for condition monitoring of systems that operate in a repetitive manner. A data driven method is proposed that considers changes in the distribution of data samples obtained from multiple executions of one or several tasks. This is made possible with the use of kernel density estimators and the Kullback-Leibler distance measure between distributions. To increase robustness to unknown disturbances and sensitivity to faults, the use of a weighting function is suggested which can considerably improve detection performance. The method is very simple to implement, it does not require knowledge about the monitored system and can be used without process interruption, in a batch manner. The method is illustrated with applications to robust wear monitoring in a robot joint. Interesting properties of the application are presented through a real case study and simulations. The achieved results show that robust wear monitoring in industrial robot joints is made possible with the proposed method.

Keywords: FDI for robust nonlinear systems, Data-driven methods, Industrial robots, Wear monitoring, 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 down time, Khan and Abbasi (1999); Rao (1998). 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. Its impacts can however be considerably reduced if good maintenance practices are performed.

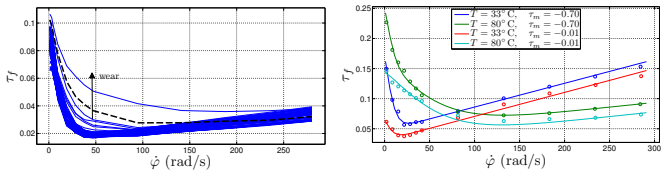
In the manufacturing industry, including industrial robots, preventive scheduled maintenance is a common approach used to improve equipment SRAM. This setup delivers high availability, reducing operational costs (e.g. small downtimes) with the drawback of high maintenance costs since unnecessary maintenance 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, preferably, this should be done automatically.

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

This work discusses the use of a data driven method for condition monitoring of machines that operate in a repetitive manner, e.g. commonly found in the manufacturing industry and in automation. The method was developed with the interest focused on diagnosis of industrial robots, where a repetitive operation is almost a requirement in most of its applications.

In robotics, condition monitoring and fault detection methods are mainly considered in the time-domain. Due to the complex dynamics of an industrial robot, the use of nonlinear observers for fault detection is a typical approach (Caccavale and Villani (2003)). Since observers are sensitive to model uncertainties and disturbances, some methods attempt to diminish these effects. In Brambilla et al. (2008) and De Luca and Mattone (2004), nonlinear observers are used together with adaptive schemes while in Caccavale et al. (2009), the authors mix the use of nonlinear observers with support vector machines. The problem has also been approached by the use of neural networks as presented in Vemuri and Polycarpou (2004) and in Eski et al. (2010), where vibration data are used for diagnosis. Parameter estimation is a natural approach because it can use the physical interpretation of the system, e.g. Freyermuth (1991). No reference was found for condition monitoring methods of industrial robots that make a direct use of the repetitive behavior of the system.

In the literature, actuator failures are typically considered as abrupt changes in the output torque signals. These fault



(a) Wear effects. (b) Disturbances effects.

Fig. 1. Static friction in a robot joint. As seen in (a), the wear causes an increase of the friction in the joint. The effects of disturbances caused by load τ_m and temperature T are however very significant as illustrated in (b). These effects were measured in similar gearboxes and are presented in directly comparable scales.

models can relate to several types of failures such as a motor malfunction, power supply drop or a wire cut. Such failures are however difficult to predict and therefore might cause damages even if detected. One example of a failure type that is not abrupt is a failure that follows after a gradual wear of a component. This type of fault develops with time/usage and might be detected at an early stage, allowing for CBM. Even if such wear is a long process of several years, it is possible to study the phenomena in accelerated wear tests by running the robot at much higher stress levels than allowed. In this work, data resulting from accelerated wear tests performed in a lab are considered for the proposed methods.

It is well known that friction changes can follow as a result of wear processes in mechanical systems, see e.g. Kato (2000). In Bittencourt et al. (2011), this dependency in an industrial robot joint is studied and modeled. A possible diagnose solution is thus to monitor the friction in the joints. The problem is however challenging since friction depends on other phenomena such as load and temperature (Bittencourt et al. (2010)), see Fig. 1. In Bittencourt et al. (2011), a method is proposed for wear identification in a robot joint based on a test cycle and a known friction model. The study shows that it is possible to achieve robust wear estimates and presents basic limitations of identification methods for wear monitoring. Its practical use is however limited since it requires a test cycle and assumes a known friction model which can describe the effects of speed, load, temperature and wear.

In this paper, a quantity suitable for condition monitoring of systems that operate in a repetitive manner is proposed. The quantity relates to the differences found in the distributions of data taken under recurring conditions, e.g. from the execution of the same task. The problem of robust wear monitoring in a robot joint is used to illustrate the method throughout the paper with an experimental case study and simulations. The basic framework is presented in Sec. 2 with an experimental study of wear monitoring in a robot joint. In Sec. 3, ideas are presented and illustrated through examples to handle the cases where the repetitive behavior of the system changes, e.g. when several tasks are executed multiple times. Ideas used to reduce the sensitivity to disturbances are presented in Sec. 4 with detailed simulation studies of the effects of temperature for the robotics application. Finally, conclusions and possible extensions are given in Sec. 5.

