On Modeling and Diagnosis of Friction and Wear in Industrial Robots

André Carvalho Bittencourt

Division of Automatic Control
Department of Electrical Engineering
Linköping University, SE-581 83 Linköping, Sweden
http://www.control.isy.liu.se
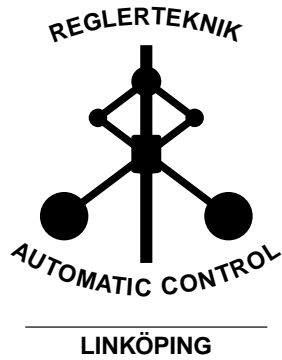andrecb@isy.liu.se

Linköping 2012

Linköping studies in science and technology. Thesis.
No. 1516

On Modeling and Diagnosis of Friction and Wear in Industrial Robots

André Carvalho Bittencourt

andrecb@isy.liu.se
www.control.isy.liu.se
Department of Electrical Engineering
Linköping University
SE-581 83 Linköping
Sweden

To the memory of my brother.

Abstract

Industrial robots are designed to endure several years of uninterrupted operation and therefore are very reliable. However, no amount of design effort can prevent deterioration over time, and equipments will eventually fail. Its impacts can, nevertheless, be considerably reduced if good maintenance/service practices are performed. The current practice for service of industrial robots is based on preventive and corrective policies, with little consideration about the actual condition of the system. In the current scenario, the serviceability of industrial robots can be greatly improved with the use of condition monitoring/diagnosis methods, allowing for condition-based maintenance (CBM).

This thesis addresses the design of condition monitoring methods for industrial robots. The main focus is on the monitoring and diagnosis of excessive degradations caused by wear of the mechanical parts. The wear processes may take several years to be of significance, but can evolve rapidly once they start to appear. An early detection of excessive wear levels can therefore allow for CBM, increasing maintainability and availability. Since wear is related to friction, the basic idea pursued is to analyze the friction behavior to infer about wear.

To allow this, an extensive study of friction in robot joints is considered in this work. The effects of joint temperature, load and wear changes to static friction in robot a joint are modeled based on empirical observations. It is found that the effects of load and temperature to friction are comparable to those caused by wear. Joint temperature and load are typically not measured, but will always be present in applications. Therefore, diagnosis solutions must be able to cope with them.

Different methods are proposed which allow for robust wear monitoring. First, a wear estimator is suggested. Wear estimates are made possible with the use of a test-cycle and a friction model. Second, a method is defined which considers the repetitive behavior found in many applications of industrial robots. The result of the execution of the same task in different instances of time are compared to provide an estimate of how the system changed over the period. Methods are suggested that consider changes in the distribution of data logged from the robot. It is shown through simulations and experiments that robust wear monitoring is made possible with the proposed methods.

Populärvetenskaplig sammanfattning

Moderna industrirobotar konstrueras för att kunna arbeta oavbrutet under flera år och är därför väldigt pålitliga. Det är dock omöjligt att förhindra att roboten slits, att prestanda försämras och att slutligen fel uppstår. Effekterna kan dock reduceras med hjälp av rutiner för service och underhåll. För närvarande bygger dessa rutiner oftast på förebyggande underhåll, där relativt liten hänsyn tas till det faktiska tillståndet hos roboten. Situationen kan dock förbättras genom att i större utsträckning använda metoder för övervakning och diagnos av systemet och därmed kan tillämpa en högre grad av tillståndsbaserat underhåll.

Denna avhandling behandlar utformning metoder för tillståndsövervaknings för industrirobotar. Arbetet behandlar främst övervakning och diagnos av försämrade prestanda på grund av förslitning av mekaniska delar. Förslitningen kan ta flera år för att börja utvecklas, men därefter kan förloppet gå snabbt. Om förslitningen upptäcks i tid kan tillståndsbaserat underhåll tillämpas, vilket kan förhindra fel och öka tillgängligheten hos roboten. Eftersom förslitning är nära relaterad till friktion är därför grundidén att studera friktion för att dra slutsatser om förslitningen.

För att möjliggöra detta har en omfattande studie av friktion hos industrirobotar genomförts. Inverkan av temperatur och belastning har studerats och modellerats utgående från omfattande experimentella försök. Dessa har visat att effekterna av belastning och temperatur på friktionen är i storlek jämförbara med förslitningens inverkan. Eftersom såväl temperatur som belastning varierar när roboten arbetar, men ingen av dem normalt brukar mätas måste en diagnosmetod kunna hantera dessa variationer.

Avhandlingen föreslår två olika metoder för robust övervakning och diagnos av förslitning, där den första metoden innebär att man skattar förslitningen genom att använda en så kallad testcykel och en skattad friktionsmodell. Den andra metoden utnyttjar att en industrirobot ofta arbetar repetitivt där en viss rörelse upprepas. Beteendena hos roboten då samma rörelse upprepas hos roboten vid olika tillfällen jämförs och används för att bedöma hur robotens egenskaper förändrats. Metoden baseras på förändringar i fördelningen hos uppmätta data, t. ex. moment, från roboten. Simuleringar och experiment indikerar att det är möjligt att skapa robusta metoder för övervakning och diagnos av förslitning hos industribotar.

Acknowledgments

First, I would like to thank my supervisor Prof. Svante Gunnarsson for the excellent guidance through these years, for keeping me motivated, for the constant support and for always finding (quick and accurate) solutions to my worries. Perhaps if it was not for Dr. Shiva Sander-Tavallaey I would not have taken this journey, thank you for inviting me to graduate education and for all your guidance; thanks also to Prof. Lennart Ljung and Prof. Svante Gunnarsson for accepting me into the group. Special thanks to the valuable input I received from my co-supervisors Dr. Erik Wernholt and Prof. Mikael Nörloff.

Being a graduate student at the ISY/RT group has been a remarkable experience. I feel extremely privileged for the things I have learned, taught and the people I have met. Such excellence is the result of several years of dedicated work, and I would like to express my gratitude to everyone behind our organizational structure. To mention some, thank you Prof. Lennart Ljung and Prof. Svante Gunnarsson for your leadership; Ulla Salaneck, Åsa Karmelind and now Ninna Stensgård for the adimistrative support; all of our gurus, specially the ones behind this nice LATEX thesis template, Dr. Gustaf Hendeby and Dr. Henrik Tidefelt. The quality of this thesis was also considerably improved with the comments I received from Prof. Svante Gunnarsson, Prof. Mikael Nörloff, Dr. Torgny Broghård, Patrik Axelsson and Dr. Shiva Sander-Tavallaey; thank you for your time. Thanks also to Dr. Erik Frisk for the interesting discussions on diagnosis.

This work would not have been the same if it was not for the close collaboration with ABB. ABB not only supported me financially, via VINNOVA's Industry Excellence Center LINK-SIC, but also with priceless expertise and guidance. Thank you Dr. Shiva Sander-Tavallaey and Lic. Niclas Sjöstrand for giving me access to resources and for the constant interest in my work. Thank you Dr. Shiva Sander-Tavallaey, Dr. Karin Saarinen, MSc. Hans Andersson, Dr. Stig Moberg and many others for all the valuable input and for helping me feel home at ABB. Special thanks to Dr. Torgny Broghård, who has acted in different levels for the success of this collaboration, from his role in the LINK-SIC board to the detailed reviews of my work; thanks for your guidance, expertise, and humbleness. I was introduced to diagnosis of industrial robots already in 2007 when I took an "exjobb" at ABB. This is probably where my journey to a post graduate education started. Special thanks to Lic. Niclas Sjöstrand and Prof. Bo Wahlberg who acted as my supervisors back then. Thanks also to Prof. Alf Isaksson with whom I worked in other projects related to ABB.

The endurance of the long-dark-cold winters and the brief-bright-warm summers of Sweden was made much more enjoyable with the presence of good friends in my live. With Daniel Ankelhed I have shared most of my time at the office; thanks for your patience, company, and music knowledge exchanges. Sina Kosh···kazad is *probably the best host in the world*, thanks for all the nice time we shared. Thank you Fredrk Lindsten, Karl Granström, Emre Ozkan, Tohid Ardeshiri, Henrik Ohlsson, Peter Rosander, Carsten Fritsche, Patrik Axelsson, Lubos Váci, Saikat

Saha, Tianshi Chen, Zoran Sjanic, Ylva Jung, Jessica Escobar and everyone else for all the nice moments we shared. Thank you all for helping me counterbalance the stress loads of the PhD education with good-humored discussions during our **w**-fika-**o**-fika-**r**-fika-**k** routine, moped races, *il krakens*, gala dinner frenzies, flying limones, giant mozzarellas, pink elephants, climbing and kitesurfing adventures, (icy/bloody) barbecues, Peruvian giraffes, and so much more. I must also mention my friends from college, with whom I shared my first experiences in Engineering. Specially those of the legendary 031 class of Automatic Control, PETEEL and more. Thank you all for the Linguiços, churrascos, Oktobers, pelican competitions, etc. I can only hope that we will always keep in touch.

Thank you Alícia for your love and *presence* in my life. I will always be grateful for all the efforts you have made for our relationship. I am sorry that it has meant that you had to interrupt your studies in Brazil, say goodbye to your family, learn a foreign language, freeze, and so many other things that happened and will happen; I can only hope that I will always be worthy. Thank you for your patience, your cosy company when having a beer, relaxing at home, going on unforgettable climbing adventures, and so much more. I love you Bosa.

Acima de tudo eu gostaria de agradecer ao suporte e amor incondicionais da minha família. De todos os aprendizados que eu acumulei, os mais importantes foram vocês quem ensinaram. Obrigado pela dedicação e amor sempre presentes em minha vida. Obrigado pelo suporte que recebi e recebo desde que chegou a hora de buscar meus sonhos, mesmo que isso resultasse em nos separarmos geograficamente. Peço desculpas pelas minhas falhas em manter maior frequência no contato e por ser tão esquecido. Com amor,

André Carvalho Bittencourt,
Linköping, January 2012.

Contents

I   Background

Notation

ABBREVIATIONS

Abbreviation	Meaning
ISO	International Organization for Standardization
ABB	Asea Brown Boveri Ltd.
SRAM	Safety, Reliability, Availability and Maintainability
CBM	Condition Based Maintenance
KLD	Kullback-Liebler Divergence
KL	Kullback-Liebler distance
CRB	Cramér-Rao lower Bound
NSEDE	Nonparametric Smooth Density Estimate
ROC	Receiver Operating Characteristic
DOF	Degree of Freedom
BL	Boundary Lubrication region of the friction curve
ML	Mixed Lubrication region of the friction curve
EHL	Elasto-Hydrodynamic Lubrication region of the friction curve

Operators and Miscellaneous

Notation	Meaning
$\frac{\mathrm{d}}{\mathrm{d}x}f(x)$	Derivative of $f(x)$ with respect to x
$\dot{x}(t)$	Derivative of $x(t)$ with respect to time
$\triangleq$	Equal by definition
$\sim$	Denotes "is distributed according to"
$\propto$	Denotes "is proportional to"
$\mathbf{T}\{f(x)\}$	Integral transform of $f(x)$
$\mathbf{F}\{f(x)\}$	Fourier transform of $f(x)$
$\mathbf{F}^{-1}\{f(\nu)\}$	Inverse Fourier transform of $f(\nu)$
$\mathbf{Z}\{f(x)\}$	Z-transform of $f(x)$
$\underset{x}{\operatorname{argmin}}\, f(x)$	The value of x that minimizes $f(x)$
$\operatorname{sign}(\cdot)$	Sign function
$\lvert x \rvert$	Modulus of x
$\lVert \cdot \rVert_\delta$	The δ vector or induced matrix norm
$\mathbb{N}$	Set of natural numbers
$\mathbb{R}$	Set of real numbers
$\mathbb{C}$	Set of complex numbers
i	The complex number, i.e. $\sqrt{-1}$
$d(\cdot,\cdot)$	A distance measure
$s(k)$	A test quantity
$g(k)$	A test statistic
i, j	Auxiliary indexes in $\mathbb{N}$

NOTATION FOR ROBOTICS

Notation	Meaning
φ	Joint(s) angular position(s)
φ_a	Joint(s) angular position(s) at the arm side
φ_m	Joint(s) angular position(s) at the motor side
φ^i	ith joint angular position
φ^r	Reference joint(s) angular position(s)
p^i	ith coordinate frame
R_i^{i-1}	Rotation from frame i to $i-1$
d_i^{i-1}	Translation from frame i to $i-1$
H_i^{i-1}	Homogeneous transformation from frame i to $i-1$
P^i	ith augmented coordinate frame
X	End-effector pose (position and orientation)
$\mathcal{J}(\cdot)$	Analytical Jacobian
$\mathcal{L}(\cdot,\cdot)$	Lagrangian function
$\mathcal{K}(\cdot,\cdot)$	Kinetic energy
$\mathcal{P}(\cdot)$	Potential energy
$M(\cdot)$	Inertia matrix
$C(\cdot)$	Coriolis and centrifugal torques
$K(\cdot)$	Stiffness matrix
$D(\cdot)$	Damping matrix
τ_g	Gravity-induced load torque(s)
τ_f	Friction torque(s)
$\tau_{f,a}$	Friction torque(s) at the arm side
$\tau_{f,m}$	Friction torque(s) at the motor side
τ	Applied torque(s)
τ_a	Applied torque(s) at arm side, as in (2.12)
τ_m	Applied torque(s) at motor side, as in (2.12)
τ_m^{ffw}	Feedforward torque(s) at motor side, as in Section 2.4.1
i_m	Motor(s) applied current(s), as in Section 2.4.1
i_m^r	Reference motor(s) current(s), as in Section 2.4.1
η	Inverse gearbox ratio(s)
$\mho$	Trajectory
$\mho^r$	Reference trajectory

Notation for Friction Modeling

Notation	Meaning
$\mathcal{F}$	Generalized friction
$\mathcal{X}$	Generalized friction states
τ_l	Resulting component of the load torques which is to the joint DOF
τ_p	Resulting component of the load torques which is perpendicular to the joint DOF
w	Wear level(s)
T	Joint(s) temperature(s)
$g(\cdot)$	Function describing the velocity weakening behavior of the friction curve, as in (3.2)
$h(\cdot)$	Function describing the velocity strengthening behavior of the friction curve, as in (3.2)
z	Internal friction state in a dynamic friction model
σ_0, σ_1	Stiffness and damping parameters of the LuGre model as in (3.1)
F_c	Coulomb parameter
F_s	Standstill parameter
$\dot{\varphi}_s$	Stribeck speed parameter
α	Exponent parameter describing the Stribeck nonlinearity
F_v	Viscous parameter
F_μ	Viscous parameter describing the non-Newtonian behaviour of the lubricant
β	Exponent parameter describing the non-Newtonian behaviour of the lubricant

NOTATION FOR MODELS OF SYSTEMS AND IDENTIFICATION

Notation	Meaning
z	Deterministic set of inputs
v	Random unknown set of inputs
u	Deterministic known set of inputs
d	Deterministic uninteresting unknown set of inputs
f	Deterministic interesting unknown set of inputs
y	Set of known outputs
x	Set of states
k	Sample index
z	Complex variable following from the **Z**-transform
$\left[\begin{array}{c\|c} A & B \\ \hline C & D \end{array}\right]$	State space realization of the system $x(k+1)=Ax(k)+Bu(k),\ y(k)=Cx(k)+Du(k)$
θ	Vector of parameters
θ_0	True vector of parameters
$\mathcal{M}$	Model structure
$\mathcal{M}(\theta)$	A model instance of the model structure $\mathcal{M}$ with parameter θ
$\phi(\cdot)$	Regression vector function
$\phi_i(\cdot)$	ith component of the regression vector function
$\Phi(\cdot)$	Matrix of stacked regressors
η	Vector of parameters that are linear in the regression
ρ	Vector of parameters that are nonlinear in the regression
$\hat{\theta}$	Estimate of parameter θ_0
$\hat{\theta}_N$	Estimate of parameter θ_0 from N data samples
$\hat{\theta}(k)$	Estimate of parameter θ_0 at sample index k
$\hat{y}(k,\theta)$	Predictor function
$\varepsilon(k,\theta)$	Prediction error, defined as $y(k) - \hat{y}(k\|\theta)$, sometimes called a residual
$\psi(k,\theta)$	Gradient of the prediction error with respect to θ
$\hat{R}_N$	Information matrix estimate from N data samples
γ_0	True noise variance
$\hat{\gamma}_N$	Noise variance estimate from N data samples
P	Asymptotic covariance matrix of $\sqrt{N}(\hat{\theta}_N - \theta_0)$
$\hat{P}_N$	Covariance estimate from N data samples
$\hat{P}(k)$	Covariance estimate at sample index k

Notation for Probability, Statistics and Decision Theory

Notation	Meaning
Y	Random variable
$\mathbf{y}$	Sample from the random variable Y
$\mathbb{E}[Y]$	Expectation of random variable Y
$p(y)$	Probability distribution (density) function of Y
$p(y\|x)$	Probability distribution (density) function of Y conditional on X
$\Phi_Y(\cdot)$	Characteristic function of Y
$K_h(\cdot)$	Weighting function, as in (4.27)
$k_h(\cdot)$	Kernel function, as in (4.27)
h	Smoothing parameter, as in (4.27)
$\hat{p}(y)$	Kernel density estimate of $p(y)$ from a sample $\mathbf{y}$, as in (4.27)
$\mathcal{N}(\mu, \gamma)$	The normal (Gaussian) distribution with mean μ and variance γ
$\mathcal{U}(\underline{y}, \overline{y})$	The uniform distribution with lower limit $\underline{y}$ and upper limit $\overline{y}$
$D_{\mathrm{KL}}(p\|q)$	Kullback-Leibler divergence between densities $p(y)$ and $q(y)$
$\mathrm{KL}(p\|q)$	Kullback-Leibler distance between densities $p(y)$ and $q(y)$
$x \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \hbar$	A binary test. Chooses $\mathcal{H}_1$ if $x \geq \hbar$ and $\mathcal{H}_0$ otherwise
$\hbar$	A threshold
$\mathcal{H}_0$	Null hypothesis in a binary test
$\mathcal{H}_1$	Alternative hypothesis in a binary test
P_f	Probability of false detection in a binary test
P_d	Probability of correct detection in a binary test

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 high 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 (Thompson (1999)), 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 order to support maintenance actions, the use of methods to determine the condition of the equipment is desirable. Condition monitoring methods can be used to increase SRAM and minimize maintenance costs, allowing for condition-based maintenance (CBM). Preferably, these methods should perform automatically and with no interruption of the equipment's operation.

This thesis addresses the design of condition monitoring methods for an equipment which is many times of crucial importance in manufacturing, *industrial robots*. The main focus is on the monitoring and diagnosis of excessive degradation caused by wear of the mechanical parts. The wear processes may take several years to be of significance, but can evolve rapidly once it starts to appear. An early detection of excessive wear levels can therefore allow for CBM and increased SRAM. Since wear is related to friction, the basic idea pursued here is to analyze the friction behavior to infer about wear. To allow this, an extensive study of friction in robot joints is considered in this work, and different solutions for wear monitoring are proposed and evaluated. This chapter presents an introduction and motivation to the problem, followed by the outline and main research contributions of the thesis.

1
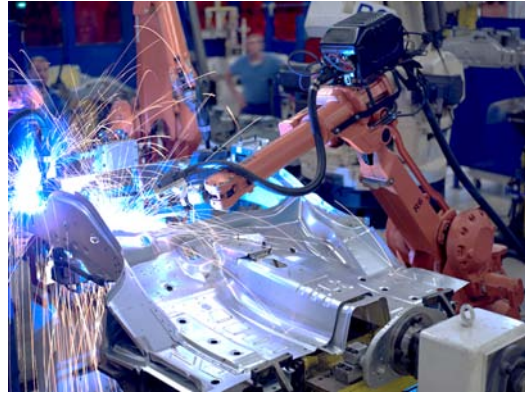
(a) Pick and place.   *(b)* Spot welding.

Figure 1.1: *Examples of applications of industrial robots where high avail-*
ability is critical. The economical damages of an unpredicted robot stop in a
production line are counted by the second.

1.1   Motivation

Industrial robots are used as a key factor to improve productivity, quality and
safety in automated manufacturing. Robot installations are many times of cru-
cial importance in the processes where they are used. As illustrated by the ap-
plications found in Figure 1.1, an unexpected robot stop or malfunction has the
potential to cause downtimes of entire production lines, with consequent produc-
tion losses and economical damages. Increased availability and maintainability
are therefore critical for industrial robots. In practice, robot supervision is still
mostly related to protection and safety. Functionalities such as collision detection
and brake monitoring are already available in some commercial platforms. How-
ever, there are currently little commercial solutions for automated monitoring of
the mechanical parts of the robot.

For industrial robots, the requirements on high availability are most of the times
achieved based on preventive and corrective maintenance policies. Service rou-
tines are typically performed on-site, with a service engineer. Service actions are
based on specific on-site tests or simply from a pre-determined schedule. Such
pre-determined scheduled maintenance is based on the estimated components'
lifespan, with considerable margins. These maintenance solutions can deliver
high availability, reducing downtimes. The drawbacks are however the high costs
due to on-site inspections by an expert and/or due to unnecessary maintenance
actions that might take place.

In the current scenario, the serviceability of industrial robots can be greatly im-
proved with the use of condition monitoring methods, allowing for CBM. There
are however requirements from both the robot user and the service contractor.

The robot user seeks for improved SRAM. Therefore, the monitoring solution should be reliable and accurate, with minimal, preferably none, intervention with the robot's operation.

The service contractor seeks for reduced costs for the service operation. Therefore, the monitoring solution should be performed remotely, it should be as automated as possible and use no extra sensors than what are available in a typical robot setup.

Achieving these compromises is however a very challenging task. This is partly because some faults are difficult to predict, or affect the operation of the system abruptly, e.g. a wire cut or a power supply drop. These type of faults, even when detected, might still cause damages. Therefore, with focus on service, the interest is limited to the monitoring of faults that can be diagnosed before a critical degradation takes place, so that appropriate maintenance actions can take place.

An important type of such fault is related to the *wear* processes in a robot joint. Wear develops with time/usage and might be detected at an early stage, allowing for CBM. The wear processes inside a robot joint cause an eventual increase of wear debris in the lubricant. A possible solution is therefore to monitor the iron content in the lubricant. For a typical robot setup, this type of approach will however contradict most of the user's and service contractor's requirements.

An important characteristic of wear is that affects friction in the robot joint. An alternative solution, explored in this work, is thus to *monitor friction changes to infer about wear*. Since the friction torques must be overcome by the motor torques during its operation, it is possible to extract information about friction from available signals. Friction is however dependent on other factors than wear. In fact, the changes caused, e.g., by temperature are typically at least as significant as those caused by wear.

1.2   Problem Formulation and Approach

The main objective of this work is to develop methods for *friction (wear) monitoring in industrial robot joints*. This work is in the overlap of three main research areas, namely: *industrial robotics*, *tribology* and *diagnosis*. To consider a problem in their intersection will require understanding of the available techniques from each of these fields. Therefore, much of this thesis is dedicated to provide an overview of these research areas. This will help to motivate the research presented and to identify needs for innovative solutions.

The outline of the thesis and the main contributions are described next. The presentation is, of course, focused on aspects that are relevant to the problems addressed in the thesis.

1.3 Thesis Outline

The thesis is divided into two parts. Part I gives an overview of the related research areas and provides a background to the research contributions. The main research contributions are presented in Part II, which contains edited versions of published papers.

The outline for Part I is summarized below,

Chapter 2 provides an introduction to industrial robotics. The purpose is to provide an overview of important aspects to consider when working with industrial robots, the main limitations and challenges.

Chapter 3 focuses on describing the friction and wear phenomena in industrial robot joints. It provides an overview of the challenges and motivations behind this work.

Chapter 4 provides an overview of the diagnosis process. It includes a description of the tasks and challenges involved. Attention is given to provide an overview of different methods for wear monitoring in a robot joint.

Chapter 5 presents a summary of the work and a discussion of next steps to come.

Each chapter in Part I is concluded by presenting connections to the research papers of Part II. A summary of the main research contributions of Part II is given below.

Extensive studies of friction in a robot joint are presented in Papers A and B. The effects of joint angle, load torques, temperature and wear are analyzed through detailed empirical studies.

Friction modeling, the effects of load torques and temperature to friction in a robot joint are modeled and identified in Paper A.

Wear modeling, the effects of wear to friction in a robot joint are also modeled and identified in Paper B. Based on Blau (2009) and the observed wear behavior in accelerated wear tests, a model for the evolution of wear with time is also suggested in Paper C.

Wear identification is proposed as a method for wear monitoring in Paper B based on a test-cycle.

Monitoring of repetitive systems is considered in Paper C. Methods tailored for systems that operate in a repetitive manner are presented with applications to robust wear monitoring in a robot joint. The methods are suitable for use with no interruption of the system operation.

Studies of limitations imposed by disturbances, specially those caused by temperature, are presented for the wear monitoring methods of Papers B and C.

1.3.1   Publications

Edited versions of the following papers are included in Part II of this thesis. The background between the research contributions for each paper are discussed next.

Paper A: Static Friction in a Robot Joint - Modeling and Identification of Load and Temperature Effects

> A. C. Bittencourt and S. Gunnarsson. Static friction in a robot joint - modeling and identification of load and temperature effects. *ASME Journal of Dynamic Systems, Measurement, and Control.* Accepted for publication.

Background. Several reports can be found in the literature regarding the dependency of friction in a robot joint to others factors than speed. However, to the best of the authors knowledge, no detailed empirical studies of these effects had been previously performed in a robot joint.

This work provides a deeper understanding of these phenomena based on experiments that were carried out during the summer of 2009 at ABB. The main motivation for the studies was to gather understanding of these phenomena. This would serve as a pre-requisite to the development of wear monitoring methods based on studies of friction. As a result, a model that can explain the relevant effects of temperature and load to static friction was developed and validated. The developed model is important not only for the design and validation of diagnosis methods but also for control and simulation.

Paper B: Modeling and Identification of Wear in a Robot Joint under Temperature Disturbances

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

Background. Different approaches had been previously proposed for diagnosis of friction changes in a robot joint. However, no report could be found that considers the effects of wear changes explicitly. Moreover, no detailed studies of the undesired disturbances caused by temperature and load to friction were found. This is partly because there were no available models to explain these phenomena. Another important aspect is that performing experiments for wear monitoring is a very time consuming and expensive task.

Based on accelerated wear experiments performed in ABB, the effects of wear to friction were studied and a wear model was developed. This wear model, combined with the model of Paper A, is very important for the design and evaluation of wear diagnosis solutions. They are used extensively through Part II of this thesis.

The wear identification method developed was carried out as a project for the System Identification course given by Prof. Lennart Ljung during the spring of

2010. The developed wear estimator is based on friction observations achieved from a test-cycle. A framework favourable to identification methods was adopted, with a known friction model, to reveal the challenges and restrictions of such methods for wear diagnosis. As it is shown, a careful experiment design can lead to a robust wear monitoring solution.