2. MONITORING OF SYSTEMS THAT OPERATE IN A REPETITIVE MANNER

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

$$\mathbf{Y}^M = [\mathbf{y}^0, \cdots, \mathbf{y}^j, \cdots, \mathbf{y}^{M-1}],$$

where $\mathbf{y}^j = [y_1^j, \cdots, y_i^j, \cdots, y_N^j]^T$ denotes the N dimensional vector of measurements, which is sequentially repeated M times.

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

$$\mathbf{y}^j = h(\mathbf{z}^j, \mathbf{v}^j), \tag{1}$$

where $h(\cdot)$ is a general function. Let the set of deterministic inputs $\mathbf{Z}^M$ be categorized in three distinct groups, $\mathbf{U}^M$, $\mathbf{D}^M$ and $\mathbf{F}^M$. The sequences $\mathbf{f}^j$ are unknown and of interest (a fault [2]), while $\mathbf{u}^j$ and $\mathbf{d}^j$ are respectively known and unknown (e.g. inputs and disturbances). For the purpose of monitoring $\mathbf{y}^j$ to detect changes in $\mathbf{f}^j$, the following assumptions are taken:

Assumption 2.1. (Faults are observable). Changes on $\mathbf{f}^j$ affect the measured data $\mathbf{y}^j$.

Assumption 2.2. (Regularity of $\mathbf{y}^j$ if no fault). It is considered that the monitored data $\mathbf{y}^j$ change only slightly along j, unless in the presence of a nonzero fault $\mathbf{f}^j$.

Assumption 2.3. (Regularity of $\mathbf{d}^j$). The deterministic disturbance $\mathbf{d}^j$ is such that it changes only slightly along j. Notice that this follows partly from Assumption 2.2.

Assumption 2.4. (Nominal data are available). At $j = 0$, $\mathbf{f}^0 = 0$ and the sequence $\mathbf{y}^0$ is always available.

Notice that if $\mathbf{u}^j$ satisfies the Assumptions 2.1, 2.2 and 2.4, it can be included in the monitored sequence $\mathbf{y}^j$.

The rationale is then to simply compare the nominal data $\mathbf{y}^0$ (always available from Assumption 2.4) against the remaining sequences $\mathbf{y}^j$. While Assumption 2.1 is necessary, Assumption 2.2 ensures that two given sequences $\mathbf{y}^k$, $\mathbf{y}^l$ are comparable and might differ significantly only if there is a fault. Two basic questions arise which are answered in the next subsections

- How to characterize $\mathbf{y}^j$?
- How to compare two sequences $\mathbf{y}^k$, $\mathbf{y}^l$ for monitoring?

Furthermore, Assumptions 2.2, 2.3 and 2.4 are too restrictive in many applications. In Sections 3 and 4, alternatives are presented in order to relax these assumptions.

For an industrial robot executing a regular task under wear changes, the basic framework applies as follows. An industrial robot can be described as a multi body dynamic mechanism by

$$\tau = M(\varphi)\ddot{\varphi} + C(\varphi, \dot{\varphi}) + D\dot{\varphi} + \tau_g(\varphi) + \tau_s(\varphi) + \tau_f(\dot{\varphi}, \tau_m, T, \mathsf{w}), \tag{2}$$

[2] The terminology adopted in this paper defines a fault as a deviation of at least one characteristic property of the system from the acceptable / usual / nominal condition.

where τ is the applied torque, φ is the vector of angular positions (at motor and arm sides), $M(\varphi)$ is the inertia matrix, $C(\varphi, \dot{\varphi})$ relates to speed dependent terms (e.g. Coriolis and centrifugal), D is a damping matrix, $\tau_g(\varphi)$ are the gravity-induced torques, $\tau_s(\varphi)$ is a nonlinear stiffness. The function $\tau_f(\cdot)$ contains the joint friction components and is dependent on joint speed $\dot{\varphi}$, the manipulated load τ_m, the temperature inside the joint T and the wear levels w.

Using the introduced notation, the deterministic input of interest, $\mathbf{f}$, is the wear level w, which is considered to be zero when the robot is new and to increase with time/usage. In typical industrial robots applications, angular position at the motor side and motor current are measured quantities. Angular position measurements φ are achieved with high resolution encoders and can be differentiated to achieve motor angular speed $\dot{\varphi}$. The current is the control input to the motor and it is common to assume that the relationship between current and applied torque τ is given by a constant [3]. Since from (2) it is clear that τ is affected directly by w (satisfying Assumption 2.1), only τ is considered of interest and included in $\mathbf{y}$. The remaining variables, φ and its derivatives, load torque τ_m and joint temperature T are considered as disturbances and included in $\mathbf{d}$.

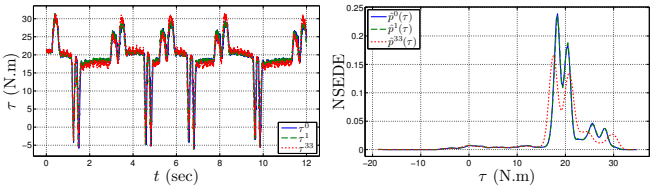
Notice that the effects of φ, its derivatives, and τ_m are defined by the task, $\mho$, executed by the manipulator. If the monitored sequences $\mathbf{y}^j$ are achieved from the operation of the same task $\mho$, these disturbances satisfy Assumption 2.3, notice that they considerably vary along i. If this behavior is also valid for T, then $\mathbf{y}^j$ satisfies Assumption 2.2 and the framework is valid. Joint temperature is however the result of complicated losses mechanisms in the joint and heat exchanges with the environment and might not satisfy the assumption. The effects of T are in fact comparable to those caused by w, recall Fig. 1. The problem of robust monitoring of w is therefore challenging.

2.1 Characterizing the Measured Data – NSEDE

There are several ways to characterize a sequence $\mathbf{y}^j$. It could be represented by a single number, such as its mean, peak, range, etc. Summarizing the whole sequence into single quantities might however hide many of the signal's features. A second alternative would be to simply store the whole sequence and try to monitor the difference $\mathbf{y}^0 - \mathbf{y}^j$ but this requires that the sequences are synchronized, which is a limitation in many applications. Sometimes, looking at the data spectra are helpful, but this type of analysis requires the data to be ordered.

The alternative pursued in this work is to consider the distribution of $\mathbf{y}^j$, which does not require ordering or synchronization and reveals many of the signal's features. Because the mechanisms that generated the data are considered unknown, the use of a nonparametric estimate of the distribution of $\mathbf{y}^j$ is a suitable alternative. A nonparametric estimate of the distribution $p(\cdot)$ of $\mathbf{y}^j$ can be achieved with the use of kernel density estimators (Bishop (2007)),

[3] This is due to the fast dynamics of the current control loop compared to the arm.



(a) Torque signal. (b) Estimated NSEDEs.

Fig. 2. (a), torque signals at a joint under accelerated wear tests and their NSEDEs, (b), related to the execution of a task $\mho$. The sequences $\boldsymbol{\tau}^0$ and $\boldsymbol{\tau}^1$ are fault free, $\boldsymbol{\tau}^{M-1}$ was achieved with increased wear levels in the gearbox. A Gaussian kernel was used for computing the NSEDEs.

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

where $k_h(\cdot)$ is a kernel function, satisfying $k_h(\cdot) \geq 0$ and that integrates to 1 over $\mathbb{R}$. The bandwidth $h > 0$ is a smoothing parameter and y includes the domain of $\mathbf{Y}^M$. It is typical to choose kernels 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 Bowman and Azzalini (1997) for more details on kernel density estimators and criteria/methods for choosing h. From the definition, it follows that $\int \hat{p}(y)\,\mathrm{d}y = 1$, that is, the distribution is normalized to 1. The quantity $\hat{p}^j(y)$ is a nonparametric smooth empirical distribution estimate (NSEDE) of $\mathbf{y}^j$.

Example 2.1. (NSEDEs of Experimental Data from a Robot Executing a Regular Path under Wear Changes:) Accelerated wear tests were performed in a robot joint with the objective of studying the wear effects. During these experiments, the joint temperature T was kept constant to satisfy Assumption 2.3. Throughout the tests, a task $\mho$ was executed regularly a total of $M = 33$ times yielding a data set $[\boldsymbol{\tau}^0, \cdots, \boldsymbol{\tau}^{M-1}]$. The tests were executed until the wear levels were considered significant, so that maintenance should be performed. For an illustration, the torque sequences $\boldsymbol{\tau}^0$, $\boldsymbol{\tau}^1$ and $\boldsymbol{\tau}^{M-1}$ are shown in Fig. 2(a), together with their estimated NSEDEs, in Fig. 2(b). The sequences $\boldsymbol{\tau}^0$ and $\boldsymbol{\tau}^1$ are considered to be fault free while $\boldsymbol{\tau}^{M-1}$ was achieved with increased wear levels. Notice how the NSEDEs are similar for the fault free data and how they considerably differ from $\boldsymbol{\tau}^{M-1}$.

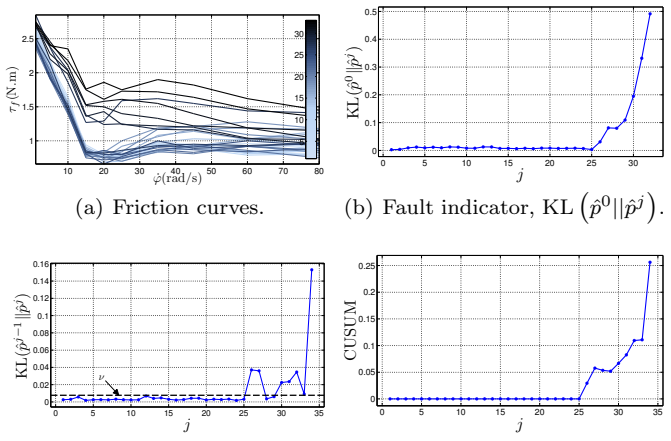
From Ex. 2.1 and Fig. 2, it is possible to see that Assumptions 2.2 and 2.1 are valid and that it might be possible to monitor the changes in the NSEDEs to infer about a fault. In the next subsection, a distance measure is defined between NSEDEs.