Paper C: A Data-driven Method for Monitoring Systems that Operate Repetitively - Applications to Robust Wear Monitoring in an Industrial Robot Joint

> A. C. Bittencourt, K. Saarinen, and S. Sander-Tavallaey. A data-driven method for monitoring systems that operate repetitively - applications to robust wear monitoring in an industrial robot joint. In *Proc. of the 8th IFAC SAFEPROCESS*, 2012. Under review.

Background. Surprisingly, no references were found in the literature that considered the diagnosis of systems that operate in a repetitive manner. This type of system is however quite common, e.g. in automation, or for any system from which a test-cycle can be repeated periodically. The repetitive execution of a system provides redundancies about the system's behavior which are directly found in the data. For example, it is possible to compare the result of the execution of a test-cycle performed today to how it is performed in a year. This comparison would allow to infer the system's deterioration over the period.

The methods were developed with the interest focused on diagnosis of industrial robots, where a repetitive operation is almost a requirement in most of its applications. The ideas behind the methods emerged via a combination of development and testing of methods in collaboration with ABB and new knowledge and insights from the Machine Learning area.

1.3.2   Relevant and Additional Publications

The author was introduced to the wear monitoring problem already in 2007 during a Master Thesis project taken at ABB:

> A. C. Bittencourt. *Friction change detection in industrial robot arms.* Msc. thesis, The Royal Institute of Technology, 2007.

In the contribution, a method for friction change detection was developed. The basic idea was to monitor the changes directly on the friction curves. A test-cycle was required in order to estimate the friction curve, in a similar way as in Paper B. The effects of load, lubricant and temperature were briefly investigated during the work and motivated the more thorough experiments of Paper A.

An early version of Paper A was presented in:

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

The version added in this thesis includes a more detailed analysis of the modeling assumptions, together with a more general framework for identification of static friction models.

A preliminary version of Paper C was presented in:

> A. C. Bittencourt, K. Saarinen, S. Sander-Tavallaey, and H. A. Andersson. A method for monitoring of systems that operate in a repetitive manner - application to wear monitoring of an industrial robot joint. In *Proc. of the 2011 PAPYRUS Workshop*, Corsica, France, Oct 2011b.

The version added in this thesis includes methods to reduce sensitivity to disturbances as well as validation with real data. Paper C also includes analyses of the achievable performance of the methods under temperature disturbances.

Part I

Background

2

Basics of Industrial Robotics

The International Organization for Standardization, ISO, proposes the following definitions in ISO 8373 (1994).

Definition 2.1 (ISO 8373:1994 No. 2.15 – Robotics).
    Robotics is the practice of designing, building and applying robots.

Definition 2.2 (ISO 8373:1994 No. 2.6 – Manipulating industrial robot).
    A manipulating industrial robot is an automatically controlled, re-programmable, multipurpose, manipulator programmable in three or more axes, which may be either fixed in place or mobile for use in industrial automation applications. Note: The robot includes the manipulator (including actuators) and the control system (hardware and software).

The above definitions make a clear distinction of *industrial* robots in the manner that they are used for, "industrial automation applications". The first industrial robot was operating in 1961 in a General Motors automobile factory in New Jersey. It was Devol and Engelberger's UNIMATE. It performed spot welding and extracted die castings, Wallén (2008). Since then, many new applications of industrial robots have been introduced, e.g. welding, cutting, forging, painting, assembly, etc. Industrial robots penetrated quite rapidly in manufacturing and specially in the automotive industry, which is still the largest robot user. In 2007, there were more than 1.000.000 industrial robots in operation worldwide. This number is expected to double only in electronics manufacturing in less than a decade, Yee and Jim (2011).

(a) An ABB IRB 6 from 1973.   *(b)* A modern ABB IRB 6640.

Figure 2.1: *The 5 axes robot IRB 6 was the first robot controlled by a microcomputer. The 6 axes robot IRB 6640 is a high payload general purpose manipulator.*

Industrial robots are a key factor to improve productivity, flexibility, quality and safety of technical systems. The history of industrial robotics development is filled with technological milestones. In 1971, the first all-electrically actuated robot was introduced by Cincinnati Millacron, whose robotics development team was later acquired by ABB in 1990. In 1973, ABB released IRB 6, the first microcomputer-controlled robot, which was also all-electrically actuated. Remarkably, this setting is still dominant in modern industrial robots, see Figure2.1.

The mechanical structure of a standard industrial robot is composed by *links* and *joints*. Links are the main bodies that make up the mechanism and the links are connected by joints to each other. A joint adds constraints to the relative motion of the connecting links and are categorized accordingly. The configuration of links and joints defines the *kinematic chain* of the robot. The number of joints defines the number of manipulated *degrees of freedom*, DOF, of a robot. The most common configuration of industrial robots is the 6 DOF with serial kinematics, meaning that links and joints are mounted serially. This type of robots is also known as "elbow" manipulators for its resemblance with the upper arm of a human. For elbow manipulators, the first three axes, also called *main axes*, are used to achieve a desired position of the end-effector. The links of the main axes are bigger since they drive more load compared to the last three, *wrist axes*, which are used to manipulate the orientation of the end-effector.

The main developments in industrial robotics have been directly connected to its main user, the automotive industry. This resulted in products with high cost efficiency, reliability and performance (Brogårdh (2007)). A cost-driven develop-

ment means the need of cost reduction of the components used. This leads to a more difficult control design to handle the larger variations in kinematic and dynamic parameters, lower mechanical resonance frequencies and larger nonlinearities. In order to meet the performance required from industrial robots, a broad understanding of the system is needed. In this chapter, the basics of industrial robotics is reviewed.

2.1   Actuators and Sensors

An industrial robot is a complete system that interacts with its surroundings. Its degree of autonomy is directly related to the sensory information available, the knowledge built in the system (e.g. models/learning), and the possibilities to perform actions. Following demands on cost efficiency and reliability, the amount and variety of sensors are remarkably small in typical applications of industrial robots. With the development of new applications and higher demands on autonomy, alternative sensors are becoming more common (Brogårdh (2009)).

2.1.1   Basic Setup

As mentioned in the beginning of this section, modern industrial robots are most commonly actuated with electrical motors. The permanent magnet synchronous motor, PMSM, is a popular choice due to its high power density, easy operation and performance. The output torque of such motor can be divided into two parts:

- an ideal mutual torque, the dominant part, arising from the interaction of the stator and rotor, and

- *torque ripple*, an angular dependent component arising from geometry imperfections which can be amplified by feedback.

The torque ripple leads to challenges in control of machines actuated with PMSM, see, e.g., Proca et al. (2003); Mohamed and El-Saadany (2008). Furthermore, the relation between applied current and output torque varies with temperature due to a reversible demagnetization of the magnets (Sebastian (1995)). A power amplifier is used to modulate the power used as input to the motors. In order to provide high torques and low speeds, a *gearbox transmission* is used at the motor output. The rotary vector (RV) type is a popular choice of compact gearboxes due to their low backlash, high gear ratio (in the order of $100 - 300$) and size. This type of transmissions is commonly found in the main axes of a manipulator. In the wrist axes, also harmonic drive gears are used as well as special gear solutions. See Figure 2.2 for examples of motor and gear units used in industrial robots.

Typically, only the rotation angle of the motor shaft, electrical quantities (voltage and current), and winding temperature are measured. Optical encoders and resolvers are the most commonly used sensors for the angular measurements. With these types of sensors, it is possible to compute the motion and its direction in relation to a reference position with high accuracy. The high accuracy allows
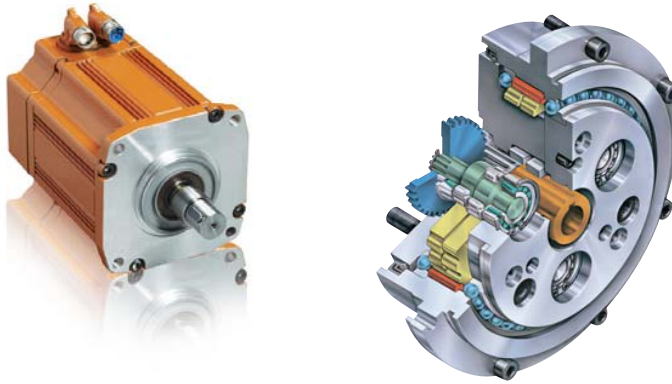
Figure 2.2: *An* ABB *motor (left) and a Nabtesco* RV *gear unit scheme (right, picture courtesy of Nabtesco.)*

for differentiation of the measured positions to provide estimates of speed and acceleration.

2.1.2   Application Dependent Sensors

With the basic sensors and refined models of the system, it is possible to achieve high path and positioning performances. This allows robots to be used in applications with a controlled/predictable environment. In more demanding applications, where the workpiece and environment are changing or where the robot is used in contact applications, the use of alternative sensors is required.

Six DOF force/torque sensors can be used in applications such as high precision assembly of drive trains. This type of sensor is also important in machining applications, such as grinding and polishing. The use of high speed cameras combined with image processing algorithms is also important in pick and place applications. Applications demanding very high accuracy might require the use of additional sensors on the arm side of the robot. Measurements of the arm variables help to reduce the influence of backlash and compliance of the gears on the accuracy of the robot. This can be achieved, e.g., with the use of encoders, torque sensors and inertial measurement units, IMU's, in the actuator transmissions and the arm system. For a review, see Brogårdh (2009); for an example on the use of IMU's to improve accuracy, see Axelsson et al. (2011).

Remark 2.1.   While the use of additional sensors can increase the robot autonomy, performance and safety, it also means higher cost and increased complexity of the system.

2.2   Modeling

Given the limited sensory information from the measurements of the angles of the motor shafts, the high demands on accuracy and performance expected from
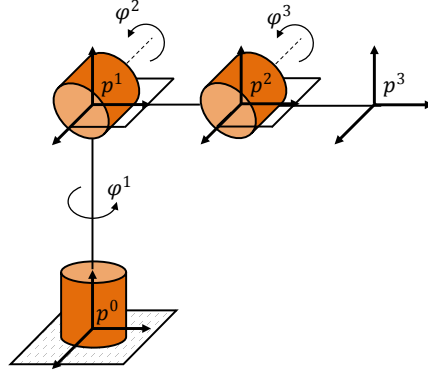
Figure 2.3: *Joint positions, φ^i, and coordinate frames, p^{i-1}, for an elbow manipulator with $i = [1, 2, 3]$ joints. The end-effector is fixed at frame p^3.*

industrial robots are only possible with the use of reliable models and model-based control (Brogårdh (2009)). Models are also important for design, simulation, diagnosis, etc. They play a significant role in all industrial robotics.

In this section, modeling of industrial manipulators is reviewed. The presentation follows standard textbooks, see, e.g., Spong et al. (2006); Sciavicco and Siciliano (2000).

2.2.1 Kinematics

The kinematics describes the motion without considering the forces and torques causing it. A kinematic model only depends on the geometric description of the robot. Let φ^i be the ith joint position at the arm side, and let us define by convention a frame p^{i-1} at each joint. For a configuration with n joints, there are $n + 1$ frames and the end-effector is considered fixed at frame p^n. See Figure 2.3 for an illustration.

By using a coordinate transformation, it is possible to describe a point attached to coordinate frame i in the coordinate frame $i-1$ by

$$p^{i-1} = R_i^{i-1} p^i + d_i^{i-1} \tag{2.1}$$

where R_i^{i-1} and d_i^{i-1} are a rotation and a translation from frame i to frame $i-1$ respectively. The above transformation can be written as a homogeneous transformation

$$P^{i-1} \triangleq \begin{bmatrix} p^{i-1} \\ 1 \end{bmatrix} = \underbrace{\begin{bmatrix} R_i^{i-1} & d_i^{i-1} \\ 0 & 1 \end{bmatrix}}_{\triangleq H_i^{i-1}} P^i, \tag{2.2}$$

which facilitates calculations since consecutive frame transformations simplify to multiplications of matrices. Notice that the homogeneous transformation H_i^{i-1} is

a function of φ^i and the links' geometry.

Forward Kinematics

The forward kinematics is the problem of finding the end-effector pose X (position and orientation) relative to the base frame given the joint variables φ. This can be achieved with the use of a homogeneous transformation from the tool pose to the base frame. For a configuration with n joints, the transformation is described as

$$P^0 = H_n^0(\varphi)P^n, \tag{2.3}$$

from which it is possible to extract the pose, X, of the end-effector. The Denavit-Hartenberg convention provides a manner to choose the reference frames that allows for a systematic analysis. For a serial robot, the direct kinematics always has a unique solution.

Taking the time derivative of the end effector pose, gives a relation between the joint velocities $\dot{\varphi}$ and the linear and angular velocities of the end-effector as

$$\dot{X} = \mathcal{J}(\varphi)\dot{\varphi}, \tag{2.4}$$

where $\mathcal{J}(\varphi)$ is known as the analytical Jacobian matrix. The accelerations can be found by taking the time derivative again, yielding

$$\ddot{X} = \mathcal{J}(\varphi)\ddot{\varphi} + \left(\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{J}(\varphi)\right)\dot{\varphi}. \tag{2.5}$$

The Jacobian matrix is a very important tool in robotics and it can be used to find singular configurations, transformation of tool forces to joint torques, etc.

Inverse Kinematics

The reverse problem, finding the joint positions φ given the end-effector pose X is known as the inverse kinematics. The inverse kinematics problem is important for trajectory generation, when a desired tool path needs to be transformed to joint positions. For the serial robot, it can be expressed as solving the nonlinear equations

$$H_1^0(\varphi^1)H_2^1(\varphi^2)\cdots H_n^{n-1}(\varphi^n) = H(X) \tag{2.6}$$

for a given right-hand side, where φ^i is ith joint position and H_i^{i-1} is given by (2.2). An analytical solution is not always possible, in which case a numerical solver must be used, and even if a solution exists it is typically not unique.

2.2.2   Dynamics

A dynamic model describes the relation between the robot motion and the forces and torques that cause it. Dynamic models are important for simulation, trajectory generation and control. In feed-forward control, the motor torques required to achieve a certain path are computed from the *inverse dynamics*.

The simplest modeling approach is to consider all links as rigid bodies. From this

simplification, there are different possible methods to derive rigid-body models. The Euler-Lagrange formulation considers the Lagragian equation

$$\mathcal{L}(\varphi, \dot{\varphi}) = \mathcal{K}(\varphi, \dot{\varphi}) - \mathcal{P}(\varphi), \tag{2.7}$$

where the Lagragian $\mathcal{L}(\varphi, \dot{\varphi})$ is defined as the difference between kinetic, $\mathcal{K}(\varphi, \dot{\varphi})$, and potential energies, $\mathcal{P}(\varphi)$. By writing the kinetic energy as a quadratic function $\mathcal{K}(\varphi, \dot{\varphi}) = \frac{1}{2}\dot{\varphi}^T M(\varphi)\dot{\varphi}$, where $M(\varphi)$ is the total inertia matrix, the equations of motion are given from the Euler-Lagrange equations

$$\frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial}{\partial \dot{\varphi}^i} \mathcal{L}(\varphi, \dot{\varphi}) - \frac{\partial}{\partial \varphi^i} \mathcal{L}(\varphi, \dot{\varphi}) = \tau^i, \quad \text{for } i = 1, \dots, n \tag{2.8}$$

where τ^i is the applied torque at joint i. By gathering gravitational terms of the form $\tau_g^i(\varphi) = \frac{\partial}{\partial \varphi^i}\mathcal{P}(\varphi)$ into the vector $\tau_g(\varphi) = [\tau_g^1, \cdots, \tau_g^n]^T$ and terms involving $(\dot{\varphi}^i)^2$ and cross-products of $\dot{\varphi}^i \dot{\varphi}^j$ in $C(\varphi, \dot{\varphi})$, the resulting multi-body rigid model is of the form

$$M(\varphi)\ddot{\varphi} + C(\varphi, \dot{\varphi}) + \tau_g(\varphi) = \tau \tag{2.9}$$

where τ is the vector of applied torques. This model can be extended by including a dissipative friction term, τ_f, which is typically modeled as a nonlinear function of $\dot{\varphi}$, see Chapter 3 for more on friction.

Including Flexibilities

In most cases when modeling robots, a rigid-body model is not sufficient to describe the system in a realistic manner. The approximation of a rigid gear-box is specially unrealistic for compact gearboxes. Also, with a trend of lighter robots, the flexibilities of bearings- and links are also becoming significant. The model for a flexible robot structure can, as a first approximation, be described by lumped masses connected by springs and dampers.

For instance, including one flexibility for each joint gives a 2-masses system connected by a torsional spring-damper for each joint as shown in Figure 2.4. Neglecting possible inertial couplings between motor and arm[i], the resulting model can be described as

$$M_a(\varphi_a)\ddot{\varphi}_a + C(\varphi_a, \dot{\varphi}_a) + \tau_g(\varphi_a) + \tau_{f,a}(\dot{\varphi}_a) = \tau_a \tag{2.10}$$

$$\tau_a = K(\eta \varphi_m - \varphi_a) + D(\eta \dot{\varphi}_m - \dot{\varphi}_a) \tag{2.11}$$

$$\tau_m - \eta \tau_a = M_m \ddot{\varphi}_m + \tau_{f,m}(\dot{\varphi}_m) \tag{2.12}$$

where the subscripts $_a$ and $_m$, relate to variables at the arm motor sides respectively, η is the inverse gear ratio matrix, $K(\cdot)$ and $D(\cdot)$ are the stiffness and damping matrices. The friction torque is here divided between the motor and arm side, $\tau_{f,m}(\dot{\varphi}_m)$ and $\tau_{f,a}(\dot{\varphi}_a)$ respectively. The friction occurs at different components in the gearbox, at different gear ratios, meaning different reductions when seen at the motor side. See, e.g. Moberg (2010), for a detailed treatment on modeling of flexible robots.

---

[i]According to Spong (1987) this is a reasonable approximation if the transmission ratio is large.
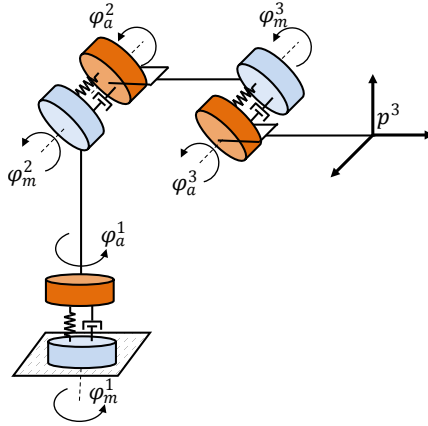
Figure 2.4: *Illustration of a flexible robot structure, where the flexibilities are modeled as lumped masses connected by springs and dampers.*

2.3   Identification

The described models depend on a number of parameters that are most often unknown or partly known. In order to make use of models, e.g. for control and simulation, the modeling process must be complemented with identification procedures. Identification is used to find and verify the parametric description of the models from experiments. As introduced in the previous section, the different models can relate to: *kinematics*, *dynamics* and *joint*-related phenomena. A summary of these identification problems is given below.

Kinematic models. An accurate dynamic models is important for positioning of the end-effector. The parameters in the model relate to the geometric description of the kinematic chain. These parameters can be partly obtained during the design process, e.g. available from CAD models. There are however errors that could relate, amongst other sources, to tolerances during production and assembly of the robot. An identification procedure can be used to correct for these errors, considerably improving the volumetric accuracy of the robot. The process of identifying these parameters is also known as *kinematic calibration* or *robot calibration*, and requires measurements of the end-effector position. For a survey on the topic, see Hollerbach (1989).

Dynamic models are important for simulation and feed-forward motion control of robots. The identification of dynamic models of robots is a much studied problem and several approaches can be found, see Wu et al. (2010) for an overview. An important consideration is the type of dynamic model considered. Rigid-body models are typically parametrized as a function which is linear in the parameters. For example, the model in (2.9) can be rewritten as a linear regression

$$u(k) = \phi^T(k, \varphi, \dot{\varphi}, \ddot{\varphi})\theta, \tag{2.13}$$

where k is the sample index, $u(k)$ are the applied torques, $\phi(\cdot)$ is a regressor vector function, dependent on φ and its derivatives, and θ are the rigid-body parameters. Based on data from an identification experiment, the parameters θ can be found, e.g., based on a weighted least squares minimization

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} \left(\mathbf{u} - \Phi^T \theta\right)^T W \left(\mathbf{u} - \Phi^T \theta\right) = \left(\Phi^T W \Phi\right)^{-1} \Phi^T W \mathbf{u}, \qquad (2.14)$$

where $\mathbf{u}$ and Φ^T are the stacked input and regressors data achieved from the identification experiment. The choice of weight matrix W will affect the solution and different criteria are possible, see, e.g., Gautier and Poignet (2001); Swevers et al. (1997). Finally, the trajectory must be chosen carefully to avoid excitation of flexible modes and improve the estimation performance, see, e.g., Wernholt and Moberg (2011); Gautier and Poignet (2001); Swevers et al. (1997). Identification of parameters describing the flexibilities is a more involving problem since only a subset of the states can be measured and a linear regression cannot be formed. These models are however important for improved performance of robot control. For a detailed treatment on identification of dynamic models and flexibilities, see Wernholt (2007); Moberg (2010); Wernholt and Moberg (2011).

Joint models. Due to the complex construction of a robot joint, its characteristics are often uncertain and introduces diverse nonlinearities to the system. Nonlinearities that can be of significant influence in a robot joint are related to friction, backlash and nonlinear stiffness. Available parametric models are often achieved from empirical observations on a specific platform since it is difficult to predict the characteristics of these nonlinearities. For example, the amount of backlash and friction will depend on how the joints are assembled. Therefore, these models must be found invariably from an identification procedure. It is important to notice that identification of, e.g., a dynamic model is facilitated if an accurate joint model is available. For example, in Wernholt (2007) it is reported that the friction at low speeds makes it difficult to identify the resonances related to a flexibility. This is because friction adds damping to the system. With a known friction model, its effects can be analytically removed from the data, making the identification of dynamic parameters more reliable. See, e.g., Wernholt and Gunnarsson (2006) where a three step identification procedure is suggested.

2.4   Reference Generation and Control

From the perspective of a robot user, it is convenient to be able to program the robot in a high level of abstraction. Typically, objectives can be defined in the task space, and the user does not need to worry about how each joint is controlled. A robot manufacturer dependent programming language is used where instructions to the robot can be given in task (or joint) space. This can be done manually by typing the code or in some cases by demonstration. This process can also be partly automated with the use of CAD/CAM softwares allowing greater flexibility. An example of a robot task program is given in Algorithm 1. In order to perform a task, different problems must be solved.

Algorithm 1 My spot-welding task.

Move to point A_0 as fast as possible.
Approach point A_1 slowly.
Perform spot weld.
Move to point B_0 as fast as possible.
. . .

Motion planing. First, given a task, e.g. the one defined in Algorithm 1, a path to be executed by the robot must be generated. This is made by a *motion planner*, which calculates the movements that the robot must make. At first, the programmed movements are interpreted with respect to what geometry that the path will have (line, circle, spline etc.) and then the path is interpolated to consist of discrete steps, which are transformed from task space to joint space using the inverse kinematic model.

Trajectory generation. The time dependence of the robot movements, i.e. a trajectory, can be calculated either in the task space or in the joint space. Finding a trajectory involves optimization of the use of the dynamic capabilities of the robot with respect to speed- and acceleration performance. Let $\mho$ denote a trajectory, the trajectory generation is basically an optimization problem including,

$$\mho^r = \underset{\mho}{\operatorname{argmin}} \quad \text{Objective}(\mho)$$
$$\text{subject to} \quad \text{Kinematics}(\mho)$$
$$\text{Dynamics}(\mho)$$
$$\text{Mechanical stress limitations}(\mho)$$

where the solution, $\mho^r$, is used in the next stage as a reference for the motion control. The objective can be, e.g., minimal cycle-time or minimal energy. The constraints involve, of course, the kinematics and dynamics of the manipulator as well as knowledge of the mechanical limitations such as joint ranges, motor speeds, maximum allowed forces in the joints, etc. Notice that the solution for this optimization problem can considerably affect the time and performance of the task execution and is highly dependent on the models used. For example, in Ardeshiri et al. (2011) the inclusion of speed dependent constraints in a convex formulation of the problem allowed for reductions of the path tracking time by $5-20\%$. Speed-dependent constraints are motivated from physical modeling of the motors and the drive system, they can, e.g., relate to viscous friction.

Motion Control. Finally, when the reference trajectory is generated, it is possible to execute the task with the help of the servo control. Important features of the servo are trajectory tracking, robustness and disturbance rejection. Different control strategies and structures are possible depending on the sensors available, controlled variables, etc., see Moberg (2010); Brogårdh (2009) and available textbooks for details. Here, a common control approach is discussed for the typical setup, with measurements only at the motor side.
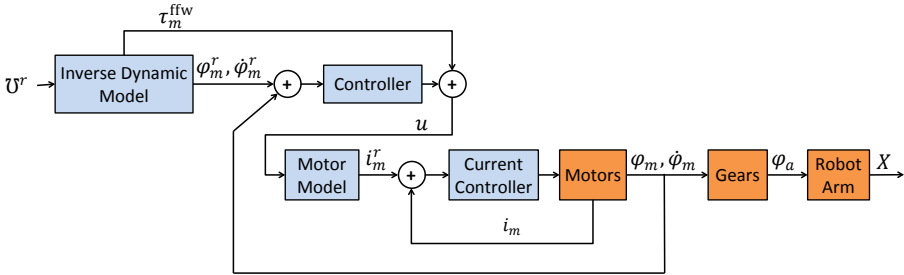
Figure 2.5: *A model-based control scheme for trajectory tracking. A feed-forward action τ_m^{ffw} and motor references $\varphi_m^r, \dot{\varphi}_m^r$ for the outer feedback loop are computed based on the reference trajectory $\mho^r$ using an inverse model. An inner control loop is used to control the motor current according to i_m^r which is achieved from a desired torque u using a motor model.*

2.4.1   Model-based Control for Trajectory Tracking

An overview of one possible robot control scheme can be seen in Figure 2.5. The desired trajectory $\mho^r$ contains the joint information through time at the arm side, that is, φ_a^r and its derivatives. With measurements only available at the motor side, $\varphi_m, \dot{\varphi}_m$, the arm side references are transformed to the motor side, yielding $\varphi_m^r, \dot{\varphi}_m^r$. For a rigid joint, this will depend only on the gearboxes ratios.

To improve performance, an inverse dynamic model is used to generate feed-forward motor torques, τ_m^{ffw}. The input u is the total torque the motor should generate to drive the robot in the desired manner and is composed of both feed-forward and feedback actions. Since the motor torque is not measured, a motor model is used to transform u to the current reference, i_m^r, for the inner current control loop. The motor variables $\varphi_m, \dot{\varphi}_m$ are fed back to the outer control loop. At the output is the end-effector pose X.