2.2 Fault Indicator – Kullback-Leibler distance

In statistics and information theory, the Kullback-Leibler divergence (KLD) is 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

$$D_{\mathrm{KL}}\left(p||q\right) = -\int_{-\infty}^{\infty} p(y) \log \frac{q(y)}{p(y)}\,\mathrm{d}y. \tag{4}$$

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 not a



(a) Friction curves.

(b) Fault indicator, KL $\left(\hat{p}^0||\hat{p}^j\right)$.

(c) Increments KL $\left(\hat{p}^{j-1}||\hat{p}^j\right)$.

(d) Fault indicator, CUSUM of KL $\left(\hat{p}^{j-1}||\hat{p}^j\right)$.

Fig. 3. Monitoring of a wear fault in an industrial robot joint under accelerated wear tests. The friction changes caused by wear for the fault are shown in (a) for a comparison, the colormap relates to j. The fault indicator using KL $\left(\hat{p}^0||\hat{p}^j\right)$ from Ex. 2.2 is shown in (b). The lower row presents the resulting quantities when monitoring the accumulated changes from Ex. 3.1. The incremental changes and the drift parameter ν are shown in (c). The fault indicator from the CUSUM filtered increments is displayed in (d), notice its robustness compared to (b).

metric, since in general it is not symmetric, $D_{\mathrm{KL}}\left(p||q\right) \neq D_{\mathrm{KL}}\left(q||p\right)$. The quantity

$$\mathrm{KL}\left(p||q\right) \triangleq D_{\mathrm{KL}}\left(p||q\right) + D_{\mathrm{KL}}\left(q||p\right), \tag{5}$$

known as the Kullback-Leibler distance is however symmetric. For an up to date review of divergences, see Reid and Williamson (2011).

Although the KL distance is defined for probability functions, it can also be used for NSEDEs since they are normalized to 1. An answer to the second question outlined in the beginning of this section can therefore be given with the use of the KL measure defined in (5). From Assumption 2.4, fault free data are always available, so that $\mathbf{y}^0$ is known and $\hat{p}^0$ can be evaluated. The quantities KL $\left(\hat{p}^0||\hat{p}^j\right)$ can therefore be used as a fault indicator.

Example 2.2. (Application of the KL $\left(\hat{p}^0||\hat{p}^j\right)$ for Experimental Wear Monitoring in a Robot Joint:)
The same sequence $[\boldsymbol{\tau}^0, \cdots, \boldsymbol{\tau}^{M-1}]$ used in Ex. 2.1 is considered here. First, their respective NSEDEs are computed, resulting in $[\hat{p}^0, \cdots, \hat{p}^{M-1}]$. Considering $\boldsymbol{\tau}^0$ to be fault free, the quantities KL $\left(\hat{p}^0||\hat{p}^j\right)$ are computed for $j = 1, \ldots, M-1$. As shown in Fig. 3(b), these quantities show a clear response to how the wear level increases and can therefore be used as a wear indicator (recall that temperature was kept constant during these experiments). For an illustration of the wear behavior during the experiments, the friction curves in the joint were estimated using a dedicated experiment (see Bittencourt et al. (2010)) at each jth execution of $\mho$ and are shown in Fig. 3(a).

The above example illustrates how the basic framework can be successfully used to monitor systems that operate in a repetitive manner. The regularity requirements described in Assumptions 2.2 and 2.3 are however limiting in many practical applications. The next sections discuss approaches to relax these assumptions.

3. MONITORING THE ACCUMULATED CHANGES

Since KL $\left(\hat{p}^{k-1}||\hat{p}^k\right)$ measures the difference between consecutive sequences, the sum of these increments over $1, \ldots, j$ gives the accumulated changes up to j, which is related to a fault and can therefore be used for monitoring, without requiring the assignment of nominal data.

Because of the noise components $\mathbf{v}$, the increments KL $\left(\hat{p}^{j-1}||\hat{p}^j\right)$ will also have a random behavior when there is no fault. The simple summation of the increments will therefore behave like a random walk and drift away. An alternative is to use the cumulative sum (CUSUM) algorithm (Gustafsson (2000)), defined as

Algorithm 1 CUSUM

$$g^j = g^{j-1} + s^j - \nu \tag{6}$$
$$g^j = 0 \text{ if } g^j < 0. \tag{7}$$

The test statistic g^j adds up the signal to be monitored s^j, which in the context presented here is $s^j = \mathrm{KL}\left(\hat{p}^{j-1}||\hat{p}^j\right)$. To avoid positive drifts, the drift parameter ν is subtracted from the update rule (6), if on the other hand g^j becomes negative, g^j is reset, avoiding negative drifts. The resulting quantity, g^j is suitable for condition monitoring and *does not require assignment of a nominal data*, that is, Assumption 2.4 is relaxed. The drift parameter can be chosen as

$$\nu = \kappa\sigma + \mu, \tag{8}$$

where μ and σ are the mean and the standard deviation of the increments KL $\left(\hat{p}^{j-1}||\hat{p}^j\right)$ under no fault and κ is a positive constant.