The inner current control loop has much faster dynamics than the outer loop. It is therefore common to accept a constant relation between the measured motor currents and the motor torques, that is $u = \tau_m = K i_m$. As pointed out in Section 2.1.1, this static relation actually depends on temperature since the nominal performance of the motors degrades with increased temperature.

2.5   Summary and Connections

This chapter provided an overview of important aspects to consider when working with industrial robots. The purpose was to provide an introduction to the technologies and their limitations. Two aspects are particular about the development of industrial robots, the limited sensory information available and consequently the importance of using different types of robot models.

The purpose of this introduction has also been to provide a background to the

research described in this thesis. In Paper C, the 2 axes flexible model described in Moberg et al. (2008) and Axelsson et al. (2011) was used to simulate friction faults in a robot joint with the friction models proposed in Papers A and B. Such simulation studies allowed for detailed analysis of the proposed methods which otherwise would have been too expensive and time consuming to perform. The assumption of an ideal current loop, giving $\tau_m = K i_m$, is also important for the identification of friction levels in Paper A and in any diagnosis methods that rely on torque estimates. However, since the identification experiments for friction estimation were made at constant speeds it would have been possible to perform most of the experiments even if the current control had not been much faster than the speed control. It is nevertheless important that K is not speed dependent (or that the speed dependence is known) and that the temperature dependence of K will not considerably disturb the friction identification and modeling.

3

Joint Friction and Wear

Friction exists in all mechanisms to some extent. It can be defined as the tangential reaction force between two surfaces in contact. There are different types of friction, e.g. dry friction, viscous friction, lubricated friction, skin friction, internal friction. Friction is not a fundamental force but the result of complex interactions between contacting surfaces in down to a nanoscale perspective. Due to its complex nature, it is difficult to described it from physical principles. Tribology, the science of interacting surfaces in relative motion, is therefore mostly based on empirical studies.

One reason for the interest in friction in manipulator joints is the need to model friction for control purposes. A precise friction model can considerably improve the overall performance of a manipulator with respect to accuracy and control stability, see, e.g., Kim et al. (2009); Guo et al. (2008); Olsson et al. (1998); Bona and Indri (2005); Susanto et al. (2008). Since friction can relate to the wear down process of mechanical systems (Blau (2009)), including robot joints, there is also interest in friction modeling for robot condition monitoring and fault detection, see, e.g., Caccavale et al. (2009); Namvar and Aghili (2009); McIntyre et al. (2005); Vemuri and Polycarpou (2004); Brambilla et al. (2008); Mattone and Luca (2009); Freyermuth (1991).

In a robot joint, with several components interacting such as gears, bearings, and shafts, which are rotating/sliding at different velocities and under different lubrication levels, it is difficult to separate and model friction at a component level. A typical approach is to consider these effects collectively, as a "lumped" *joint friction*. For examples of friction models at a component level, see SKF (2011).

Friction always opposes motion, converting kinetic energy into heat. Another outcome of friction is wear. Wear is defined as "the progressive loss of material
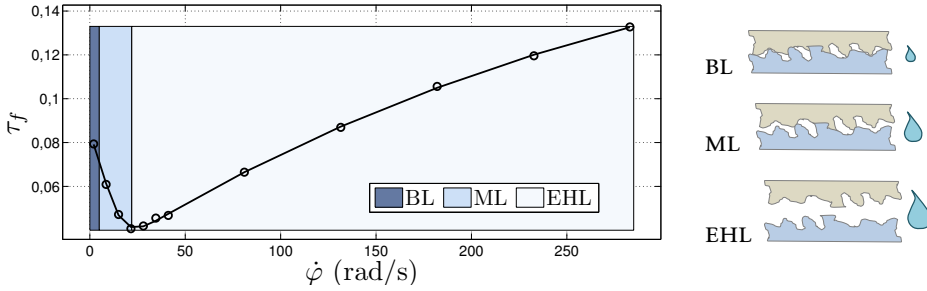
Figure 3.1: *Friction curve for constant speed movements and the lubrication regimes illustrated at contact level.*

from the operating surface of a body occurring as a result of relative motion at its surface" (Lansdown et al. (1987)). The need for relative motion between surfaces implies that the wear mechanisms are related to mechanical action between surfaces. This is an important distinction to other processes with a similar outcome and very different nature, e.g. corrosion.

3.1   Basics of Tribology

The most important friction characteristics for control applications are usually described by a so-called *friction curve*, which is a plot of friction levels as function of speed[i]. An example of such plot achieved from experiments in a robot joint can be seen in Figure 3.1[ii,iii]. The nonlinear behavior from low to high speeds is typical in lubricated friction and is known as the Stribeck effect. This phenomenon is considered to have been first observed by Stribeck in 1902 (Jacobson (2003); Woydt and Wäsche (2010)). This behavior is present in a robot joint due to the presence of lubricant in the gearboxes and motor shaft. Notice that the friction in the motor is dry. The use of lubricant is essential to decrease the wear processes. It acts as a separation layer between the surfaces. With the use of additives, it can even create a chemical barrier between the contact surfaces under high contact pressure, reducing low speed friction and wear.

The friction curve is divided in three regions according to the lubrication regime: boundary lubrication (BL), mixed lubrication (ML) and elasto-hydrodynamic lubrication (EHL). The phenomenon present at very low speeds (BL) is mostly related to interactions between the asperities of the surfaces in contact. With the increase of velocity, there is a consequent increase of the lubricant layer between

---

[i]In fact, as presented originally in Stribeck (1902), a friction curve is plotted as a function of speed normalized by the ratio of load pressure and lubricant viscosity. For simplicity however, it is many times shown only as a function of speed.

[ii]In the figure, the friction torques are normalized to the maximum allowed torque to the joint and are displayed as a dimensionless quantity, this convention is followed in the whole thesis.

[iii]This type of curve is obtained when the speed levels are stable and include no transient phenomena. There are also dynamic effects related to friction, see Section 3.3.

the surfaces with a decrease of contact friction (ML). The decrease of contact friction continues until it reaches a full lubrication profile (EHL), with a total separation of the surfaces by the lubricant. In EHL, friction is proportional to the force needed to shear the lubrication layer, and it is thus dependent on the lubricant properties (e.g. viscosity).

The wear processes are most significant in BL and ML, where contact friction is significant. In a full-film lubrication, there is theoretically no wear taking place, but it still happens because of eventual breakdown of this layer. It is important to notice that due to the high gear ratio of the gearboxes used in industrial robots, the components closer to the output will be moving slower in comparison to the ones closer to the input. Therefore, at a component level, wear might occur even in the EHL region of the joint friction curve.

3.2   Friction Dependencies in Robot joints

At a contact level, friction is dependent on the contact geometry, topology, properties of the materials, relative velocity, lubricant, etc. (Al-Bender and Swevers (2008)). Depending on the setup, each of these factors will be more or less significant to the total friction. In a robot joint, it is known that friction can relate to

- temperature,
- force/torque levels,
- position,
- velocity,
- acceleration,
- lubricant properties.

These dependencies will significantly differ depending on the size and type of joint considered. Since the main axes undertake significant load levels, the wear processes in these axes are usually more significant than in the wrist axes. Therefore, the focus in this thesis is on the study of friction in the main axes. Bigger robot types, as the ones studied in this thesis, are today typically equipped with RV gearboxes, recall Section 2.1.1. Figure 3.2 presents a summary of the effects of relevant variables to the friction curves of robot joints. It should be noted that, except for Figures 3.2a and 3.2b, the curves are obtained for different robots. The effects are summarized below.

Load. The effects of *load* follow from the consequent increase of contact pressure between the surfaces in contact, leading to a generalized increase of the friction curve, with a more significant increase at very low speeds, i.e. in the BL regime.

Lubricant. In lubricated mechanisms, both the thickness of the *lubricant* layer and its viscosity play an important role in the resulting friction properties. The oil lubricants 1 to 3 for the curves shown in Figure3.2c have an increasing viscosity. The higher viscosity leads to higher shear forces and therefore higher friction levels in the EHL regime.

Temperature. The viscosity is also dependent on the *temperature* of the lubricant (Seeton (2006)), the higher the temperature, the lower the viscosity. This can be observed in Figure 3.2b with a decrease of friction in the EHL regime at higher
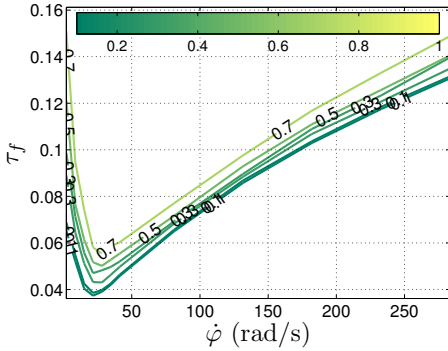
(a) *Normalized load torques. Increases in the BL regime and a bias-like increase.*
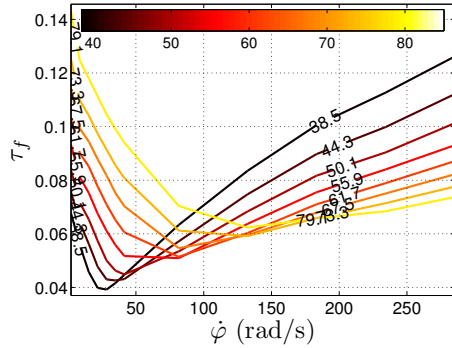
(b) *Temperature (C°). Significant increases in the BL and ML regions and decreases in the EHL region.*
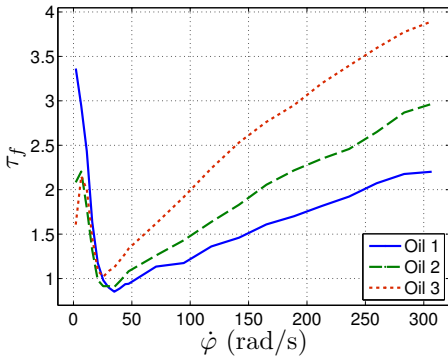
(c) *Gearbox lubricant. Decreases in the EHL region with viscosity.*
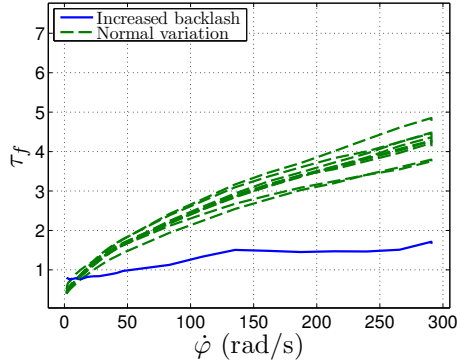
(d) *Backlash. Decreases in the ML and EHL regions.*
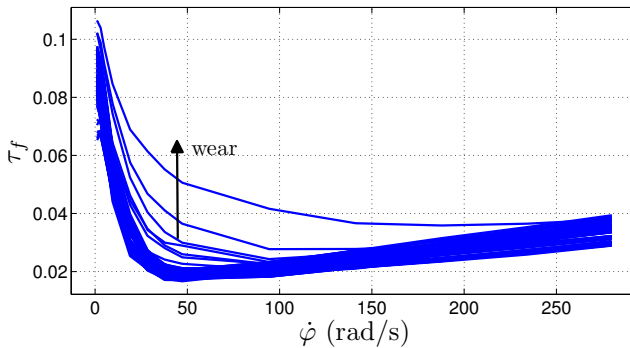
(e) *Wear. Increases concentrated in the ML region.*

Figure 3.2: *Effects of different factors to the friction curve. The effects of backlash are shown for different robot individuals of the same family. The wear effects are achieved from accelerated wear tests.*

temperatures. The effects of temperature are however more complex, changing also the BL and ML regimes. A possible explanation is that temperature also considerably affects the interaction forces of the surfaces in contact. This could be caused, e.g., by assymetric dilations of the gearbox components.

Wear. The increase of friction with *wear* as seen in Figure 3.2e is related to the accumulation of wear debris in the circulant lubricant, which increases friction specially in the ML regime.

Backlash. The decrease of friction with *backlash* seen in Figure 3.2d can possibly be explained by a consequent loosening of the gearbox components, yielding lower contact pressures. Notice that backlash might follow from a degenerate wear process, where the amount of material removed by wear starts to be significant enough to create undesired clearances between the surfaces.

3.3 Modeling

Due to the complex nature of friction in a robot joint, it is common to accept models based on empirical observations of the phenomena. The history of the development of empirical friction models is extensive, see Dowson (1998) for a historical perspective. At a contact level, the surfaces' asperities can be compared to bristles on a brush. Each of these (stiff) bristles can be seen as a body with its own dynamics which are connected by a similar bulk. Different models have been proposed to model this dynamic behavior of friction, and some examples are presented in Olsson et al. (1998); Al-Bender and Swevers (2008); Harnoy et al. (2008); Åström and Canudas-de Wit (2008). A typical approach is to consider all the dynamics into a single state (Dupont et al. (2002)).

A dynamic friction model commonly found in robotics is the LuGre model (Åström and Canudas-de Wit (2008)). For a revolute joint, the friction torque is given by the LuGre model as

$$\tau_f = \sigma_0 z + \sigma_1 \dot{z} + h(\dot{\varphi}) \tag{3.1a}$$

$$\dot{z} = \dot{\varphi} - \sigma_0 \frac{|\dot{\varphi}|}{g(\dot{\varphi})} z, \tag{3.1b}$$

where the state z captures the average dynamic behavior of the asperities. It can be interpreted as their average deflection, with stiffness σ_0 and damping σ_1.

Since z is not measurable, it is difficult to estimate the parameters describing the dynamic behavior of friction, i.e. $[\sigma_0, \sigma_1]$. In practice, it is common to accept only a static description of (3.1). For constant velocities, (3.1) is equivalent to the *static* model:

$$\tau_f(\dot{\varphi}) = g(\dot{\varphi})\text{sign}(\dot{\varphi}) + h(\dot{\varphi}) \tag{3.2}$$

which is fully described by the g- and h functions. In fact, (3.1) simply adds dynamics to (3.2).

The function $h(\dot{\varphi})$ represents friction in the EHL regime, where friction has a
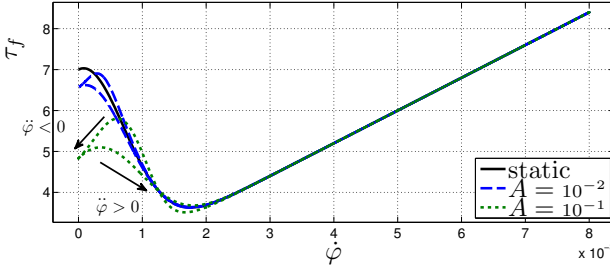
Figure 3.3: *Simulation of a LuGre model under different acceleration levels A and the related static friction model. The parameters are chosen for illustration purposes with static parameters $[F_c, F_s, F_v, \dot{\varphi}_s, \alpha] = [2, 5, 8\,10^2, 1\,10^{-3}, 2]$ and dynamic parameters $[\sigma_0, \sigma_1] = [1.4\,10^6, 2.42\,10^3]$.*

velocity strengthening behavior. For Newtonian fluids this behavior is directly proportional to speed, yielding the relationship

$$h(\dot{\varphi}) = F_v \dot{\varphi} \tag{3.3}$$

for the viscous behavior of friction. The function $g(\dot{\varphi})$ captures the BL and ML regimes, where friction has a *velocity weakening* behavior. Motivated by the observations mainly attributed to Stribeck (Jacobson (2003); Woydt and Wäsche (2010); Bo and Pavelescu (1982)), $g(\dot{\varphi})$ is usually modeled as

$$g(\dot{\varphi}) = F_c + F_s e^{-\left|\frac{\dot{\varphi}}{\dot{\varphi}_s}\right|^\alpha}, \tag{3.4}$$

where F_c is the Coulomb friction, F_s is defined as the standstill friction parameter[iv], $\dot{\varphi}_s$ is the Stribeck velocity, and α is the exponent of the Stribeck nonlinearity. The resulting static fricion model is given by

$$\tau_f(\dot{\varphi}) = \left[F_c + F_s e^{-\left|\frac{\dot{\varphi}}{\dot{\varphi}_s}\right|^\alpha}\right]\text{sign}(\dot{\varphi}) + F_v \dot{\varphi}. \tag{3.5}$$

which can describe many of the friction characteristics with speed. Notice that the $g-$ and h functions can be chosen differently.

Figure 3.3 shows the response of the LuGre model and the corresponding static model with $g-$ and h chosen according to (3.4) and (3.3). The simulation was performed with $\dot{\varphi}$ as half a period of a triangular wave with different slopes A. When accelerating, the transition from BL to EHL gives less friction torques than in deacceleration. The higher A, the more pronounced are the dynamic effects.

[iv] F_s is commonly called static friction parameter. An alternative nomenclature was adopted to make a distinction between the dynamic/static friction phenomena.

3.4   Summary and Connections

This chapter presented an overview of friction and wear from both empirical and phenomenological perspectives. The summary of the effects of different factors to the friction curves in Figure 3.2 gives a good idea behind the motivation and challenges of this work. The effects of temperature, load and wear in the figure are in comparable scales. Therefore, monitoring friction to perform wear diagnosis is a challenging problem. Load and temperature changes will always be present in applications and the diagnosis solutions must be able to cope with them.

The models presented in this chapter are only dependent on speed (and z). In Papers A and B, static friction models are proposed to describe the effects of load, temperature and wear. These models are found to be of utmost importance in the design of wear diagnosis solutions, where it is costly and time consuming to perform experiments. They are also used to design and validate solutions for wear diagnosis in industrial robot joints in Papers B and C.

4

Basics of Diagnosis

Diagnosis is a multidisciplinary science. Due to its relevance in different fields of applications, much can be found in the literature that relates to the diagnosis of systems, some examples are Reiter (1987); Rao (1998); Gertler (1998); de Kleer and Williams (1992); Basseville and Nikiforov (1993); Gustafsson (2000); Mobley (2002); Isermann (2006); Bishop (2007); Ding (2008). The notation and terminology can differ considerably across the different fields. Therefore, it is difficult to find a common framework for the presentation and categorization of the different methods and solutions. An important contribution towards a common framework is presented in an special issue of the IEEE Transactions on Systems, Man and Cybernetics in 2004 (Biswas et al. (2004)). In the issue, two research communities are identified that relate to diagnosis of systems:

The Fault Detection and Isolation (FDI) community, with origins closer to control theory and statistical decision making, and

The Diagnosis (DX) community, with origins in computer science and artificial intelligence.

However, these are not the only communities interested in diagnosis. Another community has origins in Mechanical Engineering, where the topic is known as *condition monitoring*. Also, in the Machine Learning community, diagnosis relates to *classification* problems.

It is outside of the scope of this thesis to provide an active discussion of the different diagnosis solutions and terminology; for a detailed review, see Venkatasubramanian et al. (2003c,a,b). Instead, the presentation here aims to familiarize the reader with the problems and contextualize the methods developed in this work. The presentation is mainly based on Nyberg (1999); Gustafsson (2000);
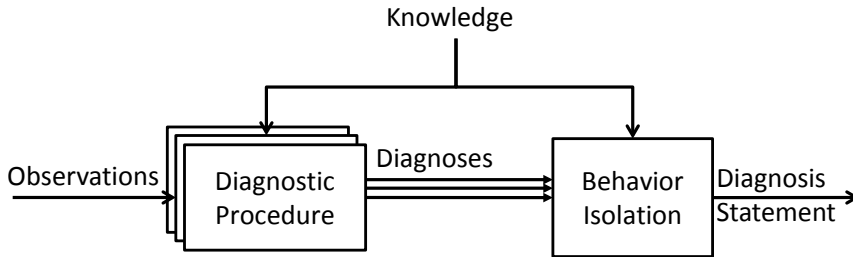
Figure 4.1: *Overview of the diagnosis process. Each diagnostic procedure produces one or more diagnosis based on the observations and available knowledge. The behavior isolation outputs a diagnosis statement, which is consistent to all knowledge and observations available.*

Van Trees (2001); Isermann (2006). Because it is difficult to contextualize the proposed methods in an available framework, the terminology and notation used are defined throughout the text.

4.1   Overview

The following terminology is defined for the diagnosis process.

Definition 4.1 (Diagnostic Procedure).
> An inference from observations and knowledge to a *diagnosis*.
>
> NOTE   A *diagnostic test* is a diagnostic procedure that can be performed automatically, with no human involvement.

Definition 4.2 (Diagnosis).
> A statement about the possible behavioral modes present given all or parts of the available knowledge and observations.

Definition 4.3 (Behavioral Mode).
> A state representing a possible behavior of an object being diagnosed.
>
> NOTE   A behavioral mode might be, e.g., healthy, faulty, broken, unknown. A behavioral mode related to a fault is a *fault mode*.

Definition 4.4 (Behavior Isolation).
> Determination of which behavioral modes are plausible given the observations and knowledge. Generates a *diagnosis statement*.

Definition 4.5 (Diagnosis statement).

The result of the diagnosis process.

NOTE   This is a statement regarding the behavioral modes which are consistent to all knowledge and observations available.

An overview of the diagnosis process can be seen in Figure 4.1. The objective is to decide the plausible behavioral mode(s) of the object(s) being diagnosed by considering the available knowledge and observations. The object under diagnosis can be, e.g., a component or a system, although this work is only interested in diagnosis of systems.

The diagnosis process makes extensive use of knowledge about the system under diagnosis. This knowledge might come from models, assumptions, data, an operator, an expert, etc. The available knowledge is used to provide means to infer the behavioral modes which are consistent to the observations, that can explain the observations. Each inference mechanism is called a *diagnostic procedure* and is based on parts (or the whole) of the available knowledge and observations.

The design and choice of diagnostic procedures are critical for the diagnosis process. For instance, they must be chosen as to allow certain isolation and robustness requirements. Diagnosis of complex systems will typically be composed of several diagnostic procedures, each of which generates at least one *diagnosis*. Based on all diagnoses, a *diagnosis statement* is provided which is consistent to all knowledge and observations available.

When a diagnosis solution is performed with no perturbation of the system's functions, during its normal operation, it is called an *on-line* diagnosis solution. If the diagnosis method requires the system to operate outside the scope of its functions, reducing its availability, the diagnosis is called *off-line*, as "off-the-line". When the diagnosis is performed by actively exciting the system, it is called *active*. If it is performed by passively studying the system, it is called *passive*. A diagnosis which is performed at each new observation (e.g. at each data sample) is a *real-time* diagnosis solution, otherwise it is a *batch* solution.

── Example 4.1: An off-line passive method for wear monitoring ──

The wear processes inside a robot joint cause an eventual increase of wear debris in the lubricant. Monitoring the iron content of lubricant samples taken from the robot joint can thus be used as an indication of the joint condition. The study of wear debris particles is known as *ferrography* and was first introduced by Seifert and Westcott (1972). Since then, the science has evolved and helped to understand wear related phenomena, Roylance (2005). In Figure 4.2, different types of wear particles are shown from ferrography studies.

In most applications however, the collection of lubricant samples can only be performed when the system is turned off, in an *off-line* manner, followed by laboratory analyzes. Notice that no dedicated excitation of the system is needed, so this is also a *passive* method.

(a) Spherical.   *(b) Laminar.*   *(c) Cutting.*
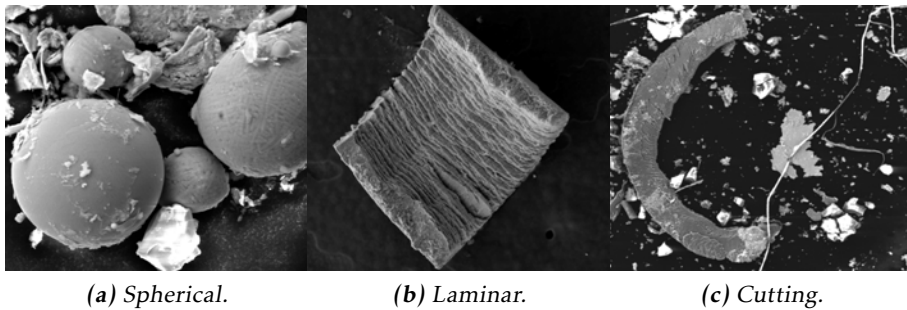
Figure 4.2: *Images of different types of wear particles from ferrography studies. The condition of the mechanical system may be determined from analyses of the characteristics of the wear particles, e.g. the type, shape, frequency, etc. (Pictures extracted from Machalíková et al. (2010)).*

The main interest in this work is on diagnosis solutions to *support for maintenance actions* and that can be *performed automatically*, i.e. based on *diagnostic tests*. It is considered that no detailed information about the fault type is necessary, only whether there is a fault present or not. This is motivated since in order to support maintenance of an industrial robot joint based on wear diagnosis, it is only necessary to determine the severity of the wear levels. A precise determination of which component in the joint is undergoing wear, e.g. whether a bearing or shaft, is not needed. It is important, though, that a faulty state is detected in an early stage, so that appropriate maintenance actions can take place before a critical degradation. In such cases, there are only few different behavioral modes and the isolation task is simple. Therefore, the discussion for the remaining part of this chapter is focused on the *design and evaluation of diagnostic tests*. For details on fault isolation, see, e.g., Nyberg (1999, 2011); Krysander and Frisk (2008); Basseville (2001).

4.2   Diagnostic Tests

To automate a diagnostic procedure, the available knowledge must be transformed into quantities that can be interpreted by a computer. Consequently, the observations are represented as *data*. A diagnosis algorithm, i.e. a diagnostic test, must be designed such that it processes the data and provides a diagnosis which is consistent with the available knowledge. This knowledge will therefore be built in the diagnostic test and can take different forms, e.g., models, assumptions or even as data sequences whose behavioral modes are known.

Three functions are identified for a diagnostic test and are defined below, see also Figure 4.3. Notice that observations are now written as data.
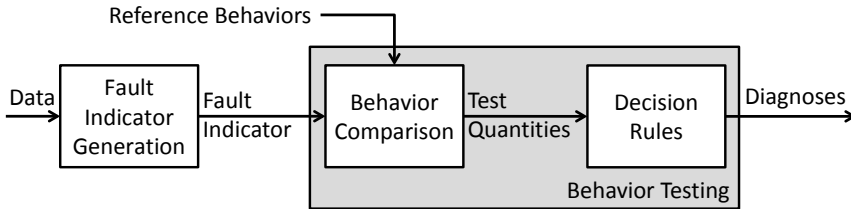
Figure 4.3: Overview of a diagnostic test, i.e. a diagnostic procedure which can be performed automatically (recall Figure 4.1). For a diagnostic test, observations are denoted data. Based on the data, a fault indicator, i.e. a quantity sensitive to a fault, is generated. The behavior of the fault indicator is compared to (known) reference behaviors. The result of the comparison provides one or more test quantities from which a decision about the behaviors present is made, i.e. diagnoses are generated.

Definition 4.6 (Fault Indicator Generation).
> Given data, provides a quantity which can be affected by a fault mode. Generates a *fault indicator*.

Definition 4.7 (Behavior Comparison).
> Compares the current behavior of the fault indicator with *reference behaviors*. Generates *test quantities*.