Example 3.1. (Application of the CUSUM to KL $\left(\hat{p}^{j-1}||\hat{p}^j\right)$ for Experimental Wear Monitoring in a Robot Joint:)
The real failure case in Ex. 2.2 is considered again. Instead of using KL $\left(\hat{p}^0||\hat{p}^j\right)$ as a fault indicator, the increments KL $\left(\hat{p}^{j-1}||\hat{p}^j\right)$ are computed and the CUSUM filter is used. The drift parameter is chosen as in (8), with $\kappa = 3$ and σ, μ estimated from the first 5 sequences. The resulting quantities are shown Figs. 3(c) and 3(d), with a clear response to the wear increases.

3.1 Monitoring Irregular Data

Let $\mho^j$ denote the conditions under which a sequence $\mathbf{y}^j$ was generated. Assumption 2.2 requires the whole sequence $\mathbf{Y}^M$ to have been generated under the same $\mho$, so that they are comparable. The alternative solution of monitoring the accumulated consecutive increments KL $\left(\hat{p}^{j-1}||\hat{p}^j\right)$ requires, in principle, that only $\mho^{j-1}$ and $\mho^j$ are the same, relaxing Assumption 2.2.

Since the behavior of the increments might differ depending on $\mho$, special care should be taken when monitoring their accumulated changes. If the CUSUM algorithm is used, the drift parameter ν can be set differently according to the executed task, that is, ν will be a function of $\mho^j$.

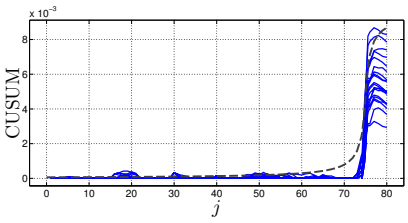


Fig. 4. CUSUM taken over sequential increments KL $\left(\hat{p}^{j-1}||\hat{p}^{j}\right)$ resulting from 3 different tasks. The increments are mixed at random, 10 cases are presented (solid lines). Also shown are the scaled values of w (dashed) for a comparison.

Example 3.2. (Simulation of Wear Monitoring for a Robot Executing Several Tasks:)
To illustrate the idea, a simulation study is carried out (see the appendix for details of the simulation model). The simulations are carried out considering 3 different tasks $\mho^t$, $t=0,1,2$, which are taken from real applications of an industrial robot. A realistic friction model is used that can explain, amongst others, the effects of wear w. A wear fault scenario is considered where, motivated by Blau (2009), the wear quantity w is assigned with a time-profile as

$$\mathsf{w}^j = \mathsf{w}^0 + \frac{\mathsf{w}^f - \mathsf{w}^0}{2}\xi(j) \tag{9a}$$

$$\xi(j) = 1 + \frac{j - j_m}{\left(1 + |j - j_m|^b\right)^{-b}} \tag{9b}$$

where j is the measurement sequence index, w^0 is the wear level prior to wear increases, w^f is the wear level after the increases. To illustrate a partial damage of the joint, the values $\mathsf{w}^0=0$ and $\mathsf{w}^f=50$ are chosen. The transition function $\xi(j)$ models the time behavior of the wear with an exponential factor. The variable j_m assigns the index where the transition from w^0 to w^f is half the way, the constant b changes the transition behavior. The remaining parameters are adjusted according to the wear evolution in a real fault scenario, with $j_m=75$ and $b=1$. The behavior of w is shown as the dashed line in Fig. 4. The figure also displays the CUSUM statistic for increments which are mixed at random for the different tasks $\mho^t$ in 10 different cases. The drift parameters are chosen as $\nu^j = \sigma^t + \mu^t$, where σ^t, μ^t are estimated from the fault free execution of task $\mho^t$. As it can be seen, monitoring is still possible even when data are generated under different conditions.

4. REDUCING SENSITIVITY TO DISTURBANCES

An alternative to achieve robustness to disturbances is to consider weighting the raw data $\mathbf{y}^j$ according to prior knowledge of the fault and disturbances. Defining a weighting vector $\mathbf{w} \in \mathbb{R}^N$, the weighted data are written as

$$\bar{\mathbf{y}}^j = \mathbf{w} \circ \mathbf{y}^j, \tag{10}$$

where $\circ$ is the Hadamard product (element-wise multiplication). The idea is to choose $\mathbf{w}$ to maximize the sensitivity to faults while increasing the robustness to disturbances.