Definition 4.8 (Decision Rule).
> Makes a decision about the reference behavioral modes present. Generates *a diagnosis*.

The fault indicator is essentially an algorithm designed to provide a quantity sensitive to a fault mode, a fault indicator. The behavior of this indicator will vary according to the behavioral mode present in the system. With *knowledge* of the behavior of the fault indicator under certain *reference modes*, a *behavior comparison* can be made to provide *test quantities*. A *decision rule* is applied *to each test quantity* to provide a diagnosis. The combined tasks of behavior comparison and decision rule are called *behavior testing*. This is because they are often designed jointly. In fact, many times there is only a subtle distinction between the different tasks of a diagnostic test. For example, a *classifier* makes a direct map from data to a diagnosis, Bishop (2007).

The next section presents an important type of knowledge representation, models of systems and faults. In general, the extent to which a system model is known can considerably affect the type of diagnosis solutions and the amount of extra knowledge required for the diagnostic test, see e.g. Svärd and Nyberg (2011) where an automated diagnosis design is presented. The discussion follows with a more detailed presentation of each function in a diagnostic test and its evalua-
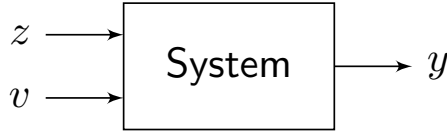
Figure 4.4: *A system is treated as a general mechanism generating data under the effects of deterministic and random inputs, z and v respectively. The data are the measured output y and the known components of the inputs z.*

tion.

4.3   Models of Systems and Faults

In order to choose the diagnosis algorithm, it is important to understand the behavior of the system and its dependencies on the faults. This can be achieved with the use of models. In this work, the approach is to consider a system as a mechanism that generates data from which a diagnosis can be performed.

Consider the representation in Figure 4.4, where the system represents the mechanism generating measured output data y, which is the result of deterministic inputs z, and random inputs v (e.g. noise). The inputs v are assumed unknown, while z could have known and unknown components. The set of deterministic inputs z, is further categorized in three distinct groups, u, d and f. The sequence u is known (e.g. control inputs), d is unknown and uninteresting (e.g. disturbances), and f is unknown and of interest (e.g. faults). The known signals y and u are of course important for the design of diagnosis solutions. It is also important to know the effects of v, d and f to the observed data y and u, so that the different effects can be identified correctly.

In general terms, a *system model* is a map from z and v to y,

$$y = g(z, v). \tag{4.1}$$

When this map is parametrized with parameters θ, this map is called a *model structure*, $\mathcal{M}$,

$$\mathcal{M}: \quad y = g(z, v, \theta). \tag{4.2}$$

A particular parameter choice, $\bar{\theta}$, leads to a model instance, $\mathcal{M}(\bar{\theta})$, of the model structure $\mathcal{M}$.

A system model structure can be achieved in different manners. It can be the result of careful *modeling* based on physical principles and well-established relations. The parameters in such models will be related to some physical meaning. If all parameters are known, the model is called a *white box* and if some parameters are unknown, it is called *gray box*. In the case of gray box models, the unknown parameters must be determined, e.g. from an identification procedure.

An alternative to modeling from physical principles is to determine the model

structure directly from available data. The resulting model is called *black box* and its parameters have no obvious physical interpretation.

Remark 4.1. An important characteristic of a system model is its *generalization capacity*, i.e. the scope to which a model behavior is consistent to that of the actual system. Typically, models derived from physical principles will have larger generalization capacities than black box models.

4.3.1   Fault Models

Of special importance is the modeling of faults. The fault model chosen must reflect the physical effects of the fault and how they appear in the available data. Faults can be categorized by its time behavior and by the manner they affect the system. With respect to the time behavior, a fault might be:

Abrupt, faults that affect the system abruptly, stepwise.

Incipient, faults that develop gradually with time.

Intermittent, faults that affect the system with interruptions.

Corresponding to the manner a fault affects the system, a fault might be:

Additive, faults that are effectively added to the system's inputs or outputs.

Multiplicative, faults acting on a parameter of the system. For example, changing a parameter θ of the model structure $\mathcal{M}$.

Structural, faults introducing new governing terms to the describing equations of the system. For example, changing the model structure $\mathcal{M}$.

Additive faults are typically used to model sensor faults, while multiplicative faults are mostly used to model a system fault.

───── **Example 4.2: Linear discrete time-invariant dynamic state space models** ─────
The class of *linear discrete time-invariant dynamic state-space* models can be described by

$$x(k+1) = A(\theta)x(k) + B(\theta)u(k) + B_{\bar{d}}(\theta)\bar{d}(k) + B_f(\theta)f(k) \tag{4.3a}$$

$$y(k) = C(\theta)x(k) + D(\theta)u(k) + D_{\bar{d}}(\theta)\bar{d}(k) + D_f(\theta)f(k), \tag{4.3b}$$

where $x(k)$ are states and the matrices $A(\theta)$, $B(\theta)$, $C(\theta)$, $D(\theta)$ describe the relation between u and y. The augmented disturbance and noise vector is $\bar{d}(k) = [d^T(k), \ v^T(k)]^T$ with effects described by the matrices $B_{\bar{d}}$ and $D_{\bar{d}}$. The faults $f(k)$ affect the inputs and outputs additively through $B_f(\theta)$ and $D_f(\theta)$. The matrices describe the model structure $\mathcal{M}$. A particular choice of parameters $\bar{\theta}$ gives a model instance $\mathcal{M}(\bar{\theta})$. Multiplicative faults are also possible by letting $\theta = \theta_0 + \Xi f(k)$, with faults affecting the nominal parameters θ_0 through the matrix Ξ.

┌─── **Example 4.3: An industrial robot under wear and temperature effects** ───┐

With references to Section 2.2 and Chapter 3, a manipulator can be described in a simplified manner by a multi-body rigid model

$$M(\varphi)\ddot{\varphi} + C(\varphi, \dot{\varphi}) + \tau_g(\varphi) + \tau_f(\dot{\varphi}, \tau_l, T, \mathsf{w}) = \tau, \qquad (4.4)$$

where the parametric dependencies are not shown for simplicity. The friction torques are described as function of angular speed $\dot{\varphi}$, manipulation load torques τ_l, the temperatures of the joints T, and the wear levels w. The fault f relates to the wear levels w. Wear changes the behavior of friction in a gradual manner, and is therefore an *incipient multiplicative* fault. The measured (known) outputs y are the angular positions φ, from which is also possible to achieve speed $\dot{\varphi}$, and acceleration $\ddot{\varphi}$. The known inputs u are the applied torques τ^{i}. The measured quantities are corrupted by noise v (not shown in (4.4)). The loads, τ_l, and temperatures, T, are unknown and considered as disturbances d.

└──┘

4.4   Fault Indicator Generation

The objective of this task can be formulated as "given data (u, y), design an algorithm that generates a fault indicator, i.e. a quantity which is affected by a fault f". In case knowledge related to the unknowns (v, d, f) is available, it can be used for the design.

An important tool for the design of fault indicators is a system model structure $\mathcal{M}$. The model structure contains very important information regarding the system behavior. Its availability can substantially support the design of fault indicators. Considering that the actual model instance generating the data is given by $\mathcal{M}(\theta_0)$, two methods are presented here:

Output observer: The model instance, $\mathcal{M}(\theta_0)$, is assumed known. $\mathcal{M}(\theta_0)$ is used to reconstruct the output from the data (u, y), creating an analytical redundancy $\hat{y}(\theta_0)$ which provides a fault indicator that is given by $\varepsilon(\theta_0) = y - \hat{y}(\theta_0)$, also known as *residual*.

Notice that in this case the residual indicates a deviation of the observations from a reference behavioral mode related to θ_0. That is, the residual $\varepsilon(\theta_0)$ can already be seen as a test quantity. For example, when θ_0 are the nominal parameters, the residual indicates deviations from a *no-fault* reference behavioral mode. In case knowledge about other reference modes is available, further processing of $\varepsilon(\theta_0)$ can provide additional test quantities.

Parameter estimation: Only the model structure, $\mathcal{M}$, is assumed known. Using the measured data (u, y), it generates an estimate $\hat{\theta}$ of the parameter θ_0. This is a natural choice for multiplicative faults. In case external knowledge

---

[i]Based on the simplification that the relation between current and applied torque is given by a constant. See Section 2.4.1 for details.

about θ_0 is available, a test quantity can be defined as $\hat{\theta} - \theta_0$ to indicate deviations from the reference behavior related to θ_0.

Notice that given an estimate $\hat{\theta}$, it may also be possible to provide an output estimate $\hat{y}(\hat{\theta})$ and the residual $\varepsilon(\hat{\theta}) = y - \hat{y}(\hat{\theta})$. The residual $\varepsilon(\hat{\theta})$ is a test quantity suitable for diagnosis of structural faults. As in the output observer case, if other reference behavioral modes are know, additional test quantities can be defined.

Remark 4.2. Parameter estimation techniques are a natural choice for multiplicative and structural faults, while the typical formulation for output observers considers additive faults. Nevertheless, these methods can be used interchangeably, Isermann (2006).

When a model structure is not available, alternative solutions are possible. These solutions will typically require expert knowledge about the data or extra (redundant) sensory information. An example of such expert knowledge is found in the analysis of the characteristics of measured signals, e.g. its frequency response. Another common approach is the use of *labeled data*, i.e. data sets under which the behavioral modes are known. Essentially however, any fault indicator generation method attempts to provide a quantity sensitive to faults given the available knowledge and observations. Some approaches to the generation of fault indicators are described next.

4.4.1   Output Observer

Different approaches are possible depending for instance on the model structure, Frank and Ding (1997). The trade-offs of the design are illustrated here for the case of a linear discrete model as described in (4.3). The presentation is based on Liu and Zhou (2008), where the residual is achieved from a filter. This filter is sometimes known as *fault detection filter*, but is directly related to an output observer, as will be illustrated.

By taking the **Z**-transform, (4.3) can be rewritten as

$$y(z) = G_u u(z) + G_{\bar{d}} \bar{d}(z) + G_f f(z), \quad \text{where} \tag{4.5}$$

$$\begin{bmatrix} G_u & G_{\bar{d}} & G_f \end{bmatrix} = \left[\begin{array}{c|ccc} A & B & B_{\bar{d}} & B_f \\ \hline C & D & D_{\bar{d}} & D_f \end{array} \right], \tag{4.6}$$

and the dependencies on θ were dropped for simplicity. The state-space realization (4.6) can be rewritten using left coprime factorization as (Liu and Zhou (2008)),

$$\begin{bmatrix} G_u & G_{\bar{d}} & G_f \end{bmatrix} = M^{-1} \begin{bmatrix} N_u & N_{\bar{d}} & N_f \end{bmatrix}, \quad \text{where} \tag{4.7}$$

$$\begin{bmatrix} M & N_u & N_{\bar{d}} & N_f \end{bmatrix} = \left[\begin{array}{c|cccc} A + LC & L & B + LD & B_{\bar{d}} + LD_{\bar{d}} & B_f + LD_f \\ \hline C & I & D & D_{\bar{d}} & D_f \end{array} \right], \tag{4.8}$$

and L is such that the realization is stable, i.e. $A + LC$ is stable. In Frank and Ding (1997), it has been shown that a residual generation filter (fault detection filter)

can be written, without loss of generality, as,

$$\varepsilon(z) = Q\left(My(z) - N_u u(z)\right), \tag{4.9}$$

where the free matrix Q (to be designed) is a square stable transfer matrix with same dimension as $y(z)$. In Ding et al. (1994) it was shown that

$$My(z) - N_u u(z) = y(z) - \hat{y}(z), \tag{4.10}$$

where the output estimate $\hat{y}(z)$ is given by the observer

$$\dot{\hat{x}}(k) = A\hat{x}(k) + Bu(k) - L\left(y(k) - \hat{y}(k)\right) \tag{4.11}$$

$$\hat{y}(k) = C\hat{x}(k) + Du(k). \tag{4.12}$$

This gives that any residual generation filter has the form

$$\varepsilon(z) = Q\left(y(z) - \hat{y}(z)\right), \tag{4.13}$$

and is composed of an output observer that is, if needed, filtered by Q (Frank and Ding (1997)).

By substituting $y(z)$ in (4.9) using (4.5) and (4.7), it can also be shown that

$$\varepsilon(z) = QN_{\bar{d}}\bar{d}(z) + QN_f f(z), \tag{4.14}$$

and the residual depends on the unknown disturbance $\bar{d}$ and unknown fault f. In case no disturbance is present, the residual is only dependent on the fault. In practice, disturbances are always present and the choice of Q can be seen as a multi objective optimization problem. The objectives are to choose Q that maximizes sensitivity to faults while achieving good robustness to disturbances, formulated for example as

$$\max_Q \quad \left\|QN_f\right\|_\vartheta \tag{4.15}$$

$$\text{subject to} \quad \left\|QN_{\bar{d}}\right\|_\delta < \upsilon, \tag{4.16}$$

for a constant υ. The choice of norms represents different compromises of the design. For example, $\vartheta = 2$ and $\delta = \infty$ provides a maximization of the average fault sensitivity with respect to a worst case disturbance. In Liu and Zhou (2008), solutions are found for different choices of ϑ and δ. In Li and Zhou (2009) the results are extended to the time-varying case. See also Liu (2008); Li (2009).

4.4.2 Parameter Estimation

The objective is to identify the unknown parameters of a known model structure from the data. Algorithmically, the solutions will depend on the model structure and whether a real-time method is sought or not, Gustafsson (2000). To illustrate this class of methods, the *prediction error method* is described next. The presentation is based on Ljung (1998); Peretzki et al. (2011).

Given data up to time index $k-1$, a (one-step ahead) predictor is a function that

generates estimates of the output at time k,

$$\hat{y}(k, \theta) = g(k, u^{k-1}, y^{k-1}, \theta), \tag{4.17}$$

where $u^{k-1} = [u(k-1), u(k-2), \cdots, u(1)]$ and similarly for y^{k-1}, $g(\cdot)$ is a predictor function and defines the model structure $\mathcal{M}$. A prediction error is thus defined as $\varepsilon(k, \theta) = y(k) - \hat{y}(k, \theta)$. For a data sequence of size N, the objective is to choose θ that best describes the data in a prediction error sense,

$$\hat{\theta}_N = \underset{\theta}{\operatorname{argmin}} \left\| y - \hat{y}(k, \theta) \right\|_\delta \tag{4.18}$$

where the choice of norm is a design choice. The least squares problem is written with the square of the Euclidian norm, i.e. $\|\cdot\|_2^2$. Since the least squares objective is differentiable, gradient methods can be used to solve the minimization.

In case the predictor is a linear (or affine) function of the parameters, a *linear regression* is achieved,

$$\hat{y}(k, \theta) = \phi^T(k)\theta, \tag{4.19}$$

where $\phi(k) = \phi(k, u^{k-1}, y^{k-1})$ is a vector function with same dimension as θ and defines the model structure $\mathcal{M}$. In this case, the solution to the least squares problem is given in closed-form by

$$\hat{\theta}_N = \underbrace{\left[\frac{1}{N} \sum_{k=1}^{N} \phi(k)\phi^T(k) \right]^{-1}}_{\hat{R}_N} \frac{1}{N} \sum_{k=1}^{N} \phi(k)y(k), \tag{4.20}$$

where the feasibility of this solution depends on whether the *information matrix* $\hat{R}_N$ is invertible. Notice that this algorithm can also be defined in a recursive form, providing an estimate $\hat{\theta}(k)$ at each sample time, allowing for real-time methods, see e.g. Ljung (1998); Gustafsson (2000).

Assuming that the true system generating the data is described by a linear regression with same model structure as the predictor, with true parameters θ_0, and with additive white noise with variance γ_0. Then, as $N \to \infty$, the estimate $\hat{\theta}_N$ will be asymptotically normally distributed, Ljung (1998). More precisely

$$\sqrt{N}(\hat{\theta}_N - \theta_0) \in As \, \mathcal{N}(0, P), \quad P \triangleq \gamma_0 \left[\lim_{N \to \infty} \hat{R}_N \right]^{-1}. \tag{4.21}$$

For a finite number of data, N, an estimate of the covariance matrix $\hat{P}_N$ is

$$\hat{P}_N = \frac{1}{N} \hat{\gamma}_N \, [\hat{R}_N]^{-1}, \quad \hat{\gamma}_N = \frac{1}{N} \sum_{k=1}^{N} \varepsilon^2(k, \hat{\theta}_N). \tag{4.22}$$

The matrix $\hat{R}_N$ therefore determines the quality of the estimate $\hat{\theta}_N$. Notice that $\hat{R}_N$ is a function of the data and the model structure $\mathcal{M}$. In order to make the covariance small, $\hat{R}_N$ should be made large in some sense. This idea is explored, for instance, in experiment design (choosing u) for system identification. For

on-line solutions, this restriction might be complicating since u can not be set freely.

4.4.3   Signal-driven Methods

In many applications, the available data are in fact measured signals. It is possible to extract information about the fault by only considering the characteristics of these signals.

With the objective of making it easier to reveal relevant characteristics of the signals, *transforms* are widely used in signal-driven methods. A transform is used to "map" a signal from its original domain to an alternative domain. The alternative domain is hopefully more suitable for the indication of faults.

An integral transform is any transform $\mathbf{T}$ of the form

$$y(v) = \mathbf{T}\{y(t)\} = \int_{t_0}^{t_1} k(t, v) y(t) \, \mathrm{d}t. \tag{4.23}$$

where $y(t)$ is a signal with domain on t, $y(v) = \mathbf{T}\{y(t)\}$ is the transformed signal with domain on v and $k(t, v)$ is a *kernel function*. Several types of integral transforms and discrete transforms can be defined, e.g. Fourier transform, Wavelet transform, Karhunen-Loève transform, Radon transform, etc. Each transform will provide different properties in the transformed domain. For example, the Fourier transform $\mathbf{F}\{y(t)\}$ is a transform with

$$k(t, v) = e^{-itv}, \quad t_0 = -\infty, \quad t_1 = \infty. \tag{4.24}$$

When t is time, v is frequency and $y(v)$ relates to the coefficients of the Fourier series of $y(t)$. The transformed signal $y(v) = \mathbf{F}\{y(t)\}$, is thus said to be the frequency representation of $y(t)$. The analysis of data in the frequency-domain has found particular success in the monitoring of rotating machines, Taylor (1994); de Silva (2007). Example 4.4 illustrates the use of frequency domain analysis for diagnosis of rotating machines.

Remark 4.3. A disadvantage with signal-driven methods is that the characteristics of the signal will typically depend on the system operational points, restricting the scope to which the method can be applied. This is specially difficult if on-line solutions are sought.

Example 4.4: An off-line active method for backlash monitoring

This example is based on Sander-Tavallaey and Saarinen (2009). In the work, the diagnosis of increased backlash is studied in drives equipped with compact gearboxes. An increase of backlash will introduce additional resonance peaks to the frequency content of the drive response. The analysis of the signals' spectra can therefore be used to indicate backlash changes.

A dedicated *test-cycle*, displayed in Figure 4.5a, is used to excite the drive unit. In Figure 4.5b, spectra estimates for the torque signals are shown for a healthy unit and for a unit with increased backlash levels. As it can be seen, there is an
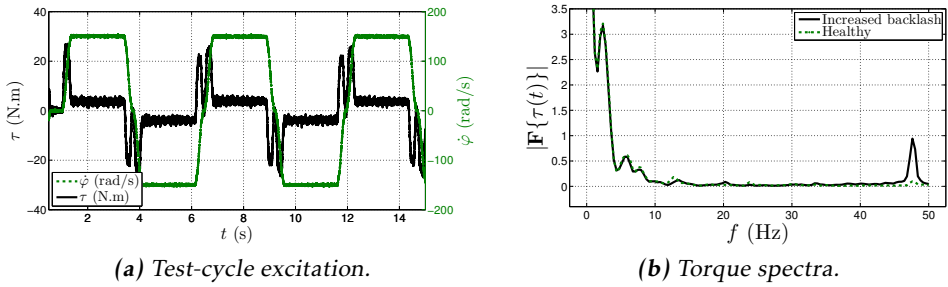
(a) Test-cycle excitation.

(b) Torque spectra.

Figure 4.5: *Backlash monitoring through signal analysis. The drive is excited with a test-cycle in an off-line manner as displayed in (a). The torque signals' spectra are shown in (b). Notice the increased resonance peak around 47 Hz for the unit with increased backlash compared to the healthy unit.*

increase of the frequency response around 47 Hz. In the paper, this deviation is used to generate a backlash indicator which is used for backlash monitoring. The proposed method takes only a few seconds to execute and does not consider additional vibration measurements, which are common for this type of method. Notice that this is an *off-line active* solution since it is based on a test-cycle.

4.4.4 Data-driven Methods

The methods developed in this work only consider data with domain on $\mathbb{R}$, the real numbers. Data can however occur in a variety of forms, e.g. characters, images, etc. Data-driven methods are therefore rather general. Perhaps the simplest illustrative example is to check for domain consistency. For example, consider that the domain of the data is "strings containing names of countries", a datum containing a number or the string "Linköping" can be interpreted as a fault.

Probability and statistics provide useful tools that can, at least in principle, be used for a large variety of data types. A possible alternative to generate a fault indicator is to consider analysis of the data *probability density function*. An estimate of the data density function can be found, e.g., using histograms. In the following, a density estimator relevant to this work is introduced.

Kernel density estimator

For a scalar random variable Y with probability density function $p(y)$, the characteristic function $\Phi_Y : \mathbb{R} \to \mathbb{C}$ is defined as

$$\Phi_Y(\nu) = \mathbb{E}\left[e^{i\nu Y}\right] = \int_{-\infty}^{\infty} e^{i\nu y} p(y)\, dy = \mathbf{F}^{-1}\{p(y)\}2\pi, \qquad (4.25)$$

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 $\Phi_Y(\nu)$ is given by

$$\hat{\Phi}_Y(\nu) = \frac{1}{N} \sum_{i=1}^{N} e^{i\nu y_i}, \tag{4.26}$$

the objective is then to estimate the density function from $\hat{\Phi}_Y(\nu)$. This is essentially a spectrum estimation problem. A direct estimation of the density function from the Fourier transform of $\hat{\Phi}_Y(\nu)$ will however lead to an estimate with increased variance for large values of ν, Ljung (1998).

To avoid this problem, the empirical estimate of the characteristic function is multiplied with a weighting function $K_h(\nu) = K(h\nu)$. The weighting function is typically symmetric, satisfying $K(0) = 1$ and tends to zero when ν tends to infinity. The density estimate is then given by

$$\hat{p}(y) = \frac{1}{2\pi} \mathbf{F} \left\{ \hat{\Phi}_Y(\nu) K(h\nu) \right\} = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\nu y} \hat{\Phi}_Y(\nu) K(h\nu) \, d\nu$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{1}{N} \sum_{i=1}^{N} e^{i\nu(y_i - y)} K(h\nu) \, d\nu = \frac{1}{Nh} \sum_{i=1}^{N} \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\nu\left(\frac{y_i - y}{h}\right)} K(h\nu) \, d(h\nu)$$

$$= \frac{1}{Nh} \sum_{i=1}^{N} k \left(\frac{y_i - y}{h} \right) = \frac{1}{N} \sum_{i=1}^{N} k_h(y - y_i), \tag{4.27}$$

where $k_h(y)h = \mathbf{F}^{-1}\{K_h(\nu)\}$. The function $k_h(y)$ is a *kernel function*, satisfying $k_h(\cdot) \geq 0$ and that integrates to 1. The resulting estimator is known as a *kernel density estimator*. It is typical to choose kernel functions with a low pass behavior, where the bandwidth parameter h controls its cutoff frequency. Common kernel functions and their Fourier transforms are shown in Figure 4.6. See Parzen (1962); Bowman and Azzalini (1997); Jones and Henderson (2009) for a detailed treatment of kernel density estimators and criteria/methods for choosing h.

The Karhunen-Loève transform - PCA

Another important type of transform is the Karhunen-Loève transform, which leads to a representation of a stochastic process in an orthonormal basis providing the minimal mean squared error, Hotelling (1933); Karhunen (1947); Loève. (1948). If Y_t is a zero mean stochastic process with covariance function $K_Y(t, u)$, the Karhunen-Loève theorem states that Y_t can be represented by the infinite

(d) Uniform   *(e) Triangular*   *(f) Gaussian*
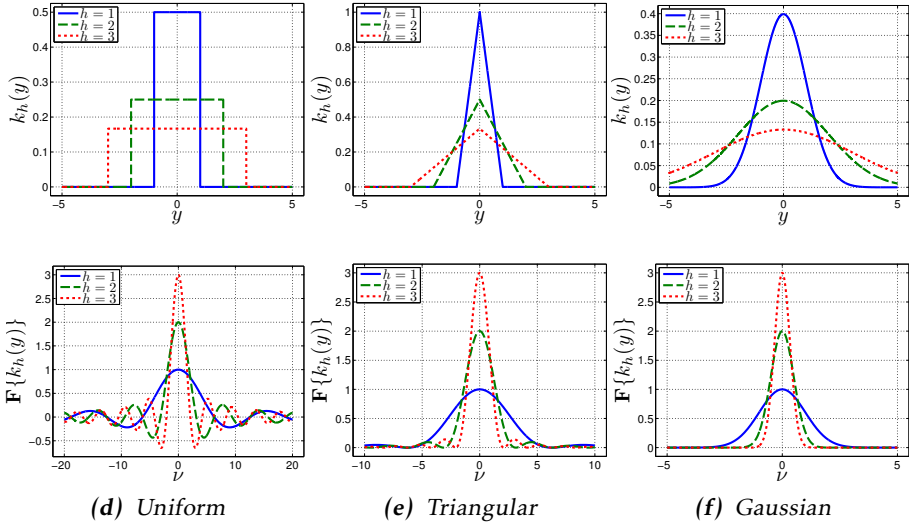
Figure 4.6: *Kernel functions (top) and their respective Fourier transforms (bottom).*

series,

$$Y_t = \sum_{n=1}^{\infty} Z_n f_n(t) \, dt, \quad \text{with } Z_n = \int_{t_0}^{t_1} Y_t f_n(t) \, dt, \tag{4.28}$$

where $f_n(t)$ in (4.28) is given by $\int_{\nu_0}^{\nu_1} K_Y(t, \nu) f_n(\nu) \, d\nu = \lambda_n f_n(t),$ (4.29)

and Z_n are independent random variables. The $f_n(t)$ are eigenfunctions with eigenvalues λ_n which are found by solving the homogeneous Fredholm integral equation of second kind in (4.29). Solving (4.29) is a difficult problem, but in the discrete case, the integral equation is simplified to an eigenvalue decomposition of the empirical covariance matrix. This discrete counterpart leads to what is known as principal component analysis (PCA), which finds broad use in diagnosis applications, Yoon and MacGregor (2004); Gertler and Cao (2004); Yin et al. (2010).

4.5   Behavior Testing

Given a fault indicator, a decision should be made regarding the behavioral modes that might be currently present, i.e. a diagnosis should be generated. In order to proceed, knowledge must be available about the *reference behaviors* which are compared to the observed behavior of the fault indicator.

Remark 4.4. In practice, it is difficult to know a priori how every fault mode affects the fault indicator. What is usually known is how the fault indicator behaves under no fault. So the reference behavior is usually related to the *no-fault* mode.

Once the reference behaviors have been decided, the next step is to define an algorithm to *compare* the current behavior with the reference behaviors, generating *test quantities*. This is typically, but not necessarily, done with a *distance measure*. A distance measure $d(x, y)$ between x and y is a function, $d : X \times X \to \mathbb{R}$, satisfying

$$d(x, y) \geq 0 \text{ (non-negativity)} \tag{4.30}$$
$$d(x, y) = d(y, x) \text{ (symmetry)} \tag{4.31}$$
$$d(x, x) = 0 \text{ (reflexivity)}. \tag{4.32}$$

A *metric* is a distance satisfying the extra requirements

$$d(x, y) = 0 \text{ if and only if } x = y \text{ (identity of indiscernibles)} \tag{4.33}$$
$$d(x, y) \leq d(x, z) + d(z, y) \text{ (triangle inequality)}. \tag{4.34}$$

Independent of how the behavior comparison is done, a test quantity is a scalar that measures deviations from one or more reference behaviors[ii]. Each test quantity is input to a *decision rule* that will accept or reject the reference behaviors related to the test quantity, i.e. a diagnosis is generated.

In practice, because of noise and disturbances, a test quantity will typically present a random behavior, which may complicate the decision. It is therefore important to evaluate the achievable performance of a test quantity. If the performance is not satisfactory, a *filtering step* can be added to the decision rule to improve robustness. The introduction of a filter will however add a delay to the decision, i.e. more evidences need to be collected before a satisfactory decision can be made.

In the following, a brief discussion on distance measures, evaluation of test quantities and decision rules is considered.

Remark 4.5. Behavior testing can be formulated as a statistical hypothesis testing, where a statistical model is known for the behavior of the fault indicator under different modes, see e.g. Nyberg (1999); Gustafsson (2000); Basseville and Nikiforov (1993). It can also be seen as a classification problem, where labeled data are used, see e.g. Mackay (2003); Bishop (2007).

4.5.1 Behavior Comparison

The choice of algorithm to generate a test quantity will depend on the type of fault indicator and on the reference behaviors. Here, some relevant cases for the discussion are illustrated, for a detailed presentation see Basseville (1989); Gustafsson (2000); Deza and Deza (2009).

[ii]Notice that several test quantities can be produced from the same fault indicator, e.g. when several reference behaviors are considered.

Comparing residuals

There are several ways to achieve a residual. For example, it can be found for a known model instance as $\varepsilon(k) = y(k) - \hat{y}(k, \theta_0)$, using an estimate of θ according to $\varepsilon(k) = y(k) - \hat{y}(k, \hat{\theta})$, or as the result of estimates from different models $\varepsilon(k) = \hat{y}(k, \theta_1) - \hat{y}(k, \theta_2)$. In any case, a reference behavior must be assigned, and is either known or estimated from the data.

It is common to consider that a residual $\varepsilon(k)$ *resembles white noise before any change has occurred*. The assumption is well based from estimation theory, e.g. when the residuals are taken as the innovations of a Kalman filter when the model used is correct. In such case, a test quantity $s(k)$ can take different forms, some examples are

$$s(k) = \varepsilon(k), \text{ for changes in the mean}, \tag{4.35}$$

$$s(k) = \varepsilon^2(k) - \gamma, \text{ for changes in the variance}, \tag{4.36}$$

$$s(k) = \varepsilon(k)\varepsilon(k-1), \text{ for changes in correlation sign}, \tag{4.37}$$

where γ is a known residual variance in the no-fault mode. The different test quantities will be more or less suitable depending on the actual fault mode. For instance, if the fault is known to affect the mean of the residual, then (4.35) should be used.

Distances for parameter estimates

According to (4.21), after enough data was collected, the test quantity

$$s(k) = \left(\hat{\theta}(k) - \theta_0\right)^T \hat{P}(k)^{-1} \left(\hat{\theta}(k) - \theta_0\right) \tag{4.38}$$

is *chi-square* distributed with degrees of freedom equal to the number of parameters. In case θ_0 is unknown, a filtering scheme can be used to provide an estimate $\hat{\theta}_0(k)$, for example using a moving average. This measure is used for example in Peretzki et al. (2011) as an estimate of the data quality.

Distances for spectra

Several distances can be defined to compare differences between the spectra of signals, Gray and Markel (1976); Basseville (1989). Let $y^0(\nu)$ and $y^1(\nu)$ be spectra estimates, the log-spectral distance is written as

$$\left\|\log \frac{y^0(\nu)}{y^1(\nu)}\right\|_\delta. \tag{4.39}$$

The choice of norm will provide different characteristics, e.g. $\delta = 2$ leads to mean quadratic distance and $\delta = \infty$ to maximum deviation.

Distances for densities

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

$$D_{\mathrm{KL}}\left(p\|q\right) = -\int_{-\infty}^{\infty} p(y) \log \frac{q(y)}{p(y)} \, \mathrm{d}y. \tag{4.40}$$

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 distance, since it is not symmetric in general. The quantity

$$\mathrm{KL}\left(p\|q\right) \triangleq D_{\mathrm{KL}}\left(p\|q\right) + D_{\mathrm{KL}}\left(q\|p\right), \tag{4.41}$$

known as the *Kullback-Leibler distance* (KL), is however symmetric and also a metric, satisfying the triangle inequality.

The KLD is in fact a specialization of an f-divergence, a family of functions that can be used as measures of the differences between densities. Other examples of divergences and distances based on f-divergences are the Hellinger distance, the total variation distance and the α-divergence, see Reid and Williamson (2011); Basseville (1989) for more.

4.5.2   Evaluation of Test Quantities

A test quantity contains important information about how far the current behavior of the system is from one or more reference behavioral modes. In case they agree, the test quantity $s(k)$ should be small (the evidence found in the test quantity accepts the reference behaviors), otherwise, it will deviate. Each of these conjectures can be seen as a hypothesis. The null hypothesis $\mathcal{H}_0$ corresponds to the case that they agree and the alternative hypothesis is $\mathcal{H}_1$. The problem of deciding which behavioral mode is present can thus be seen as a *hypothesis test*.

A decision regarding which hypothesis is present can be made with the use of a threshold $\hbar$,

$$s(k) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \hbar, \tag{4.42}$$

and reads, if $s(k) \geq \hbar$ choose $\mathcal{H}_1$, otherwise choose $\mathcal{H}_0$.

Due to noise and disturbances, $s(k)$ will typically have a random behavior, which will differ according to the hypothesis present. The probability densities of $s(k)$ for each hypothesis is written as $p(s|\mathcal{H}_0)$ and $p(s|\mathcal{H}_1)$. For the test in (4.42), the following probabilities are of interest,

$$P_f = \int_{\hbar}^{\infty} p(s|\mathcal{H}_0) \, \mathrm{d}s, \text{ assigns } \mathcal{H}_1 \text{ when } \mathcal{H}_0 \text{ is present.} \tag{4.43}$$

$$P_d = \int_{\hbar}^{\infty} p(s|\mathcal{H}_1) \, \mathrm{d}s, \text{ assigns } \mathcal{H}_1 \text{ when } \mathcal{H}_1 \text{ is present.} \tag{4.44}$$

With interest on the transition from $\mathcal{H}_0$ to $\mathcal{H}_1$, the probability P_f corresponds to a false detection, and P_d corresponds to a correct detection. The ideal test would always provide $P_f = 0$ and $P_d = 1$. This is however usually not possible because of overlaps between $p(s|\mathcal{H}_0)$ and $p(s|\mathcal{H}_1)$.

By varying the threshold $\hbar$ from $-\infty$ to ∞ and plotting P_d against P_f a *receiver operating characteristic* (ROC) curve is achieved. The ROC curve presents the performance limitations of the test quantity with respect to P_f and P_d. From a specified P_f, it is also possible to find the threshold providing the largest P_d (Neyman-Pearson criterion, see Van Trees (2001)).

Example 4.5: Evaluation of a test quantity for wear diagnosis in a robot

In Paper C, a test quantity is achieved which is sensitive to wear in a robot joint. The developed test quantity considers a distance between two sequences of applied torques, $\boldsymbol{\tau}^m$ and $\boldsymbol{\tau}^n$, and is written $s(\boldsymbol{\tau}^m, \boldsymbol{\tau}^n)$. The distance is also a metric and therefore $s(\boldsymbol{\tau}^m, \boldsymbol{\tau}^n) = s(\boldsymbol{\tau}^n, \boldsymbol{\tau}^m)$.

As presented in Example 4.3, the applied torque is affect by the wear levels w, but also by the joint temperature T, i.e. $\tau(\mathrm{w}, T)$. Temperature is considered as a disturbance since it is not measured. The objective is to decide whether $\boldsymbol{\tau}^m$ and $\boldsymbol{\tau}^n$ were generated under the same wear levels or if there was a critical wear change of size w_c between them. The corresponding hypotheses are

$$\mathcal{H}_0 : \quad \boldsymbol{\tau}^m(\mathrm{w}=0, T) \text{ and } \boldsymbol{\tau}^n(\mathrm{w}=0, T) \text{ or} \tag{4.45}$$
$$\boldsymbol{\tau}^m(\mathrm{w}=\mathrm{w}_c, T) \text{ and } \boldsymbol{\tau}^n(\mathrm{w}=\mathrm{w}_c, T)$$
$$\mathcal{H}_1 : \quad \boldsymbol{\tau}^m(\mathrm{w}=\mathrm{w}_c, T) \text{ and } \boldsymbol{\tau}^n(\mathrm{w}=0, T) \text{ or} \tag{4.46}$$
$$\boldsymbol{\tau}^m(\mathrm{w}=0, T) \text{ and } \boldsymbol{\tau}^n(\mathrm{w}=\mathrm{w}_c, T).$$

For analyses of the disturbance effects, temperature is considered random with a uniform distribution

$$T \sim \mathcal{U}(\underline{T}, \underline{T} + \Delta_T), \tag{4.47}$$

where $\underline{T} = 30°\mathrm{C}$ and the value of Δ_T relates to the size of the disturbance. The resulting distributions for each hypothesis, $p(s|\mathcal{H}_0)$ and $p(s|\mathcal{H}_1)$, are estimated using a kernel density estimator. They are shown for different levels of temperature disturbances Δ_T in Figure 4.7, together with their ROC curves. For low values of Δ_T, it is easy to distinguish between the hypotheses. The achievable performance of the fault indicator is however considerably affected by the size of Δ_T.

4.5.3   Decision Rules

A decision rule is a test used to *accept or reject* the reference behaviors that the test quantity can explain, that is, it generates a diagnosis. The simplest decision rule is to directly use a threshold as in (4.42). The choice of threshold might be motivated, e.g., from the Neyman-Pearson criterion but it can also be simpler, as illustrated in the example below.
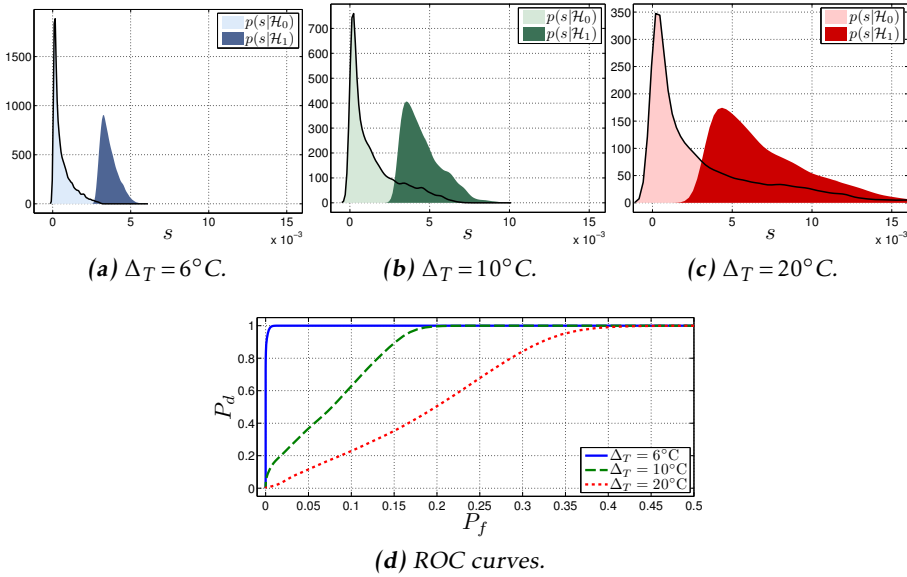
(a) $\Delta_T = 6°C$.   *(b)* $\Delta_T = 10°C$.   *(c)* $\Delta_T = 20°C$.

(d) ROC curves.

Figure 4.7: *Kernel density estimates for the hypotheses' densities under different levels of temperature disturbance Δ_T (top row), the same x-axis is used. The corresponding ROC curves (bottom).*

Example 4.6: Limit checking

One of the simplest diagnostic tests is to verify whether a variable exceeds certain limits. Let $y(k)$ be the variable of interest, consider that exceeding an upper or lower level corresponds to different fault modes $F1$ and $F2$ respectively, and that otherwise a no-fault mode NF is present. Two test quantities are then of interest

$$s_1(k) = y(k), \quad s_2(k) = -y(k), \tag{4.48}$$

and the diagnostic tests are

$$s_1(k) \underset{\{NF,F2\}}{\overset{\{F1\}}{\gtrless}} \overline{y}, \quad s_2(k) \underset{\{NF,F1\}}{\overset{\{F2\}}{\gtrless}} \underline{y} \tag{4.49}$$

where $\overline{y}$ and $\underline{y}$ are the upper and lower limits. Notice that the acceptance of the null hypothesis in either of the tests does not nullify the other.

Unfortunately, the direct use of a threshold might not be enough to meet the performance requirements for the diagnosis. For example, the susceptibility of a test to generate incorrect diagnoses can be measured by P_f in (4.43). Incorrect diagnoses are undesirable, and therefore an upper limit for P_f might be required. In such case, the detection performance, measured e.g. by P_d in (4.44), will be restricted by the achievable performance of the test quantity. This was illustrated in Example 4.5 where the temperature disturbances were a limiting factor.

In order to circumvent this problem, a filter can be applied to the test quantity *before* thresholding. Different filtering approaches are possible, but they will introduce delay on the detection, Adnan et al. (2011). That is, a correct decision is compromised with the detection time. Some common filtering approaches are discussed next.

Remark 4.6. The choice of decision rule is sometimes recognized as *alarm design*, and is related to *alarm management*. The topic has recently drawn attention due to new regulations in the process industry, see e.g. ANSI/ISA 18.2 (2009).

Low-pass filtering

A low-pass filter can be applied to $s(k)$ before thresholding. This will reduce the variability of $s(k)$ and can therefore improve performance but introduces a delay since more samples are needed. For instance, a geometric moving average filter,

$$g(k) = \lambda s(k) + (1 - \lambda)g(k - 1), \tag{4.50}$$

can be used. Where $0 \leq \lambda \leq 1$ is a tuning parameter controlling the effective averaging window. The threshold is then applied to the test statistic $g(k)$.

CUSUM filter

An alternative filtering approach is the cumulative sum (CUSUM) filter.

Algorithm 2 CUSUM

$$g(k) = g(k - 1) + s(k) - \nu \tag{4.51}$$
$$g(k) = 0 \text{ if } g(k) < 0. \tag{4.52}$$

The test statistic $g(k)$ adds up the test quantity $s(k)$. To avoid positive drifts, the drift parameter ν is subtracted from the update rule (4.51). If, on the other hand, $g(k)$ becomes negative, $g(k)$ is reset, avoiding negative drifts. In case $s(k)$ is a non-negative quantity (e.g. a distance), it is natural to include an estimate of the mean of $s(k)$ in ν. To achieve robustness to noise, a quantity proportional to an estimate of its standard deviation can also be added, yielding

$$\nu = \mu + \kappa\sigma \tag{4.53}$$

where μ and σ are estimates of the mean and standard deviation of $s(k)$ before the change and κ is a positive constant. Notice that the CUSUM is a nonlinear filter. For more filtering approaches, see Gustafsson (2000); Cheng et al. (2011).

Delay-timer

Robustness can also be achieved by postponing the decision with a delay-timer. With a delay-timer, a detection is only triggered in case the test quantity exceeds the threshold for a consecutive number of times.

4.6   Summary and Connections

This chapter presented an overview of the diagnosis process. The presentation focused on describing different methods, with special attention to methods that are suitable for the application sought, condition monitoring of industrial robots. A number of examples was presented to illustrate the common trade-offs and limitations.

Gathering knowledge about the disturbances and faults is an important aspect for the design and verification of diagnosis solutions. The achieved friction models from Papers A and B allow for a more realistic design and evaluation of solutions to wear diagnosis in industrial robot joints. These models were used in Example 4.5.

An off-line active method for wear monitoring is proposed in Paper B. The method is based on parameter estimation and makes use of the developed friction models and a dedicated test-cycle. The performance properties of the method are also verified through simulations using these models.

In Paper C, on-line passive methods are proposed for the wear monitoring problem. The repetitive behavior typically found in applications of industrial robots is explored to generate fault indicators. Data-driven methods are suggested that consider changes in the data distribution achieved from the execution of the same trajectory. It is made use of kernel density estimators, presented in Section 4.4.4, the Kullback-Leibler distance, presented in Section 4.5.1, and the CUSUM filter. The achievable performance of a suggested fault indicator was analyzed using a similar approach as presented in Section 4.5.2.

5

Concluding Remarks

The first part of this thesis provided an introduction to the research fields that are relevant for this thesis: industrial robotics, tribology and diagnosis. This served as a preparation to Part II, motivating the research contributions and contextualizing them into the different fields.

The conclusions for the thesis are given in Section 5.1. Suggestions of future research topics are presented in Section 5.2. See also the included papers for details.

5.1   Conclusions

The main objective of this work was to provide methods for wear diagnosis in a robot joint. Since wear can affect the friction in the joint, the basic idea was to monitor friction to infer about wear. Because the friction torques must be overcome by the motors during the robot operation, it is possible to extract information about friction from available measurements. As presented in Chapter 2, there are different aspects of industrial robotics that make this a challenging problem:

Complex dynamics. Industrial robots are nonlinear, multi-variable, uncertain systems operating in closed-loop.

Limited sensory information. In a typical setup, only motor angular position φ_m, and applied motor current i_m are measured. From these measurements estimates of angular speed $\dot{\varphi}_m$ and the applied motor torques $\tau_m = K i_m$ is possible.

Application-related limitations. Industrial robots are used in a wide range of

applications. Depending on the installation, there will be, e.g., restrictions on the available workspace and on how the robot is used (excited).

For various reasons, the use of models is important in industrial robotics. Due to the complexity of a robot joint, models describing, e.g., backlash and friction are difficult and are often motivated from empirical observations. With interest in the monitoring of friction for diagnosis, Chapter 3 presented an overview of friction in a robot joint.

Chapter 4 gave an overview of the diagnosis process. It described the different tasks present and the challenges involved. It mainly focused on describing different approaches for the design of diagnostic tests. A diagnostic test makes a comparison between observations and available knowledge of the system to infer about the state present in the system. The representation of this knowledge may appear in different forms, depending, e.g., on the method used to provide fault indicators. In Paper B, this knowledge appears as an accurate friction model. In Paper C, this knowledge is contained in data achieved from executions of the same trajectory.

5.1.1 Summary of Part II

In Paper A, a detailed empirical study of friction in a robot joint is presented. The study was motivated by the significant complexity of friction in a robot (recall Figure 3.2). In the paper, the typical friction related phenomena and models used in robotics are reviewed. The effects of angular position, speed, load torques and temperature to static friction in a robot joint are considered. Due to their relevance, the effects of load and temperature are modeled, identified and validated. The proposed model considerably outperforms standard friction models and is important for design and evaluation of control and diagnosis methods.

In Paper B, the effects of wear to static friction are modeled and identified. The model is based on empirical studies from accelerated wear tests performed in a robot joint. The wear model is combined with the friction model presented in Paper A. Using the resulting model, an *off-line active* method is proposed for wear diagnosis. A test-cycle is used to provide an estimate of static friction. A known model structure is considered with only temperature and wear considered as unknowns and a *wear estimator* is proposed. Special attention is given to the performance limitations imposed by the unknown temperature. With the use of a known model and friction observations, the framework is favourable for identification methods and can be seen as a proof of concept for identification based methods. As it is shown, a careful experiment design can lead to a robust wear estimation.

The main limitations are however the use of a known friction model and the need for estimates of the static friction torques. The friction estimates are achieved off-line, through a test-cycle. Off-line methods decrease the robot availability and are therefore undesirable from the perspective of the robot user. The use of on-line passive identification-based methods for diagnosis is challenging since its

performance will depend on how the robot is excited. Furthermore, additional uncertainties might also appear, e.g. related to unmodeled dynamics of the robot.

In Paper C, *on-line passive* wear monitoring methods are proposed where the repetitive behavior found in most applications of industrial robots is explored. The basic idea is that the repetitive execution of a system provides redundancies about the system's behavior which are directly found in the data. The result of the execution of the same trajectory in different instances of time can be compared to provide an estimate of how the system changed over the period.

A data-driven method is suggested that considers changes in the distribution of torque data logged from the robot. A kernel density estimator is used, in combination with the Kullback-Leibler distance. Since a robot may execute several repetitive trajectories, ideas are presented to combine measurements from different trajectories. An approach is also given to improve the robustness to disturbances with the use of a weighting function. A data-driven method is suggested for the choice of weighting function, it requires availability of labeled data.

The methods are evaluated in simulations and experiments. It is shown that robust wear monitoring is made possible with the proposed methods. The effects of unknown temperature are investigated in detail through a simulation study. It is shown that the use of the weighting function can significantly improve the robustness to temperature disturbances.

Although the methods were developed with interest in industrial robots, they can be applied for any system that operate in a repetitive manner. This repetitive behavior is commonly found in automation or can be forced with the periodic execution of a test-cycle. An important advantage of the approach is its simplicity (no models are required), and it is also easy to implement.

5.2   Future Research

Several further developments are suggested in the end of each paper in Part II. Besides those, the following aspects are emphasized for the development of solutions for wear diagnosis in robot joints.

Friction modeling. A more detailed friction model will support the design and evaluation of any wear diagnosis solution. Possible extensions are:

- *Modeling of dynamic friction.* Only static friction models were developed in this work. One possibility is to verify whether the static friction models developed can be extended to a dynamic description, e.g. using the LuGre model.

- *Validation over a larger temperature range.* The temperature effects were modeled using measurements in a temperature range of $35-80°C$. Robots may operate outside this range. It is therefore important to understand how friction would be affected.

- *Verification of the models developed in other mechanical devices.* Realistic friction models are important in many applications. The effects of load and temperature are often neglected, despite their importance. In principle, the developed models can be extended to other types of mechanical devices under lubricated friction. The use of the model structure proposed would simplify the time consuming task of developing new friction models.

  For instance, while the study of friction considered here focused on the main axes of the robot, it would be interesting to study friction also in the wrist axes.

Wear modeling. Since it is costly and time consuming to perform wear experiments, wear models are very important for the design and verification of diagnosis solutions. The further study of wear and wear modeling in a robot joint is motivated by the following,

- *Unpredictability of the wear processes.* It is not possible to assure that a wear model developed based on an observed fault is representative of future faults. The wear processes in a joint are the results of complex phenomena that cannot be predicted. With more knowledge gathered about the faults, it is possible to describe them in a more detailed manner. Perhaps different wear models can be developed to describe the most common faulty behaviors.

- *Development of lifetime models.* Understanding how the wear evolves with time and usage is important for the design and verification of diagnosis solutions. For example, it supports the decision for the experimentation time and design of the decision rules for alarm generation.

  A lifetime model also allows for *prognosis*, which is very important for the scheduling of maintenance actions.

Evaluation of the methods. Independent of which diagnosis methods are used, it is important to thoroughly evaluate the approaches before use in real world applications. This will help to identify weaknesses of the methods and will provide leads to improvements in their design. Three approaches are listed, with an increasing level of "certification",

- *Simulation studies.* With the use of the developed friction models, it is possible to setup a realistic simulation model. In such simulation model, it is important to consider the different sources of uncertainties present in practice, e.g., flexibilities, torque ripple, temperature and load variations, closed-loop effects, etc. For on-line methods, cycles that are used in real world applications must be considered.

  To allow for a comparison of different methods, a benchmark problem for robust wear diagnosis in a robot joint could be defined using such a model.

- *Accelerated wear tests.* Even though a realistic simulation model is important, it cannot substitute the validation through experiments. Since the

wear effects may take several years to occur, accelerated wear tests, performed in a lab, can be used as a first verification. It is however difficult to reproduce scenarios that are representative of what happens in the field. For example, it is difficult to control temperature in the joints.

- *Field tests.* These are irreplaceable for the evaluation of diagnosis solutions. In order to be of significance, they must be verified with several robots and in different applications. This is however only possible in cooperation with robot users and can take considerable time.

Bibliography

N. Adnan, I. Izadi, and T. Chen. Computing detection delays in industrial alarm systems. In *Proc. of the 2011 American Control Conference*, pages 786 –791, july 2011.

F. Al-Bender and J. Swevers. Characterization of friction force dynamics. *IEEE Control Systems Magazine*, 28(6):64–81, 2008.

ANSI/ISA 18.2. *Management of Alarm Systems for the Process Industries*. ISA, Raleigh, USA, 2009.

T. Ardeshiri, M. Norrlöf, J. Löfberg, and A. Hansson. Convex optimization approach for time-optimal path tracking of robots with speed dependent constraints. In *Proc. of 18th IFAC World Congress*, pages 14648–14653, Milano, Italy, Sept. 2011.

K. J. Åström and C. Canudas-de Wit. Revisiting the lugre friction model. *Control Systems Magazine, IEEE*, 28(6):101–114, Dec. 2008.

P. Axelsson, R. Karlsson, and M. Norrlöf. Bayesian state estimation of a flexible industrial robot. Technical Report LiTH-ISY-R-3027, Department of Electrical Engineering, Linköping University, SE-581 83 Linköping, Sweden, Oct. 2011.

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

M. Basseville. On fault detectability and isolability. *European Journal of Control*, 7(6):625–638, 2001.

M. Basseville and I. V. Nikiforov. *Detection of abrupt changes: theory and application*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1993.

C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 1st ed edition, 2007.

G. Biswas, M.-O. Cordier, J. Lunze, L. Trave-Massuyes, and M. Staroswiecki. Diagnosis of complex systems: Bridging the methodologies of the fdi and dx com-

munities. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 34(5):2159 –2162, oct. 2004.