Considering the basic framework presented in Sec. 2, a natural criteria for $\mathbf{w}$ would be to choose it such that KL $\left(\hat{p}^k(\mathbf{w})||\hat{p}^l(\mathbf{w})\right)$ is maximized when $\mathbf{y}^k$ is fault free and $\mathbf{y}^l$ is faulty and it is minimized in case they are both fault free or faulty. A general solution to this problem is however difficult since it depends on how $\hat{p}^j(\mathbf{w})$ was computed (e.g. the kernel function chosen) and maximization over (5). In this work, simpler criteria are used in a compromise of explicit solutions. As it will be shown, the results are directly related to linear discriminant analyses.

4.1 Choosing $\mathbf{w}$ *– Linear Discriminant Analyses*

Consider that the data set $\mathbf{Y}^M$ is available and the fault status (present or not) is known to each component $\mathbf{y}^j$ and the fault status is the same for each element in $\mathbf{y}^j$. The fault free data 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. Applying the weights $\mathbf{w}$ to the data set yields

$$\bar{\mathbf{Y}}^M \triangleq \left[\bar{\mathbf{y}}^0, \ldots, \bar{\mathbf{y}}^{M_0}, \bar{\mathbf{y}}^{M_0+1}, \ldots, \bar{\mathbf{y}}^{M_1+M_0}\right], \tag{11}$$

and the objective is to choose $\mathbf{w}$ such that the separation between the classes is maximized. A simple criterion is to consider the difference between the classes means. The cth class mean over all M_c observations is

$$\bar{m}^c \triangleq N^{-1} \sum_{i=0}^{N-1} \left[{M_c}^{-1} \sum_{j \in \mathcal{C}_c} w_i y_i^j \right] \tag{12}$$

$$= N^{-1} \sum_{i=0}^{N-1} w_i \underbrace{\left[M_c^{-1} \sum_{j \in \mathcal{C}_c} y_i^j \right]}_{\triangleq m_i^c} = N^{-1} \mathbf{w}^T \mathbf{m}^c. \tag{13}$$

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

$$\bar{m}^1 - \bar{m}^0 \propto \mathbf{w}^T(\mathbf{m}^1 - \mathbf{m}^0). \tag{14}$$

This problem is equivalently found in linear discriminant analyses, see Bishop (2007). Constraining $\mathbf{w}$ to unit length in order to achieve a meaningful solution, it is easy to see that the optimal choice is to take $\mathbf{w} \propto (\mathbf{m}^1 - \mathbf{m}^0)$, Bishop (2007).

A criterion based only on the distance between the classes mean does not consider the variability found within each class, for instance caused by disturbances. An alternative is to consider maximum separation between the classes mean while giving small variability within each class. Considering a measure of variability for each class as the mean of variances for each ith component,

$$\bar{s}^c \triangleq N^{-1} \sum_{i=0}^{N-1} \left[M_c^{-1} \sum_{j \in \mathcal{C}_c} (w_i y_i^j - w_i m_i^c)^2 \right] \tag{15}$$

$$= N^{-1} \sum_{i=0}^{N-1} w_i^2 \underbrace{\left[M_c^{-1} \sum_{j \in \mathcal{C}_c} (y_i^j - m_i^c)^2 \right]}_{\triangleq s_i^c} \tag{16}$$

$$= N^{-1} \mathbf{w}^T \mathbf{S}^c \mathbf{w}, \tag{17}$$

where $\mathbf{S}^c$ 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

$$\frac{(\bar{m}^1 - \bar{m}^0)^2}{\bar{s}_1 + \bar{s}_0} \propto \frac{\mathbf{w}^T(\mathbf{m}^1 - \mathbf{m}^0)(\mathbf{m}^1 - \mathbf{m}^0)^T\mathbf{w}}{\mathbf{w}^T(\mathbf{S}^1 + \mathbf{S}^0)\mathbf{w}}, \tag{18}$$

which is a special case of the Fisher criterion, see Bishop (2007). It can be shown that solutions for this problem satisfy

$$\mathbf{w} \propto (\mathbf{S}^1 + \mathbf{S}^0)^{-1}(\mathbf{m}^1 - \mathbf{m}^0). \tag{19}$$

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

Notice however that the solutions (14) and (19) require the data to be synchronized, which is difficult in many practical applications. In case this is possible (for instance using simulations), the result of such analyses might reveal some useful pattern of the weights. For instance, if the weights are strongly correlated to measured data, an approximate function can be used to describe the weights depending on the data, e.g. $w_i = h(y_i^j)$ for a continuous function $h(\cdot)$.

Example 4.1. (Simulation of Robust Wear Monitoring in a Robot Joint:)
To illustrate the ideas presented in this section, a simulation study is carried out (see the appendix for details of the simulation model). A path $\mho$ is simulated $M = M_1 + M_0$ times under different conditions, forming a data set $\mathbf{Y}^M$, with $M_1 = M_0 = 100$. A realistic friction model is used that represents the effects of wear w and joint temperature T. The two batches of data are generated with the following settings

$$\boldsymbol{\tau}^k : \mathsf{w} = 0,\ T \sim \mathcal{U}[\underline{\mathrm{T}}, \underline{\mathrm{T}} + \Delta_T],\ k \in \mathcal{C}_0 \tag{20a}$$

$$\boldsymbol{\tau}^l : \mathsf{w} = \mathsf{w}_\mathrm{c},\ T \sim \mathcal{U}[\underline{\mathrm{T}}, \underline{\mathrm{T}} + \Delta_T],\ l \in \mathcal{C}_1 \tag{20b}$$