A. C. Bittencourt. *Friction change detection in industrial robot arms.* Msc. thesis, The Royal Institute of Technology, 2007.

A. C. Bittencourt and S. Gunnarsson. Static friction in a robot joint - modeling and identification of load and temperature effects. *ASME Journal of Dynamic Systems, Measurement, and Control.* Accepted for publication.

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

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

A. C. Bittencourt, K. Saarinen, S. Sander-Tavallaey, and H. A. Andersson. A method for monitoring of systems that operate in a repetitive manner - application to wear monitoring of an industrial robot joint. In *Proc. of the 2011 PAPYRUS Workshop*, Corsica, France, Oct 2011b.

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

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

L. C. Bo and D. Pavelescu. The friction-speed relation and its influence on the critical velocity of stick-slip motion. *Wear*, 82(3):277 – 289, 1982.

B. Bona and M. Indri. Friction compensation in robotics: an overview. In *Proc. of the 44th IEEE International Conference on Decision and Control*, Dec 2005.

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.

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

T. Brogårdh. Present and future robot control development - an industrial perspective. *Annual Reviews in Control*, 31:69–79, 2007.

T. Brogårdh. Robot Control Overview: An Industrial Perspective. *Modeling, Identification and Control*, 30(3):167–180, 2009.

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.

Y. Cheng, I. Izadi, and T. Chen. On optimal alarm filter design. In *Proc. of the 2011 International Symposium on Advanced Control of Industrial Processes*, pages 139 –145, may 2011.

J. de Kleer and B. C. Williams. *Diagnosis with behavioral modes*, pages 124–130. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1992.

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

M. M. Deza and E. Deza. *Encyclopedia of Distances*. Springer Berlin Heidelberg, 2009.

S. X. Ding. *Model-based Fault Diagnosis Techniques: Design Schemes, Algorithms and Tools*. Springer, 2008.

X. Ding, L. Guo, and P. Frank. Parameterization of linear observers and its application to observer design. *Automatic Control, IEEE Transactions on*, 39(8): 1648 –1652, aug 1994.

D. Dowson. *History of Tribology*. Professional Engineering Publishing, London, UK., 1998.

P. Dupont, V. Hayward, B. Armstrong, and F. Altpeter. Single state elastoplastic friction models. *Automatic Control, IEEE Transactions on*, 47(5):787 –792, may 2002.

P. Frank and X. Ding. Survey of robust residual generation and evaluation methods in observer-based fault detection systems. *Journal of Process Control*, 7(6): 403 – 424, 1997.

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

M. Gautier and P. Poignet. Extended kalman filtering and weighted least squares dynamic identification of robot. *Control Engineering Practice*, 9(12):1361 – 1372, 2001.

J. Gertler. *Fault Detection and Diagnosis in Engineering Systems*. CRC Press, 1998.

J. Gertler and J. Cao. Pca-based fault diagnosis in the presence of control and dynamics. *AIChE Journal*, 50(2):388–402, 2004.

J. Gray, A. and J. Markel. Distance measures for speech processing. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 24(5):380 – 391, oct 1976.

Y. Guo, Z. Qu, Y. Braiman, Z. Zhang, and J. Barhen. Nanotribology and nanoscale friction. *Control Systems Magazine, IEEE*, 28(6):92 –100, dec. 2008.

F. Gustafsson. *Adaptive Filtering and Change Detection*. Wiley, Oct. 2000.

A. Harnoy, B. Friedland, and S. Cohn. Modeling and measuring friction effects. *Control Systems Magazine, IEEE*, 28(6), Dec. 2008.

J. M. Hollerbach. *A survey of kinematic calibration*, pages 207–242. MIT Press, Cambridge, MA, USA, 1989.

H. Hotelling. Analysis of a complex of statistical variables into principal components. *J. Educ. Psych.*, 24, 1933.

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

ISO 8373. *Manipulating industrial robots – Vocabulary*. ISO, Geneva, Switzerland, 1994.

B. Jacobson. The stribeck memorial lecture. *Tribology International*, 36(11):781 – 789, 2003.

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.

K. Karhunen. Über lineare methoden in der wahrscheinlichkeitsrechnung. *Ann. Acad. Sci. Fenn., Ser. A.1.: Math.-Phys.*, 37:3–79, 1947.

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

H. M. Kim, S. H. Park, and S. I. Han. Precise friction control for the nonlinear friction system using the friction state observer and sliding mode control with recurrent fuzzy neural networks. *Mechatronics*, 19(6):805 – 815, 2009.

M. Krysander and E. Frisk. Sensor placement for fault diagnosis. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 38(6): 1398 –1410, nov. 2008.

A. R. Lansdown, A. L. Price, and J. Larsen-Basse. Materials to resist wear—a guide to their selection and use. *Journal of Tribology*, 109(2):379–380, 1987.

X. Li. *Fault Detection Filter Design for Linear Systems*. PhD thesis, Dept. of Elec. & Computer Eng., Louisiana State University, jun 2009.

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

N. Liu. *Optimal Robust Fault Detection*. PhD thesis, Dept. of Elec. & Computer Eng., Louisiana State University, may 2008.

N. Liu and K. Zhou. Optimal robust fault detection for linear discrete time systems. *J. Control Sci. Eng.*, 2008:7:1–7:16, January 2008.

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

M. Loève. *Fonctions aléatoires du second ordre.* Gauthier-Villars, Paris,France, 1948.

J. Machalíková, M. Sejkorová, J. Chýlková, and E. Schmidová. Application of tribodiagnostics in the maintenance of vehicles. In *Proc. of the 5th International Conference on Theoretical and Practical Issues in Transport*, Feb 2010.

D. J. C. Mackay. *Information Theory, Inference and Learning Algorithms.* Cambridge University Press, 1st edition, June 2003.

R. Mattone and A. D. Luca. Relaxed fault detection and isolation: An application to a nonlinear case study. *Automatica*, 42(1):109 – 116, 2009.

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

S. Moberg. *Modeling and Control of Flexible Manipulators.* Linköping studies in science and technology. dissertations. no. 1349, Linköping Studies in Science and Technology, SE-581 83 Linköping, Sweden, Dec. 2010.

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.

R. K. Mobley. *An Introduction to Predictive Maintenance, Second Edition.* Butterworth-Heinemann, Oct. 2002.

Y.-R. Mohamed and E. El-Saadany. A current control scheme with an adaptive internal model for torque ripple minimization and robust current regulation in pmsm drive systems. *Energy Conversion, IEEE Transactions on*, 23(1):92 –100, march 2008.

M. Namvar and F. Aghili. Failure detection and isolation in robotic manipulators using joint torque sensors. *Robotica*, 2009.

M. Nyberg. *Model Based Fault Diagnosis: Methods, Theory, and Automotive Engine Applications.* PhD thesis, Linköpings Universitet, June 1999.

M. Nyberg. A generalized minimal hitting-set algorithm to handle diagnosis with behavioral modes. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 41(1):137 –148, jan. 2011.

H. Olsson, K. J. Åström, C. C. de Wit, M. Gafvert, and P. Lischinsky. Friction models and friction compensation. *European Journal of Control*, 4(3):176–195, 1998.

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

D. Peretzki, A. J. Isaksson, A. C. Bittencourt, and K. Forsman. Data mining of historic data for process identification. In *Proc. of the 2011 AIChE Annual Meeting*, oct. 2011.

A. Proca, A. Keyhani, A. El-Antably, W. Lu, and M. Dai. Analytical model for permanent magnet motors with surface mounted magnets. *Energy Conversion, IEEE Transactions on*, 18(3):386 – 391, sept. 2003.

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

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

R. Reiter. A theory of diagnosis from first principles. *Artif. Intell.*, 32:57–95, April 1987.

B. Roylance. Ferrography–then and now. *Tribology International*, 38(10):857 – 862, 2005.

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. doi: 10.1109/CCA.2009.5281173.

L. Sciavicco and B. Siciliano. *Modelling and control of robot manipulators*. Advanced textbooks in control and signal processing. Springer, 2000.

T. Sebastian. Temperature effects on torque production and efficiency of pm motors using ndfeb magnets. *Industry Applications, IEEE Transactions on*, 31(2): 353–357, oct 1995.

C. J. Seeton. Viscosity-temperature correlation for liquids. *Tribology Letters*, 22 (1):67–78, Mar. 2006.

W. Seifert and V. Westcott. A method for the study of wear particles in lubricating oil. *Wear*, 21(1):27 – 42, 1972.

SKF. Interactive engineering catalogue, Aug. 2011. URL http://www.skf.com/portal/skf/home/products?newlink=first&lang=en.

M. Spong, S. Hutchinson, and M. Vidyasagar. *Robot modeling and control.* John Wiley & Sons, 2006.

M. W. Spong. Modeling and control of elastic joint robots. *Journal of Dynamic Systems, Measurement, and Control*, 109(4):310–318, 1987.

R. Stribeck. Die wesentlichen eigenschaften der gleit-und rollenlager – the key

qualities of sliding in roller bearings. *Zeitschrift Des Vereines Deutcher Inge-nieure*, 46(36–38):1342–1348, 1432–1437, 1902.

W. Susanto, R. Babuska, F. Liefhebber, and T. van der Weiden. Adaptive friction compensation: application to a robotic manipulator. In *Proc. of 17th IFAC World Congress*, Dec 2008.

C. Svärd and M. Nyberg. Automated design of an FDI-system for the wind tur-bine benchmark. In *Proc of the 18th IFAC World Congress*, Milano, Italy, 2011.

J. Swevers, C. Ganseman, D. Tukel, J. de Schutter, and H. Van Brussel. Optimal robot excitation and identification. *Robotics and Automation, IEEE Transac-tions on*, 13(5):730 –740, oct 1997.

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

G. Thompson. *Improving maintainability and reliability through design*. Wiley, 1 edition, jan 1999.

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

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

V. Venkatasubramanian, R. Rengaswamy, and S. N. Kavuri. A review of process fault detection and diagnosis: Part ii: Qualitative models and search strategies. *Computers & Chemical Engineering*, 27(3):313 – 326, 2003a.

V. Venkatasubramanian, R. Rengaswamy, S. N. Kavuri, and K. Yin. A review of process fault detection and diagnosis: Part iii: Process history based methods. *Computers & Chemical Engineering*, 27(3):327 – 346, 2003b.

V. Venkatasubramanian, R. Rengaswamy, K. Yin, and S. N. Kavuri. A review of process fault detection and diagnosis: Part i: Quantitative model-based meth-ods. *Computers & Chemical Engineering*, 27(3):293 – 311, 2003c.

J. Wallén. The history of the industrial robot. Technical Report LiTH-ISY-R-2853, Department of Electrical Engineering, Linköping University, SE-581 83 Linköping, Sweden, May 2008.

E. Wernholt. *Multivariable Frequency-Domain Identification of Industrial Robots*. Linköping studies in science and technology. dissertations. no. 1138, Linköping Studies in Science and Technology, SE-581 83 Linköping, Sweden, Nov. 2007.

E. Wernholt and S. Gunnarsson. Nonlinear identification of a physically parame-terized robot model. In *Proc. of the 14th IFAC Symposium on System Identifi-cation*, pages 143–148, Newcastle, Australia, Mar. 2006.

E. Wernholt and S. Moberg. Nonlinear gray-box identification using local models applied to industrial robots. *Automatica*, 47(4):650 – 660, 2011.

M. Woydt and R. Wäsche. The history of the stribeck curve and ball bearing steels: The role of adolf martens. *Wear*, 268(11-12):1542 – 1546, 2010.

J. Wu, J. Wang, and Z. You. An overview of dynamic parameter identification of robots. *Robotics and Computer-Integrated Manufacturing*, 26(5):414 – 419, 2010.

L. C. Yee and C. Jim. Foxconn to rely more on robots: could use 1 million in 3 years, Aug. 2011. URL http://www.reuters.com/article/2011/08/01/us-foxconn-robots-idUSTRE77016B20110801.

S. Yin, X. Steven, A. Naik, P. Deng, and A. Haghani. On pca-based fault diagnosis techniques. In *Proc. of the 2010 Conference on Control and Fault-Tolerant Systems*, pages 179 –184, oct. 2010.

S. Yoon and J. F. MacGregor. Principal-component analysis of multiscale data for process monitoring and fault diagnosis. *AIChE Journal*, 50(11):2891–2903, 2004.

Part II

Publications

Paper A

Static Friction in a Robot Joint – Modeling and Identification of Load and Temperature Effects

Authors: André Carvalho Bittencourt and Svante Gunnarsson

Static Friction in a Robot Joint – Modeling and Identification of Load and Temperature Effects

André Carvalho Bittencourt and Svante Gunnarsson

Dept. of Electrical Engineering,
Linköping University,
SE–581 83 Linköping, Sweden
{andrecb,svante}@isy.liu.se

Abstract

Friction is the result of complex interactions between contacting surfaces in down to a nanoscale perspective. Depending on the application, the different models available are more or less suitable. Static friction models are typically considered to be dependent only on relative speed of interacting surfaces. However, it is known that friction can be affected by other factors than speed.

In this paper, the typical friction phenomena and models used in robotics are reviewed. It is shown how such models can be represented as a sum of functions of relevant states which are linear and nonlinear in the parameters, and how the identification method described in Golub and Pereyra (1973) can be used to identify them when all states are measured. The discussion follows with a detailed experimental study of friction in a robot joint under changes of joint angle, load torque and temperature. Justified by their significance, load torque and temperature are included in an extended static friction model. The proposed model is validated in a wide operating range, considerably improving the prediction performance compared to a standard model.

1 Introduction

Friction exists in all mechanisms to some extent. It can be defined as the tangential reaction force between two surfaces in contact. It is a nonlinear phenomenon which is physically dependent on contact geometry, topology, properties of the materials, relative velocity, lubricant, etc. (Al-Bender and Swevers (2008)). Friction has been constantly investigated by researchers due to its importance in several fields, Dowson (1998). In this paper, friction has been studied based on experiments on an industrial robot.

One reason for the interest in friction of manipulator joints is the need to model friction for control purposes (Kim et al. (2009); Guo et al. (2008); Olsson et al. (1998); Bona and Indri (2005); Susanto et al. (2008)), where a precise friction model can considerably improve the overall performance of a manipulator with respect to accuracy and control stability. Since friction can relate to the wear down process of mechanical systems (Blau (2009)), including robot joints (Bittencourt et al. (2011)), there is also interest in friction modeling for robot condition monitoring and fault detection, Bittencourt et al. (2011); Caccavale et al. (2009); Namvar and Aghili (2009); McIntyre et al. (2005); Vemuri and Polycarpou (2004); Brambilla et al. (2008); Mattone and Luca (2009); Freyermuth (1991).

A friction model consistent with real experiments is necessary for successful simulation, design and evaluation. Due to the complexity of friction, it is however often difficult to obtain models that can describe all the empirical observations (see Al-Bender and Swevers (2008) for a comprehensive discussion on friction physics and first principle friction modeling). In a robot joint, the complex interaction of components such as gears, bearings and shafts which are rotating/sliding at different velocities, makes physical modeling difficult. An example of an approach to model friction of complex transmissions can be found in Waiboer (2007), where the author designs joint friction models based on physical models of elementary joint components as helical gear pairs and pre-stressed roller bearings.

Empirically motivated friction models have been successfully used in many applications, including robotics, see e.g. Armstrong-Hélouvry (1991); Olsson et al. (1998); Åström and Canudas-de Wit (2008); Harnoy et al. (2008). This category of models was developed through time according to empirical observations of the phenomenon, Dowson (1998). Considering a set of states $\mathcal{X}$, and parameters θ, these model structures $\mathcal{M}$, can be described as the sum of M functions ϕ_j that describe the behavior of friction $\mathcal{F}$,

$$\mathcal{F}(\mathcal{X}, \theta) = \sum_{j=1}^{M} \phi_j(\mathcal{X}, \theta). \tag{$\mathcal{M}$}$$

The choice $\mathcal{X} = [z, \dot{q}, q]$, where z is an internal state related to the dynamic behavior of friction, q is a generalized coordinate and $\dot{q} = \frac{d}{dt}q$, gives the set of Generalized Empirical Friction Model structures (GEFM), see Al-Bender and Swevers (2008).

Among the GEFM model structures, the LuGre model (Olsson et al. (1998); Åström and Canudas-de Wit (2008)) is a common choice in the robotics community. For a revolute joint, the LuGre model structure $\mathcal{M}_L$, can be described as

$$\tau_f = \sigma_0 z + \sigma_1 \dot{z} + h(\dot{\varphi}) \tag{$\mathcal{M}_L$}$$

$$\dot{z} = \dot{\varphi} - \sigma_0 \frac{|\dot{\varphi}|}{g(\dot{\varphi})} z,$$

where τ_f is the friction torque, φ is the joint motor angle and $\dot{\varphi} = \frac{d}{dt}\varphi$. The state z is related to the dynamic behavior of asperities in the interacting surfaces and

can be interpreted as their average deflection, with stiffness σ_0 and damping σ_1.

The function $h(\dot{\varphi})$ represents the velocity strengthening (viscous) friction and is dependent on the stress versus strain rate relationship. For Newtonian fluids, the shear stress follows a linear dependency to the shear rate $\tau = \mu \frac{\mathrm{d}}{\mathrm{d}y} u$, where τ is the shear stress, μ is the viscosity and $\frac{du}{dy}$ is the velocity gradient perpendicular to the direction of shear. It is typical to consider a Newtonian behavior, yielding the relationship

$$h(\dot{\varphi}) = F_v \dot{\varphi}$$

for the viscous behavior of friction.

The function $g(\dot{\varphi})$ captures the velocity weakening of friction. Motivated by the observations mainly attributed to Stribeck (Jacobson (2003); Woydt and Wäsche (2010); Bo and Pavelescu (1982)), $g(\dot{\varphi})$ is usually modeled as

$$g(\dot{\varphi}) = F_c + F_s e^{-\left|\frac{\dot{\varphi}}{\dot{\varphi}_s}\right|^\alpha},$$

where F_c is the Coulomb friction, F_s is called the standstill friction parameter[i], $\dot{\varphi}_s$ is the Stribeck velocity and α is the exponent of the Stribeck nonlinearity. The model structure $\mathcal{M}_L$ is a GEFM with $\mathcal{X} = [z, \dot{\varphi}]$ and $\theta = [\sigma_0, \sigma_1, F_c, F_s, F_v, \dot{\varphi}_s, \alpha]$. According to Åström and Canudas-de Wit (2008) it can successfully describe many of the friction characteristics.

Since z is not measurable, a difficulty with $\mathcal{M}_L$ is the estimation of the dynamic parameters $[\sigma_0, \sigma_1]$. In Olsson et al. (1998), these parameters are estimated in a robot joint by means of open loop experiments and by use of high resolution encoders. Open-loop experiments are not always possible, and it is common to accept only a static description of $\mathcal{M}_L$. For constant velocities, $\mathcal{M}_L$ is equivalent to the static model $\mathcal{M}_S$:

$$\tau_f(\dot{\varphi}) = g(\dot{\varphi})\mathrm{sign}(\dot{\varphi}) + h(\dot{\varphi}) \tag{$\mathcal{M}_S$}$$

which is fully described by the g- and h functions. In fact, $\mathcal{M}_L$ simply adds dynamics to $\mathcal{M}_S$. The typical choice for $g-$ and h, as defined previously for $\mathcal{M}_L$, yields the static model structure $\mathcal{M}_0$:

$$\tau_f(\dot{\varphi}) = \left[F_c + F_s e^{-\left|\frac{\dot{\varphi}}{\dot{\varphi}_s}\right|^\alpha} \right] \mathrm{sign}(\dot{\varphi}) + F_v \dot{\varphi}. \tag{$\mathcal{M}_0$}$$

$\mathcal{M}_0$ requires a total of four parameters to describe the velocity weakening regime $g(\dot{\varphi})$ and one parameter to capture viscous friction $h(\dot{\varphi})$. See Fig. 3 for an interpretation of the parameters.

From empirical observations, it is known that friction can be affected by several factors, e.g.:

---

[i]F_s is commonly called static friction. An alternative nomenclature was adopted to make a distinction between the dynamic/static friction phenomena.
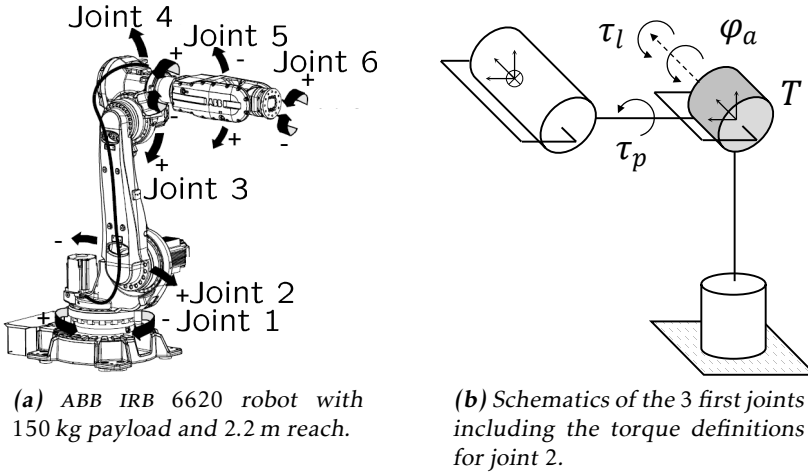
(a) ABB IRB 6620 robot with 150 kg payload and 2.2 m reach.

(b) Schematics of the 3 first joints including the torque definitions for joint 2.

Figure 1: *The experiments were made on joint 2 of the ABB robot IRB 6620.* φ_a *is the joint angle, T the joint temperature, τ_l the manipulation load torque and τ_p the perpendicular load torque.*

- temperature,
- force/torque levels,
- position,

- velocity,
- acceleration,
- lubricant properties.

A shortcoming of the LuGre model structure, as with any GEFM, is the dependence only of the states $\mathcal{X} = [z, \dot{q}, q]$. In more demanding applications, the effects of the remaining variables can not be neglected. In Waiboer et al. (2005), the author observes a strong temperature dependence, while in Olsson et al. (1998), joint load torque and temperature are considered as disturbances and estimated in an adaptive framework. In Gogoussis and Donath (1988); Dohring et al. (1993), the effects of load are modeled as a linear effect on F_c in a model structure similar to $\mathcal{M}_0$. In the recent contribution of Hamon et al. (2010) the load effects are revisited to include also a linear dependency on F_s. However, more work is needed in order to understand the influence of different factors on the friction properties. A more comprehensive friction model is needed to improve tasks related to design, simulation and evaluation for machines with friction.

The objective of this paper is to analyze and model the effects in static friction related to joint angle, load torques and temperature. The phenomena are observed in joint 2 of an ABB IRB 6620 industrial robot, see Figure 1a. Two load torque components are examined, the torque aligned to the joint DOF (degree of freedom) and the torque perpendicular to the joint DOF. These torques are in the paper named manipulation load torque τ_l and perpendicular load torque τ_p, see Figure 1b.

By means of experiments, these variables are analyzed and modeled based on the empirical observations. The task of modeling is to find a suitable model structure

according to:

$$\tau_f(\mathcal{X}^*, \theta) = \sum_{j=1}^{M} \phi_j(\mathcal{X}^*, \theta) \qquad\qquad (\mathcal{M}^*)$$

$$\mathcal{X}^* = \begin{bmatrix} \dot{\varphi}, & \varphi_a, & \tau_p, & \tau_l, & T \end{bmatrix},$$

where T is the joint (more precisely, lubricant) temperature and φ_a the joint angle at the arm side.

Ideally, the chosen model should be coherent with the empirical observations and, simultaneously, with the lowest dimension of θ, the parameter vector, and with the lowest number of describing functions (minimum M). For practical purposes, the choice of ϕ_i should also be suitable for a useful identification procedure.

The work presented here is based on Bittencourt et al. (2010), where a friction model was proposed to describe the effects of load and temperature in a robot joint. More detailed analysis of the modeling assumptions are presented, together with a more general framework for identification of friction models. The paper is organized as follows. Section 2 presents the method used to estimate static friction levels in a robot joint and consequently its friction curve, an identification procedure is also described for general parametric description of friction curves and some model simplifications are justified. Section 3 contains the major contribution of this paper, with the empirical analysis, modeling and validation. Conclusions and future work are presented in Section 4.

2   Static Friction Curve

Static friction is typically presented in a *friction curve*, a plot of static friction levels against speed. It is related to the Stribeck curve under the simplification that viscosity and contact pressure are constant. An example of a friction curve estimated from a robot joint can be seen in Figure 3.

From a phenomenological perspective, a friction curve can be divided into three regimes, according to the lubrication characteristics: *boundary (BL), mixed (ML) and elasto-hydrodynamic lubrication (EHL)*. The phenomena present in very low velocities (BL) is mostly related to interactions between the asperities of the surfaces in contact. With the increase of velocity, there is a consequent increase of the lubrication film between the surfaces and a decrease of friction (ML) until it reaches a full lubrication profile (EHL) with a total separation of the surfaces by the lubricant. In EHL, friction is proportional to the force needed to shear the lubricant layer, thus dependent on the lubricant properties, specially viscosity. Recalling the static friction model $\mathcal{M}_S$, the BL and ML regimes are described by the velocity weakening function $g-$ and the EHL regime is described by h.

In this section, an experimental procedure is suggested to estimate static friction levels at constant speeds in a robot joint and consequently its friction curve. Given static friction estimates, it is shown how the general friction model $\mathcal{M}$ can

be identified with the method described in Golub and Pereyra (1973), when the states $\mathcal{X}$ are available. Finally, the model structure $\mathcal{M}_0$ is simplified to achieve a minimal description of static friction.

2.1 Estimation Procedure

A manipulator is a multivariable, nonlinear system that can be described in a general manner through the rigid multi-body dynamic model

$$M(\varphi)\ddot{\varphi} + C(\varphi, \dot{\varphi}) + \tau_g(\varphi) + \tau_f(\dot{\varphi}) = u, \tag{1}$$

where $M(\varphi)$ is the inertia matrix, $C(\varphi, \dot{\varphi})$ relates to speed dependent terms (Coriolis and centrifugal), $\tau_g(\varphi)$ are the gravity-induced torques at the joints and τ_f contains the friction torques at joints. The system is controlled by the input torque, u, applied by the joint motor (in the experiments the torque reference from the servo was measured[i]).

For single joint movements ($C(\varphi, \dot{\varphi}) = 0$) under constant speed ($\ddot{\varphi} \approx 0$), Equation (1) simplifies to

$$\tau_g(\varphi) + \tau_f = u. \tag{2}$$

The resulting applied torque u drives only friction and gravity-induced torques. The required torques to drive a joint in forward, u^+, and reverse, u^-, directions at the constant speed level $\bar{\dot{\varphi}}$, and at a joint angle value $\bar{\varphi}$ (so that $\tau_g(\bar{\varphi})$ is equal in both directions), are

$$\tau_f(\bar{\dot{\varphi}}) + \tau_g(\bar{\varphi}) = u^+ \qquad \tau_f(-\bar{\dot{\varphi}}) + \tau_g(\bar{\varphi}) = u^-. \tag{3}$$

In the case an estimate of $\tau_g(\bar{\varphi})$ is available, it is possible to isolate the friction component in each directions using Equation (3). If such estimate is not possible (e.g. not all masses are completely known), τ_f can still be achieved in the case that τ_f is independent of the rotation direction. Subtracting the equations yields

$$\tau_f(\bar{\dot{\varphi}}) - \tau_f(-\bar{\dot{\varphi}}) = u^+ - u^-$$

and if $\tau_f(-\bar{\dot{\varphi}}) = -\tau_f(\bar{\dot{\varphi}})$, the resulting *direction independent friction* is:

$$\tau_f(\bar{\dot{\varphi}}) = \frac{u^+ - u^-}{2}. \tag{4}$$

In the experiments, each joint is moved separately with the desired speed in both directions around a given joint angle $\bar{\varphi}$. Figure 2 shows the measured joint angle-, speed- and torque[ii] signals sampled at 2 kHz[iii] for $\bar{\dot{\varphi}} = 42$ rad/s around $\bar{\varphi} = 0$. The constant speed data is segmented around $\bar{\varphi}$ and the static friction levels can be achieved using Equation (3) or (4).

---

[i]It is known that using the torque reference from the servo as a measure of the joint torque might not always hold because of the temperature dependence of the torque constant of the motors. The deviations are however considered to be small and are neglected during the experiments.

[ii]Throughout the paper all torques are normalized to the maximum manipulation torque at low speed.

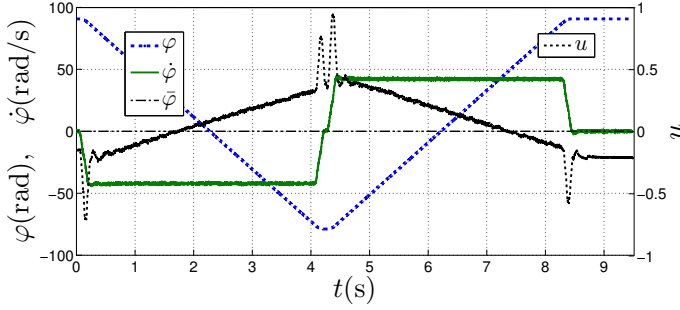[iii]Similar results have been experienced with sampling rates down to 220 Hz.

Figure 2: *Excitation signals used for the static friction estimation at $\dot{\varphi} = 42$ rad/s.*
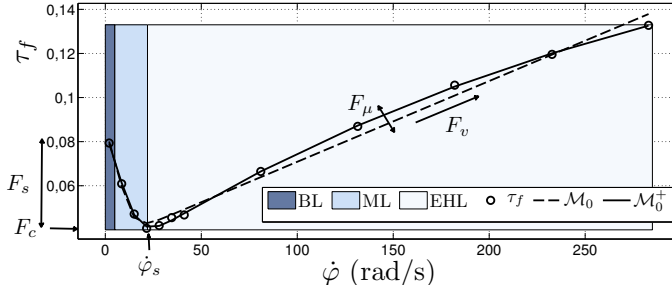

Figure 3: *Static friction curve with lubrication regimes and model-based predictions. Circles indicate friction levels achieved using Equation (4).*

The procedure can be repeated for several different speeds and a friction curve can be drawn. As shown in Bittencourt et al. (2010), there is only a small direction dependency of friction for the investigated joint. Therefore, in this paper, friction levels are achieved using Equation (4), which is not influenced by deviations in the gravity model of the robot.

2.2   General Parametric Description and Identification

The general friction models described by $\mathcal{M}$ can be written as

$$\hat{\tau}_f(\mathcal{X}_k, \theta) = \sum_{j=1}^{N_\eta} \phi_j(\mathcal{X}_k, \rho)\eta_j. \tag{5}$$

where the index k relates to the k-th measurement in the data set. The parameters' vector $\theta = \left[\eta^T, \rho^T\right]^T$ has dimension $(N_\eta + N_\rho)$ and is divided according to the manner they appear in the model, respectively linearly/nonlinearly. Notice that if there are no linear dependency in the parameters, i.e. η is empty, (5) reads directly as $\mathcal{M}$ by taking $\theta = \rho$. As it will be shown, the structure of (5), can be exploited when defining an identification method.

Considering a total of N measurements, the residuals (innovations) between predictions and measurements are written as $\varepsilon(k, \theta) = \tau_f(k) - \hat{\tau}_f(\mathcal{X}_k, \theta)$. For the following discussion, it is assumed that $\mathcal{X}_k$ is available so that it is possible to construct $\phi_j(\mathcal{X}_k, \rho)$.

The identification objective can be formulated as a prediction error minimization

$$\hat{\theta} = \arg \min_{\theta} V(\varepsilon(k, \theta)) \tag{6}$$

where $\hat{\theta}$ is the value that minimizes the cost function $V(\cdot)$. The least squares problem considers the minimization of

$$\hat{\theta} = \arg \min_{\theta} V(\varepsilon(k, \theta)) = \arg \min_{\theta} \sum_{k=1}^{N} \varepsilon^2(k, \theta). \tag{7}$$

The minimum of (7) occurs where the gradient of the innovations, $\psi(k, \theta) = \frac{\partial}{\partial \theta} \varepsilon(k, \theta)$, is zero. For the model in (5), this gradient takes the form

$$\psi(k, \theta) = \left[\phi_1(\mathcal{X}_k, \rho), \cdots, \phi_{N_\eta}(\mathcal{X}_k, \rho), \frac{\partial}{\partial \rho_1} \hat{\tau}_f(\mathcal{X}_k, \theta), \cdots, \frac{\partial}{\partial \rho_{N_\rho}} \hat{\tau}_f(\mathcal{X}_k, \theta) \right]^T, \tag{8}$$

where it is easy to realize the separable nature of the model. The solution for ρ can not be found explicitly, but can be solved numerically using an optimization routine. For instance, if η is empty, gradient based methods can be used to find an estimate of ρ (Ljung (1998)).

As presented in Golub and Pereyra (1973), the separable structure of the model can be explored. Defining the matrix $\{\Phi(\rho)\}_{k,j} = \phi_j(\mathcal{X}_k, \rho)\eta_j$, for any given ρ, the solution for η, is given by the least squares solution

$$\hat{\eta} = \Phi^{\dagger}(\rho)\boldsymbol{\tau}_f, \quad \Phi^{\dagger}(\rho) = \left[\Phi^T(\rho)\Phi(\rho) \right]^{-1} \Phi^T(\rho) \tag{9}$$

where $\boldsymbol{\tau}_f$ denotes here the vector of measurements $\{\tau_f(k)\}_1^N$ and $\Phi^{\dagger}(\rho)$ is the Moore-Penrose pseudoinverse, Golub and Pereyra (1973). Substituting this back in (7), the problem can be rewritten as a function only of ρ

$$\hat{\rho} = \arg \min_{\rho} \|\boldsymbol{\tau}_f - \Phi(\rho)\hat{\eta}\|^2 = \arg \min_{\rho} \|P^{\perp}_{\Phi(\rho)}\boldsymbol{\tau}_f\|^2, \tag{10}$$

where $P^{\perp}_{\Phi(\rho)} = I - \Phi(\rho)\Phi^{\dagger}(\rho)$ is the projector on the orthogonal complement of the column space of $\Phi(\rho)$. The idea is then to first find $\hat{\rho}$, and then plug it back in (9) to find $\hat{\eta}$. This illustrates the algorithm proposed in Golub and Pereyra (1973), where it is also shown that the resulting point $\hat{\theta} = [\hat{\eta}^T, \hat{\rho}^T]^T$ is a minimum of (7).

There is, however, no closed form solution to (10). An approach is to consider gradient based methods where information of the gradient of $P^{\perp}_{\Phi(\rho)}\boldsymbol{\tau}_f$ is relevant. In Golub and Pereyra (1973), it is shown that the gradient of $P^{\perp}_{\Phi(\rho)}\boldsymbol{\tau}_f$ requires only computation of derivatives of $\Phi(\rho)$, as in (8) (see Golub and Pereyra (1973) for a detailed treatment). In this work, a 2-step identification procedure is used, in a initial step, a coarse grid search is used to find initial estimates of ρ. The problem

Table 1: *Identified $\mathcal{M}_0$ parameters for the data shown in Figure 3.*

$F_c\,[\,10^{-2}\,]$	$F_s\,[\,10^{-2}\,]$	$F_v\,[\,10^{-4}\,]$	$\dot{\varphi}$	α
3.4 ± 0.176	4.6 ± 0.48	3.68 ± 0.12	10.68 ± 1.08	1.93 ± 0.60

(10) is then solved given the initial estimates using a trust-region reflective algorithm available in the Matlab's Optmization Toolbox. The resulting $\hat{\rho}$ estimate is finally used to find $\hat{\eta}$ as is in (9).

To assess the resulting performance of the identification procedure, it is possible to provide an estimate of the identified parameters uncertainties. For any unbiased estimator, the following relationship for its covariance P_{θ_N} holds, Ljung (1998),

$$P_{\theta_N} \geq \kappa_0 \left[\sum_{k=1}^{N} \mathbb{E}\left[\psi(k, \theta) \psi^T(k, \theta) \right] \right]^{-1} = P_{\theta_N}^* \tag{11}$$

where for Gaussian innovations with variance γ_0, $\kappa_0 = \gamma_0$ (Ljung (1998)). Under this assumption, given an estimate $\hat{\theta}_N$ of θ after N observations, $P_{\theta_N}^*$ can be estimated from the data as

$$\hat{P}_{\theta_N}^* = \hat{\gamma}_N \left[\frac{1}{N} \sum_{k=1}^{N} \psi(k, \hat{\theta}_N) \psi^T(k, \hat{\theta}_N) \right]^{-1} \tag{12}$$

$$\hat{\gamma}_N = \frac{1}{N} \sum_{k=1}^{N} \varepsilon^2(k, \hat{\theta}_N). \tag{13}$$

The quantity in (12) is used throughout this work as a covariance estimate for $\hat{\theta}$.

For the model structure $\mathcal{M}_0$ in the first quadrant, Equation (5) can be written as

$$\mathcal{X} = \dot{\varphi}, \quad \Phi(\dot{\varphi}, \rho) = \left[1,\ e^{-\left| \frac{\dot{\varphi}}{\dot{\varphi}_s} \right|^{\alpha}},\ \dot{\varphi} \right]$$

$$\eta = [F_c,\ F_s,\ F_v], \quad \rho = [\dot{\varphi}_s,\ \alpha].$$

The model parameters are identified using the direction independent data (circles) in Figure 3. The resulting identified parameters values are shown in Table 1 with one standard deviation. The dashed line in Figure 3 is obtained by model-based predictions of the resulting model, with sum of absolute prediction errors smaller than $3.0\,10^{-2}$.

A closer investigation of the friction curve in Figure 3 reveals that the behavior of friction at high speeds is slightly nonlinear with speed. This feature is related to the non-Newtonian behavior of the lubricant at high speeds, see e.g. Waiboer et al. (2005). In this case, the fluid presents a pseudoplastic behavior, with a decrease of the apparent viscosity (increase of friction) with share rate (joint speed). The

behavior motivates the suggestion of an alternative model structure

$$\tau_f(\dot{\varphi}) = F_c + F_s e^{-\left|\frac{\dot{\varphi}}{\dot{\varphi}_s}\right|^\alpha} + F_v \dot{\varphi} + F_\mu \dot{\varphi}^\beta, \qquad (\mathcal{M}_0^+)$$

where F_μ and β relates to the non-Newtonian part of the viscous friction behavior and capture the deviation from a Newtonian behavior. The parameters are identified for the friction curve in Figure 3. The resulting predictions are shown by the solid line in Figure 3, with sum of absolute prediction error as $5.5 \, 10^{-3}$.

This example illustrates that it might be worth to consider the non-Newtonian behavior of the lubricant in applications where high accuracy is needed at high speeds. However, for simplicity, this behavior is not considered further in this paper.

2.3   Fixing α

Despite the non-Newtonian behavior of the lubricant, the model $\mathcal{M}_0$ represents well the behavior of static friction with speed. From a practical perspective, it is desirable to achieve a minimal number of parameters and avoid nonlinear terms which are costly to identify.

Following the general static friction description $\mathcal{M}_S$, the model $\mathcal{M}_0$ represents the decrease of friction in the velocity weakening regime, $g(\cdot)$, through the term $e^{-\left|\frac{\dot{\varphi}}{\dot{\varphi}_s}\right|^\alpha}$. The term takes two nonlinear parameters, α and $\dot{\varphi}_s$. It is common to accept α as a constant between 0.5 and 2 (Åström and Canudas-de Wit (2008); Olsson et al. (1998); Susanto et al. (2008)). As seen in Figure 4, $\dot{\varphi}_s$ changes the constant of the decay while α changes its curvature. Notice from Figure 4a and Figure 4b that small choices of α can considerably affect friction at high speeds. This is not desirable since these effects should be described by the velocity strengthening function h. For these reasons, α is fixed as presented next.

Considering all static friction data presented in this work, in a total of 488 friction curves with more than 5800 samples, α is chosen as the value minimizing Equation (7) for the model structure $\mathcal{M}_0$ when all other parameters are free at each friction curve. Figure 5 presents the resulting relative increase in the cost for different values of α. The value with minimal cost is $\alpha^* = 1.36 \pm 0.011$.

3   Empirically Motivated Modeling

Using the described static friction curve estimation method, it is possible to design a set of experiments to analyze how the states $\mathcal{X}^*$ affect static friction. As shown in Section 2.2, the model structure $\mathcal{M}_0$ can represent static friction dependence on $\dot{\varphi}$ fairly well. $\mathcal{M}_0$ is therefore taken as a primary choice, with α fixed at $\alpha^* = 1.36$. Whenever a single instance of $\mathcal{M}_0$ can not describe the observed friction behavior, extra terms $\phi_j(\mathcal{X}^*, \theta)$ are proposed and included in $\mathcal{M}_0$ to achieve a satisfactory model structure $\mathcal{M}^*$.
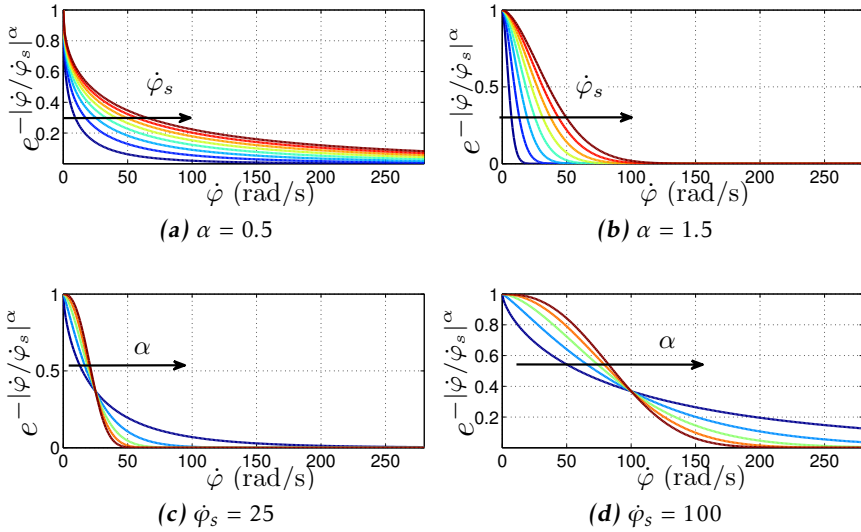
(a) $\alpha = 0.5$ *(b)* $\alpha = 1.5$

(c) $\dot{\varphi}_s = 25$ *(d)* $\dot{\varphi}_s = 100$

Figure 4: *Illustration of effects in the velocity weakening regime caused by*
$\dot{\varphi}_s$ *and* α*. Figures (a) and (b) with* $\dot{\varphi}_s = [1, \, 50]$ *rad/s. Figures (c) and (d) with*
$\alpha = [0.02, \, 3.00]$*.*
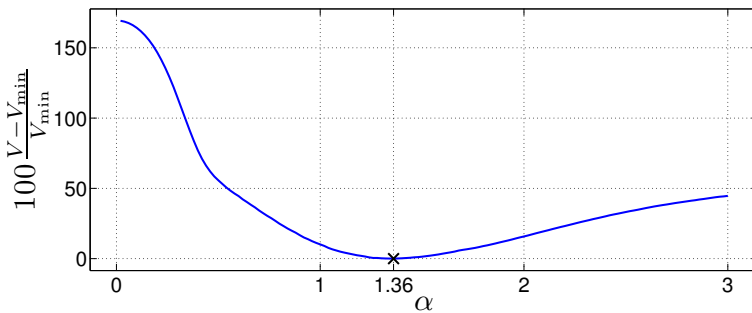
Figure 5: *Relative cost increase as a function of* α *for the model structure*
$\mathcal{M}_0$*.*

(a) Simulated τ_l.   *(b) Simulated τ_p.*
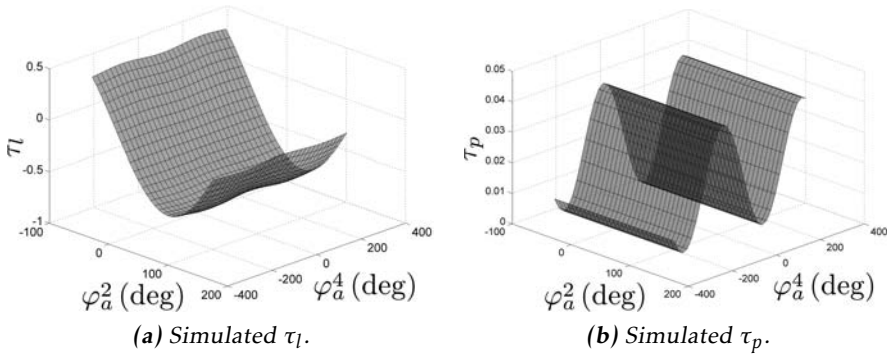
Figure 6: *Simulated load torques at joint 2 caused by angle variations of joints 2 and 4, φ_a^2 and φ_a^4 respectively. Notice the larger absolute values for τ_l when compared with τ_p.*

3.1   Guidelines for the Experiments

In order to be able to build a friction model including more variables than velocity, it is important to separate their influences. The situation is particularly critical regarding temperature as it is difficult to control it inside a joint. Moreover, due to the complex structure of an industrial robot, changes in joint angle might move the mass center of the robot arm system, causing variations of joint load torques. To avoid undesired effects, the guidelines below were followed during the experiments.

Isolating Joint Load Torque Dependency from Joint Angle Dependency

Using an accurate robot model[i], it is possible to predict the load torques at the joints for a given robot configuration (a set of all joints' angles). For example, Figure 6 shows the resulting τ_l and τ_p at joint 2 for configurations depending on the achievable angles for joints 2 and 4 (φ_a^2 and φ_a^4). Using this information, a set of configurations *can be selected a priori* in which it is possible to estimate parameters in an efficient way.

Isolating Temperature Effects

Some of the experiments require that the temperature in the joint is under control. Using joint lubricant temperature measurements[ii], the joint thermal decay constant κ was estimated to 3.04 h. By executing the static friction curve identification experiment periodically, for longer time than 2κ (i.e. > 6.08 h), the joint temperature is expected to have reached an equilibrium. Only data related to the expected thermal equilibrium was considered for the analysis.

---

[i]An ABB internal tool was used for simulation purposes.
[ii]In the studies, the robot gearbox was lubricated with oil, not grease, which gave an opportunity to obtain well defined temperature readings by having a temperature sensor in the circulating lubricant oil.
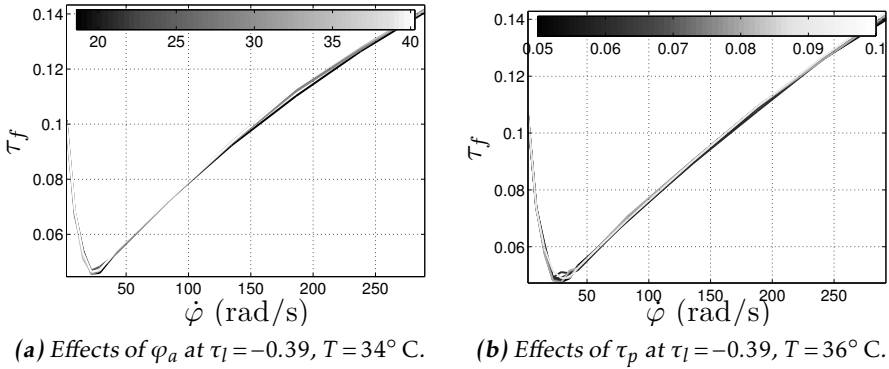
(a) Effects of φ_a at $\tau_l = -0.39$, $T = 34°$ C.   (b) Effects of τ_p at $\tau_l = -0.39$, $T = 36°$ C.

Figure 7: *Static friction curves for experiments related to φ_a and τ_p.*

3.2 Effects of Joint Angles

Due to asymmetries in the contact surfaces, it has been observed that the friction of rotating machines depends on the angular position, Al-Bender and Swevers (2008). It is therefore expected that this dependency occurs also in a robot joint. Following the experiment guidelines from the previous section, a total of 50 static friction curves were estimated in the joint angle range $\varphi_a = [8.40,\ 59.00]$ deg. As seen in Figure 7a, only small effects can be observed. The subtle deviations are comparable to the errors of the friction curve identified under constant values of $\left[\varphi_a,\ \tau_p,\ \tau_l,\ T\right]$. In fact, even a constant instance of $\mathcal{M}_0$ can describe the friction curves satisfactorily, no extra terms are thus required.

3.3 Effects of Load Torques

Since friction is related to the interaction between contacting surfaces, one of the first phenomenon observed was that friction varies according to the applied normal force. This can be explained by the increase of the true contact area between the surfaces under large normal forces. A similar reasoning can be extended to joint torques in a revolute robot joint. Due to the elaborated gear- and bearing design of the joint, it is also expected that torques in different directions will have different effects to the static friction curve[i].

Only small variations of τ_p, the perpendicular load torque, are achievable because of the mechanical construction of the robot, see Figure 6b. A total of 20 experiments at constant temperature were performed for joint 2, in the range $\tau_p = [0.04,\ 0.10]$. As Figure 7b shows, the influences of τ_p for the achievable range did not play a significant role for the static friction curve. The model $\mathcal{M}_0$ is thus considered valid over the achieved range of τ_p for this joint.

Large variations of τ_l, the manipulation load torque, are possible by simply vary-

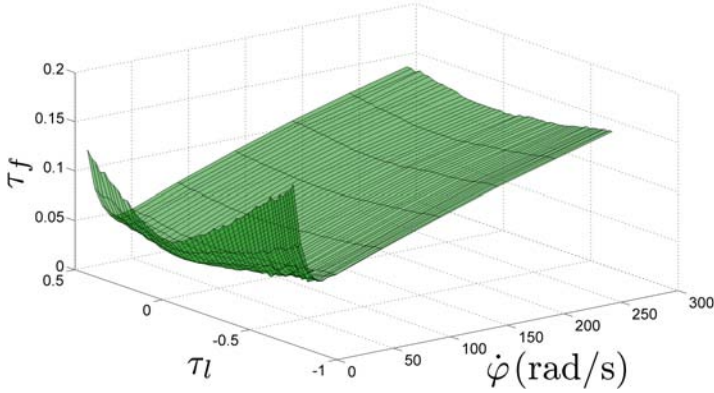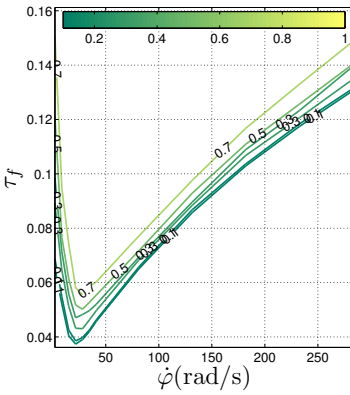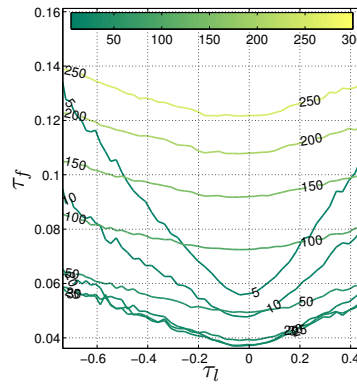[i]In fact, a full joint load description would require 3 torque and 3 force components.

(a) *Estimated friction curves for different values of τ_l.*

(b) *Friction surface cuts for different values of τ_l.*

(c) *Friction surface cuts for different values of $\dot\varphi$ rad/s.*

Figure 8: *The dependence of the static friction curves on the manipulation torque, τ_l, at $T = 34°$ C.*

ing the arm configuration, as seen in Figure 6a. A total of 50 static friction curves were estimated over the range $\tau_l = [-0.73, 0.44]$. As seen in Figure 8, the effects appear clearly.

Obviously, a single $\mathcal{M}_0$ instance can not describe the observed phenomena. A careful analysis of the effects reveals that the main changes occur in the velocity weakening part of the curve. From Figure 8c, it is possible to observe a (linear) bias-like (F_c) increase and a (linear) increase of the standstill friction (F_s) with $|\tau_l|$. Furthermore, as seen in Figure 8b, the Stribeck velocity $\dot\varphi_s$ is maintained fairly constant. The observations support an extension of $\mathcal{M}_0$ to

$$\tau_f(\dot\varphi, \tau_l) = \{F_{c,0} + F_{c,\tau_l}|\tau_l|\} + \{F_{s,0} + F_{s,\tau_l}|\tau_l|\}e^{-\left|\frac{\dot\varphi}{\dot\varphi_{s,\tau_l}}\right|^{\alpha^*}} + F_v\dot\varphi. \qquad (\mathcal{M}_1)$$

Table 2: *Identified τ_l-dependent model parameters.*

$F_{c,\tau_l}\left[10^{-2}\right]$	$F_{s,\tau_l}\left[10^{-1}\right]$	$\dot{\varphi}_{s,\tau_l}$
2.34 ± 0.071	1.26 ± 0.025	9.22 ± 0.12

In the above equation, the parameters are written with subscript $_0$ or $_{\tau_l}$ in order to clarify its origin related to $\mathcal{M}_0$ or to the effects of τ_l. The model structure $\mathcal{M}_1$ is similar to the one presented in Hamon et al. (2010), where the changes in F_c and F_s appear as linear functions of $|\tau_l|$.

Assuming that any phenomenon not related to τ_l is constant and such that the $_0$ terms can capture them, good estimates of the τ_l-dependent parameters can be achieved. The model $\mathcal{M}_1$ is identified with the data set from Figure 8 using the procedure described in Section 2.2. The resulting model parameters describing the dependency on τ_l are shown in Table 2.