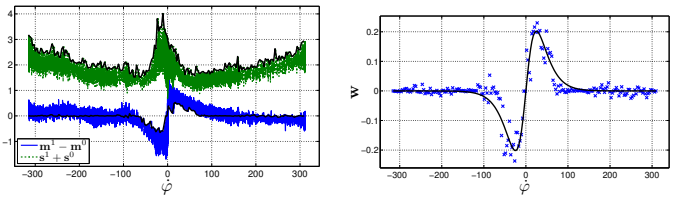
where $k \in \mathcal{C}_0$ corresponds to the first M_0 sequences and $l \in \mathcal{C}_1$ are the remaining ones, $\mathsf{w}_\mathrm{c} = 35$ is a wear level considered critical to generate an alarm (see Bittencourt et al. (2011)). Here, T is considered random, with uniform distribution given by $\underline{\mathrm{T}} = 30°\mathrm{C}$ and $\Delta_T = 40°\mathrm{C}$. This assumption is carried out for analyses purposes.

The average distance $m_i^1 - m_i^0$ and total variability $s_i^1 + s_i^0$ are displayed as a function of the joint speed $\dot{\varphi}$ in Fig. 5(a). In the same figure, a worst case estimate, largest $s_i^1 + s_i^0$ and $m_i^1 - m_i^0$ closest to zero, is also shown (solid lines). Fig. 5(b) presents the ratio for such worst case estimate, which is considered as the optimal weights according to (19). As it can be seen, the optimal weights present a strong correlation with $\dot{\varphi}$, which is not a surprise since the effects of w and T depend on $\dot{\varphi}$, recall Fig. 1. The solid line in Fig. 5(b) is a function approximation of the optimal weights given by

$$w(\dot{\varphi}) = \mathrm{sech}(\beta\dot{\varphi})\tanh(\alpha\dot{\varphi}) \tag{21}$$

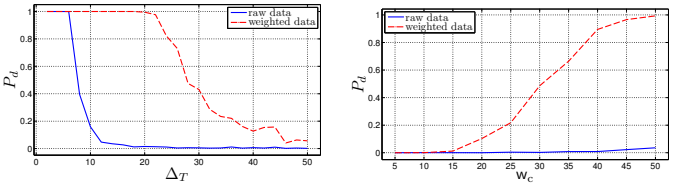
with $\alpha = 1.45\,10^{-2}$ and $\beta = 4.55\,10^{-2}$. Effectively, the optimal weighting function selects a speed region that is more relevant for robust wear monitoring. In Bittencourt et al. (2011), a similar behavior was found for the quality (variance) of a wear estimate for different speeds under temperature disturbances.

The performance improvements 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 data set generated according to (20), 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



(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 Fig.5(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 and as function of the fault size w_c for $\Delta_T = 25°\mathrm{C}$. Notice the considerably improvements when using the weighted data.

$$\mathcal{H}_0 : m, n \in \mathcal{C}_0 \text{ or } m, n \in \mathcal{C}_1 \tag{22a}$$

$$\mathcal{H}_1 : m \in \mathcal{C}_0, n \in \mathcal{C}_1 \text{ or } m \in \mathcal{C}_1, n \in \mathcal{C}_0. \tag{22b}$$

In view of the framework presented in Sec. 2, this problem is analyzed by computing the distribution of $\mathrm{KL}\,(\hat{p}^m || \hat{p}^n)$ for each hypothesis.

The overlap of these distributions gives the probability of false, P_f, and probability of detection, P_d (the problem is a binary hypothesis test, see Van Trees (2001) for more). The procedure is repeated for different values of Δ_T, with and without the use of the weighting function. For the fixed $P_f = 0.01$, Fig. 6(a) presents the achieved P_d as a function of Δ_T. 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 $\Delta_T = 25°\mathrm{C}$, data are generated according to (20) for different values of w_c. Similarly, the hypotheses distributions are computed. Fig. 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 obvious.

5. CONCLUSIONS AND FUTURE WORK

The paper presented a framework for condition monitoring of systems that operate in a repetitive manner. A data driven method was proposed that considers changes in the distribution of data samples obtained from multiple executions of one or several tasks. This was achieved with the use of kernel density estimators and the Kullback-Leibler distance measure between distributions. The suggested approach of monitoring the accumulated incremental changes allowed the framework to be extended to the

cases where fault free data are unavailable and/or the repetitive behavior of the system varies. The use of a weighting function was proposed in order to reduce sensitivity to unknown disturbances and increase sensitivity to faults. The methods were illustrated using real data and simulations for the problem of (robust) wear monitoring in an industrial robot joint. The results show that robust wear monitoring in robot joints is made possible with the proposed methods. For a complete validation however, more experiments using different cycles and with temperature variations are needed. The proposed methods should also be bench marked to existing methods.

The paper dealt only with univariate sequences $\mathbf{y}^j$. All quantities used (e.g. NSEDEs and KL) can also be defined for the multivariate case so, in principle, the framework can be extended to monitor multiple variables.

The KLD is in fact a specialization of a f-divergence Reid and Williamson (2011), a family of functions that can be used as a measure of the differences between distribution functions. It might be interesting to study the use and properties of different divergences. A similar argument is valid regarding the choice of kernel function to compute the NSEDEs and criteria for choosing the smoothing parameter.

Several filtering schemes are possible for the alternative of monitoring consecutive increments $\mathrm{KL}\left(\hat{p}^{j-1}||\hat{p}^{j}\right)$, e.g. using a moving window or a moving average. When monitoring the accumulated changes, it is important to consider how often should the sequences be compared. This issue is related to the time behavior of the fault, which is typically unknown.

While this paper focused on a method to generate a quantity sensitive to faults, the important issues of alarm generation and diagnosis were not addressed.

REFERENCES

Bishop, C.M. (2007). *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer, 1st ed. 2006. corr. 2nd printing edition.

Bittencourt, A.C., Axelsson, P., Jung, Y., , and Brogårdh, T. (2011). Modeling and identification of wear in a robot joint under temperature disturbances. In *In 18th IFAC World Congress*.

Bittencourt, A.C., Wernholt, E., Sander-Tavallaey, S., and Brogårdh, T. (2010). An extended friction model to capture load and temperature effects in robot joints. In *The 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*.

Blau, P.J. (2009). Embedding wear models into friction models. *Tribology Letters*, 34(1).

Bowman, A.W. and Azzalini, A. (1997). *Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations (Oxford Statistical Science Series)*. Oxford University Press, USA. URL `http://www.worldcat.org/isbn/0198523963`.

Brambilla, D., Capisani, L., Ferrara, A., and Pisu, P. (2008). Fault detection for robot manipulators via second-order sliding modes. *Industrial Electronics, IEEE Transactions on*, 55(11), 3954–3963.

Caccavale, F., Cilibrizzi, P., Pierri, F., and Villani, L. (2009). Actuators fault diagnosis for robot manipulators with uncertain model. *Control Engineering Practice*, 17(1), 146 – 157.

Caccavale, F. and Villani, L. (eds.) (2003). *Fault Diagnosis and Fault Tolerance for Mechatronic Systems: Recent Advances*. Springer Tracts in Advanced Robotics, Vol. 1. Springer-Verlag, New York.

De Luca, A. and Mattone, R. (2004). An adapt-and-detect actuator fdi scheme for robot manipulators. In *Robotics and Automation, 2004. Proceedings. ICRA '04. 2004 IEEE International Conference on*, volume 5, 4975 – 4980 Vol.5.

Eski, I., Erkaya, S., Savas, S., and Yildirim, S. (2010). Fault detection on robot manipulators using artificial neural networks. *Robotics and Computer-Integrated Manufacturing*.

Freyermuth, B. (1991). An approach to model based fault diagnosis of industrial robots. In *1991 IEEE International Conference on Robotics and Automation*, volume 2, 1350–1356.

Gustafsson, F. (2000). *Adaptive Filtering and Change Detection*. John Wiley & Sons, LTD.

Kato, K. (2000). Wear in relation to friction – a review. *Wear*, 241(2), 151 – 157.

Khan, F.I. and Abbasi, S.A. (1999). Major accidents in process industries and an analysis of causes and consequences. *Journal of Loss Prevention in the Process Industries*, 12(5), 361 – 378.

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

Rao, B.K.N. (1998). Condition monitoring and the integrity of industrial systems. In A. Davies (ed.), *Part 1: Introduction to Condition Monitoring*, Handbook of Condition Monitoring – Techniques and Methodology, chapter 1, 3–34. Chapman & Hall, London, UK.

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

Van Trees, H.L. (2001). *Detection, Estimation and Modulation Theory, Part I*. Wiley, New York.

Vemuri, A.T. and Polycarpou, M.M. (2004). A methodology for fault diagnosis in robotic systems using neural networks. *Robotica*, 22(04), 419–438.

Appendix A. SIMULATION MODEL

The simulation model considered is the 2 link manipulator with elastic gear transmission presented in the benchmark problem in Moberg et al. (2008). 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.

With the objective of studying friction changes related to wear in a robot joint, the static friction model described in Bittencourt et al. (2011) is included in the simulation model. The static friction model was developed from empirical studies in a robot joint (Bittencourt et al. (2010)) and describes the effects of angular speed $\dot{\varphi}$, manipulated load torque τ_m, temperature T and wear w.

In the simulation setup, a task $\mho$ is described by a set of reference joint positions 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 $\mho$ varies only slightly due to the stochastic components and feedback.

The paths $\mho$ are taken from real applications of a 6 axes industrial robot. In order to make it possible to simulate them with the 2 links robot model, the angles of joints 2 and 3 of the real robot are matched to joints 1 and 2 in the simulation. In this setup, the two main axes of the robot are studied.