3.4   Effects of Temperature

The friction temperature dependence is related to changes of both lubricant and contacting surfaces. In lubricated mechanisms, both the thickness of the lubricant layer and its viscosity play an important role for the resulting friction properties. In Newtonian fluids, the shear forces are directly proportional to the viscosity which, in turn, varies with temperature (Seeton (2006)). Dedicated experiments were made to analyze the effects of temperature. The joint was at first warmed up to $81.2°C$ by running the joint continuously back and forth. Then, while the robot cooled, 50 static friction curves were estimated over the range $T = [38.00, 81.20]\,°C$. In order to resolve combined effects of T and τ_l, two manipulation torque levels were used, $\tau_l = -0.02$, and $\tau_l = -0.72$. As it can be seen in Figure 9, the effects of T are significant.

Temperature has an influence on both velocity regions of the static friction curves. In the velocity-weakening region, a (linear) increase of the standstill friction (F_s) with temperature can be observed according to Figure 9b. In Figure 9c it can be seen that the Stribeck velocity ($\dot{\varphi}_s$) increases (linearly) with temperature. The effects in the velocity-strengthening region appear as a (nonlinear, exponential-like) decrease of the velocity-dependent slope, as seen in Figures 9b and 9c.

Combined effects of τ_l and T are also interesting to study. To better see these effects, the friction surfaces in Figure 9a are subtracted from each other, yielding $\tilde{\tau}_f$. As it can be seen in Figure 10a, the result is fairly temperature independent. This is an indication of *independence between effects caused by T and τ_l*.

Given that the effects of T and τ_l are independent, it is possible to subtract the τ_l-effects from the surfaces in Figure 9a and solely obtain temperature related phenomena. The previously proposed terms to describe the τ_l-effects in $\mathcal{M}_1$ were:

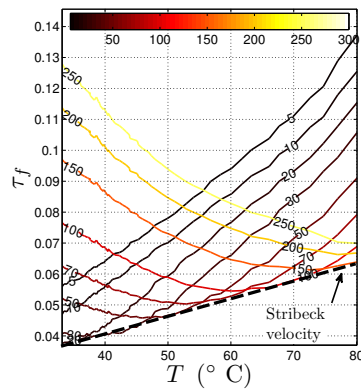$$\hat{\tau}_f(\tau_l) = F_{c,\tau_l}|\tau_l| + F_{s,\tau_l}|\tau_l|e^{-\left|\frac{\dot{\varphi}_m}{\dot{\varphi}_{s,\tau_l}}\right|^{\alpha^*}}. \tag{14}$$
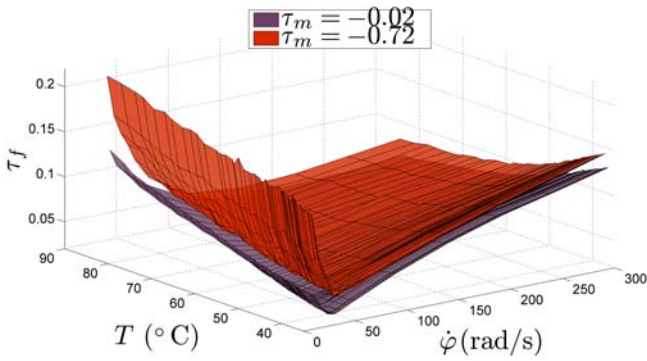
(a) Estimated friction curves for different values of T.
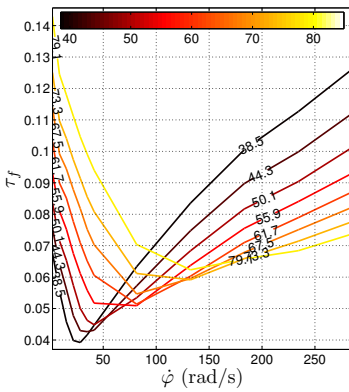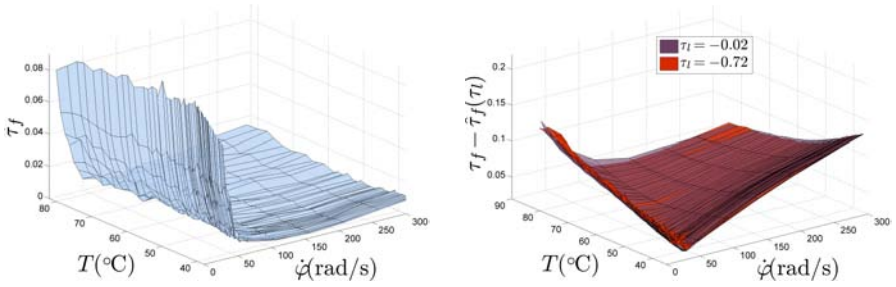

(b) Friction surface cuts for different values of T at $\tau_l = -0.02$.

(c) Friction surface cuts for different values of $\dot{\varphi}$ rad/s at $\tau_l = -0.02$.

Figure 9: *The temperature dependence of the static friction curve.*

(a) Difference $\tilde{\tau}_f$ between the two static friction surfaces in Figure 9a.

(b) Static friction surfaces in Figure 9a after subtraction of the τ_l-dependent terms.

Figure 10: Indication of independence between effects caused by T and τ_l.

With the values given from Table 2, the effects of the manipulation torque were subtracted from the friction curves of the two surfaces in Figure 9a, that is, the quantities $\tau_f - \hat{\tau}_f(\tau_l)$ were computed. The resulting surfaces are shown in Figure 10b. As expected, the surfaces become quite similar. The result can also be interpreted as an evidence that the model structure used for the τ_l-dependent terms and the identified values for the parameters are correct. Obviously, the original model structure $\mathcal{M}_0$ can not characterize all observed phenomena, even after discounting the τ_l-dependent terms.

3.5 A proposal for $\mathcal{M}^*$

From the characteristics of the T-related effects and the already discussed τ_l-effects, $\mathcal{M}_1$ is extended to:

$$\tau_f(\dot{\varphi}, \tau_l, T) = \{F_{c,0} + F_{c,\tau_l}|\tau_l|\} + F_{s,\tau_l}|\tau_l|e^{-\left|\frac{\dot{\varphi}_m}{\dot{\varphi}_{s,\tau_l}}\right|^{\alpha^*}} + \qquad (\mathcal{M}^*_{g,\tau_l})$$

$$+ \{F_{s,0} + F_{s,T}T\}e^{-\left|\frac{\dot{\varphi}_m}{\{\dot{\varphi}_{s,0}+\dot{\varphi}_{s,T}T\}}\right|^{\alpha^*}} + \qquad (\mathcal{M}^*_{g,T})$$

$$+ \{F_{v,0} + F_{v,T}e^{\frac{-T}{T_{Vo}}}\}\dot{\varphi}. \qquad (\mathcal{M}^*_{h,T})$$

The model describes the effects of τ_l and T for the investigated robot joint. The first $\mathcal{M}^*_g$ expressions relate to the velocity-weakening friction while $\mathcal{M}^*_h$ relates to the velocity-strengthening regime. τ_l only affects the velocity-weakening regime and requires a total of 3 parameters, $[F_{c,\tau_l}, F_{s,\tau_l}, \dot{\varphi}_{s,\tau_l}]$. T affects both regimes and requires 4 parameters, $[F_{s,T}, \dot{\varphi}_{s,T}, F_{v,T}, T_{Vo}]$. The 4 remaining parameters, $[F_{c,0}, F_{s,0}, \dot{\varphi}_{s,0}, F_{v,0}]$, relate to the original friction model structure $\mathcal{M}_0$. Notice that under the assumption that τ_l- and T effects are independent, their respective expressions appear as separated sums in $\mathcal{M}^*$.

The term $F_{v,T}e^{-T/T_{Vo}}$ in $\mathcal{M}^*_{h,T}$ is motivated by the exponential-like behavior of viscous friction (recall Figure 9c). In fact, the parameter T_{Vo} is a reference to

Table 3: *Identified T-dependent and $\mathcal{M}_0$-related model parameters.*

$F_{c,0}\left[10^{-2}\right]$	$F_{s,0}\left[10^{-2}\right]$	$F_{s,T}\left[10^{-3}\right]$	$F_{v,0}\left[10^{-4}\right]$
3.11 ± 0.028	-2.50 ± 0.12	1.60 ± 0.022	1.30 ± 0.056

$F_{v,T}\left[10^{-3}\right]$	$\dot{\varphi}_{s,0}$	$\dot{\varphi}_{s,T}$	T_{Vo}
1.32 ± 0.076	-24.81 ± 0.87	0.98 ± 0.018	20.71 ± 0.91

the Vogel-Fulcher-Tamman exponential description of viscosity and temperature, Seeton (2006). This description is valid for the temperature range considered here, more complex expressions may be needed for larger temperature variations. See Seeton (2006) for other possible structures.

Given the already identified τ_l-dependent parameters in Table 2, the remaining parameters of $\mathcal{M}^*$ are identified from the measurement results presented in Figure 10b, after the subtraction of the τ_l-terms. The values are shown in Table 3.

3.6   Validation

A separate data set is used for the validation of the proposed model structure $\mathcal{M}^*$. It consists of several static friction curves measured at different τ_l- and T values, as seen in Figure 11.

With an instance of $\mathcal{M}^*$ given by the parameter values from Tables 2 and 3, the distribution of the prediction errors, $p(\varepsilon)$, achieved with the validation data set is shown in Figure 12. For a comparison, the distribution of errors related to a single instance of $\mathcal{M}_0$, with parameters given in Table 1, is also shown in the figure. As it can be seen, $\mathcal{M}^*$ is able to capture considerably more of the friction behavior than $\mathcal{M}_0$, with only speed dependence. The mean, standard deviation and largest absolute error for $\mathcal{M}^*$ are $[-9.24\,10^{-4},\ 4.23\,10^{-3},\ 1.88\,10^{-2}]$, compared to $[1.09\,10^{-2},\ 1.34\,10^{-2},\ 7.58\,10^{-2}]$ for $\mathcal{M}_0$.

The proposed model structure has also been successfully validated in other joints with similar gearboxes, but it might be interesting to validate it in other robot types and even other types of rotating mechanisms.

4   Conclusions and Further Research

The main contribution of this paper is the empirically motivated modeling of static friction as a function of the variables $\mathcal{X}^* = [\dot{\varphi},\ \varphi_a,\ \tau_p,\ \tau_l,\ T]$. While no significant influences of joint angle and perpendicular torque could be found by the experiments, the effects of manipulation load torque τ_l, and temperature T, were significant and included in the proposed model structure $\mathcal{M}^*$. As shown in Figure 12, the model is needed in applications where the manipulation load torque and the temperature play significant roles.

In the studies, the friction phenomena was fairly direction independent. If this
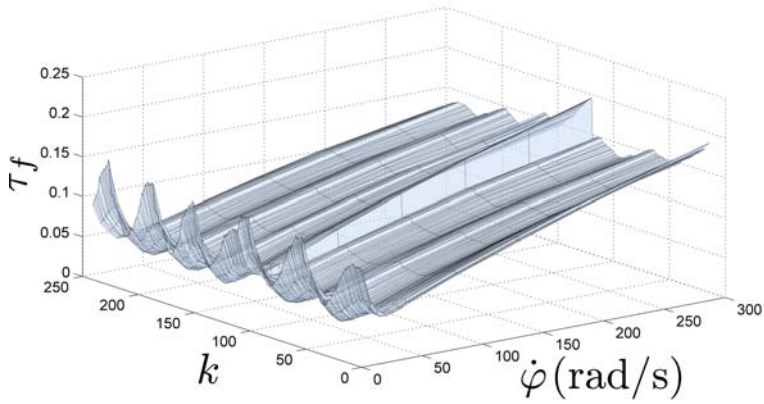
(a) Static friction curves.

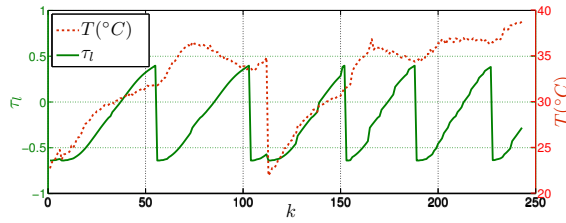*(b) τ_l - and T conditions.*

Figure 11: *Validation data set. Notice the large variations of T - and τ_l values in Figure (b) when registering the static friction curves in (a).*
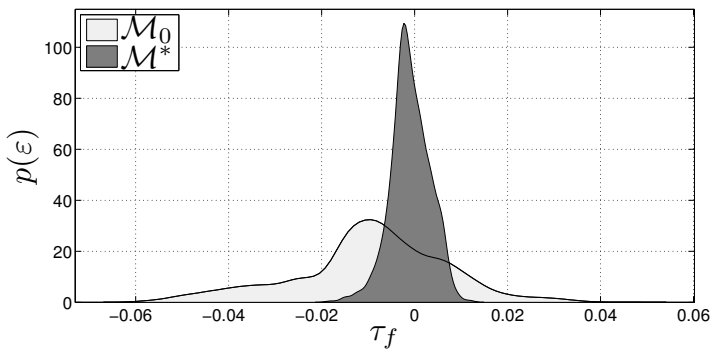

Figure 12: *Distribution of the prediction errors for the models $\mathcal{M}_0$ and $\mathcal{M}^*$ achieved using the validation data set. Notice the considerable better performance of $\mathcal{M}^*$.*

was not the case, two instances of $\mathcal{M}^*$ could be used to describe the whole speed range, but requiring two times more parameters. The model $\mathcal{M}^*$ has a total of 7 terms and 4 parameters which enter the model in a nonlinear fashion. The identification of such a model is computationally costly and requires data from several different operating conditions. Studies on defining sound identification excitation routines are therefore important.

Only static friction (measured when transients caused by velocity changes have disappeared) was considered in the studies. It would be interesting to investigate if a dynamic model, for instance given by the LuGre model structure $\mathcal{M}_L$, could be used to describe dynamic friction with extensions from the proposed $\mathcal{M}^*$. However, to make experiments on a robot joint in order to obtain a dynamic friction model is a big challenge. Probably, such experiments must be made on a robot joint mounted in a test bench instead of on a robot arm system, which has very complex dynamics.

A practical limitation of $\mathcal{M}^*$ is the requirement on availability of τ_l and T. Up to date, torque- and joint temperature sensors are not available in standard industrial robots. As mentioned in Section 3.1, the joint torque components can be estimated from the torque reference to the drive system by means of an accurate robot model. In this situation, it is important to have correct load parameters in the model to calculate the components of the load torques.

Regardless these experimental challenges, there is a great potential for the use of $\mathcal{M}^*$ for simulation-, design- and evaluation purposes. The designer of control algorithms, the diagnosis engineer, the gearbox manufacturer, etc. would benefit by using a more realistic friction model.

Bibliography

F. Al-Bender and J. Swevers. Characterization of friction force dynamics. *IEEE Control Systems Magazine*, 28(6):64–81, 2008.

B. Armstrong-Hélouvry. *Control of Machines with Friction*. Kluwer Academic Publishers, 1991.

K. J. Åström and C. Canudas-de Wit. Revisiting the lugre friction model. *Control Systems Magazine, IEEE*, 28(6):101–114, Dec. 2008.

A. C. Bittencourt and S. Gunnarsson. Static friction in a robot joint - modeling and identification of load and temperature effects. *ASME Journal of Dynamic Systems, Measurement, and Control*. Accepted for publication.

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

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

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

L. C. Bo and D. Pavelescu. The friction-speed relation and its influence on the critical velocity of stick-slip motion. *Wear*, 82(3):277 – 289, 1982.

B. Bona and M. Indri. Friction compensation in robotics: an overview. In *Proc. of the 44th IEEE International Conference on Decision and Control*, Dec 2005.

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

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.

M. Dohring, E. Lee, and W. Newman. A load-dependent transmission friction model: theory and experiments. In *Proc. of the 1993 IEEE International Conference on Robotics and Automation*, pages 430 –436 vol.3, may 1993.

D. Dowson. *History of Tribology*. Professional Engineering Publishing, London, UK., 1998.

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

A. Gogoussis and M. Donath. Coulomb friction effects on the dynamics of bearings and transmissions in precision robot mechanisms. In *Proc. of the 1998 IEEE International Conference on Robotics and Automation*, pages 1440 –1446 vol.3, apr 1988.

G. H. Golub and V. Pereyra. The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate. *SIAM Journal on Numerical Analysis*, 10(2):pp. 413–432, 1973.

Y. Guo, Z. Qu, Y. Braiman, Z. Zhang, and J. Barhen. Nanotribology and nanoscale friction. *Control Systems Magazine, IEEE*, 28(6):92 –100, dec. 2008.

P. Hamon, M. Gautier, and P. Garrec. Dynamic identification of robots with a dry friction model depending on load and velocity. In *Proc. of the 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 6187 –6193, oct. 2010.

A. Harnoy, B. Friedland, and S. Cohn. Modeling and measuring friction effects. *Control Systems Magazine, IEEE*, 28(6), Dec. 2008.

B. Jacobson. The stribeck memorial lecture. *Tribology International*, 36(11):781 – 789, 2003.

H. M. Kim, S. H. Park, and S. I. Han. Precise friction control for the nonlinear friction system using the friction state observer and sliding mode control with recurrent fuzzy neural networks. *Mechatronics*, 19(6):805 – 815, 2009.

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

R. Mattone and A. D. Luca. Relaxed fault detection and isolation: An application to a nonlinear case study. *Automatica*, 42(1):109 – 116, 2009.

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

M. Namvar and F. Aghili. Failure detection and isolation in robotic manipulators using joint torque sensors. *Robotica*, 2009.

H. Olsson, K. J. Åström, C. C. de Wit, M. Gafvert, and P. Lischinsky. Friction models and friction compensation. *European Journal of Control*, 4(3):176–195, 1998.

C. J. Seeton. Viscosity-temperature correlation for liquids. *Tribology Letters*, 22 (1):67–78, Mar. 2006.

W. Susanto, R. Babuska, F. Liefhebber, and T. van der Weiden. Adaptive friction compensation: application to a robotic manipulator. In *Proc. of 17th IFAC World Congress*, Dec 2008.

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

R. Waiboer. *Dynamic Modelling, Identification and Simulation of Industrial Robots*. PhD thesis, University of Twente, 2007.

R. Waiboer, R. Aarts, and B. Jonker. Velocity dependence of joint friction in robotic manipulators with gear transmissions. In *Proc. of the 2005 ECCOMAS Thematic Conference Multibody Dynamics*, pages 1–19, Madrid, Spain, 2005.

M. Woydt and R. Wäsche. The history of the stribeck curve and ball bearing steels: The role of adolf martens. *Wear*, 268(11-12):1542 – 1546, 2010.

Paper B

Modeling and Identification of Wear in a Robot Joint under Temperature Uncertainties

Authors: André Carvalho Bittencourt, Patrik Axelsson, Ylva Jung and Torgny Brogårdh

Modeling and Identification of Wear in a Robot Joint under Temperature Uncertainties

André Carvalho Bittencourt, Patrik Axelsson, Ylva Jung * and Torgny Brogårdh **

*Dept. of Electrical Engineering,
Linköping University,
SE–581 83 Linköping, Sweden
`{andrecb, axelsson,`
`ylvju}@isy.liu.se`

**ABB Robotics
Västerås, Sweden
`torgny.brogardh@se.abb.com`

Abstract

This paper considers the problem of wear estimation in a standard industrial robot joint. The effects of wear to the static friction of a robot joint are analyzed from experiments. An extended static friction model is proposed that explains changes related to joint speed, load, temperature and wear. Based on this model and static friction observations, a model-based wear estimator is proposed. The performance of the estimator under temperature uncertainties is found both by means of simulations and experiments in an industrial robot. Special attention is given to the analyses of the best speed region for wear estimation. As it is shown, the method can distinguish the effects of wear even under large temperature variations, opening up for the use of robust joint diagnosis for industrial robots.

1 Introduction

In the manufacturing industry, preventive scheduled maintenance is a common approach used to guarantee the reliability of a robot system, avoiding unpredicted stops. Such scheduling is in general based on the estimated robot's components lifespan and not on its actual conditions. With the development of model-based diagnosis methods, more sophisticated approaches have been proposed for manipulator diagnosis (Caccavale and Villani (2003)). These methods are based on the principle of analytical redundancy, where residuals (deviations between the system outputs and model-based predictions) are monitored to perform fault detection, see Isermann (2005) for an overview on model-based diagnosis methods. A typical approach for residual generation in robotics is the use of nonlinear observers, as presented in McIntyre et al. (2005). 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 ob-

servers 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, see for example Freyermuth (1991).

The diagnosis of actuators is a relevant research topic for industrial robots. In the literature, actuator failures are typically considered as abrupt changes in the output torque signals. These fault 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, even if detected, might still cause damages. An important type of failure is the one related to wear in robot joints. This type of fault develops with time/usage and might be detected at an early stage, allowing for *condition-based maintenance*.

The wear processes inside a robot joint cause an eventual increase of wear debris in the lubricant. Monitoring the iron content of lubricant samples taken from the robot joint can thus be used as an indication of the joint condition. The study of wear debris particles is known as ferrography and was first introduced by Seifert and Westcott (1972). Since then, the science has evolved and helped to understand wear related phenomena (Roylance (2005)). These techniques are however intrusive and costly, requiring laboratory analyses.

It is well known that friction changes can follow as a result of wear processes in mechanical systems, see for example Kato (2000). In this paper, the effects of wear to static friction in a robot joint are analyzed and modeled based on empirical observations. As the study shows, the effects are sensible and a possible diagnostic solution is thus to *monitor the joint friction*.

The friction in a robot joint is the result of complex interactions between contacting surfaces and can be affected by other factors than wear, such as,

- temperature,
- force/torque levels,
- joint angle,
- velocity,
- acceleration,
- lubricant/grease properties.

When designing a diagnostic method, it is necessary to understand the effects of the different variables to be able to distinguish them from those related to wear. The static friction model presented in Bittencourt et al. (2010) is extended in this paper to achieve a model that can represent the effects of speed, temperature, load and wear in a robot joint. With the proposed model, it is possible to predict the behavior of friction over wide operating conditions, opening up for robust diagnosis.

In this paper, a wear estimator is defined given static friction observations, τ_f, and predictions from the developed model, $\hat{\tau}_f(\cdot)$. The estimator is defined in a prediction error sense, that is

$$\hat{w} = \arg \min_{w} V\Big(\tau_f - \hat{\tau}_f\left(\dot{\varphi}, \tau_l, T, w\right)\Big), \tag{1}$$

where $V(\cdot)$ is a cost function, w is the wear level, φ is the vector of joint an-
gles, $\dot{\varphi} = \frac{\mathrm{d}}{\mathrm{d}t}\varphi$, τ_l is the load component that is manipulated by the joint and T
is the joint temperature, see also Figure 1b. Since joint temperature is seldom
measured in industrial applications, its effects are considered assuming T as a
random variable with an unknown distribution function but known lower and
upper limits. The estimator *characteristics under temperature uncertainties* are
analyzed in detail.

Static friction observations are made possible through a dedicated test cycle, which
outputs τ_f at a set of pre-defined speed levels. When designing test cycles, it is
important to distinguish between cycles used for the development of diagnosis
methods and cycles used for diagnosis of a robot in an industrial installation. In
the later case, the cycle must take only a short time to perform. This means that
the cycle must be customly designed for the estimation of the desired quantity.
In the following, it is shown that it is possible to obtain an efficient test cycle if
correct speed levels are used, even under large temperature variations.

The wear estimator and its properties are studied based on observed static fric-
tion in joint 2 of an ABB IRB 6620 industrial robot, see Figure 1a. Joint 2 is chosen
for the study as it endures great load variations for the type of robot considered.
The paper is organized as follows. Section 2 presents a method to observe static
friction in a robot joint that can be used to define a test cycle. Section 3 discusses
the static friction model presented in Bittencourt et al. (2010). The **main contri-
butions** are contained in Sections 4 to 6, where

- the empirical wear model is developed,
- the model-based wear estimator is defined,
- the best speed region for wear estimation under temperature uncertainties
  is discussed and
- the approach is validated through simulations and a case study based on
  real data.

The conclusions and proposals for further research are presented in Section 7.

2   Static Friction Observations through Experiments

A manipulator is a multivariable, nonlinear system that can be described in a
general manner through the rigid multi-body dynamic model

$$M(\varphi)\ddot{\varphi} + C(\varphi, \dot{\varphi}) + \tau_g(\varphi) + \tau_f = u \qquad (2)$$

where $M(\varphi)$ is the inertia matrix, $C(\varphi, \dot{\varphi})$ relates to speed dependent terms (Cori-
olis and centrifugal), $\tau_g(\varphi)$ are the gravity-induced torques at the joints and τ_f
contains the joint friction components. The system is controlled by the input
torque u, applied by the joint motor (in the experiments the torque reference
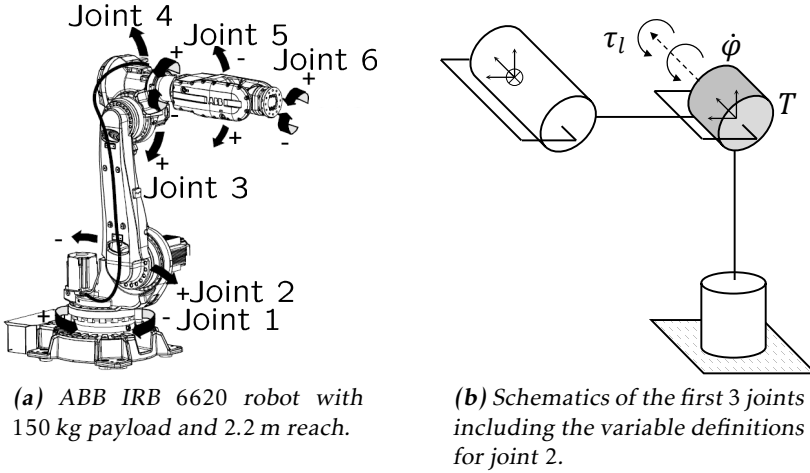
(a) *ABB IRB 6620 robot with 150 kg payload and 2.2 m reach.*

(b) *Schematics of the first 3 joints including the variable definitions for joint 2.*

Figure 1: *The study is based on joint 2 of the ABB robot IRB 6620. $\dot{\varphi}$ is the joint angular speed, T the joint temperature, the manipulated load torque τ_l is the load torque component aligned to the joint degree of freedom.*

from the servo was measured[i]).

When only one joint is moved ($C(\varphi, \dot{\varphi}) = 0$ at that joint) under constant speed ($\ddot{\varphi} \approx 0$), Equation (2) simplifies to

$$\tau_g(\varphi) + \tau_f = u. \tag{3}$$

The resulting applied torque u drives only friction and gravity-induced torques. The required torques to drive a joint in forward, u^+, and reverse, u^-, directions at the constant speed level $\bar{\dot{\varphi}}$ and at a joint angle value $\bar{\varphi}$ (so that $\tau_g(\bar{\varphi})$ is equal in both directions), are

$$\tau_f(\bar{\dot{\varphi}}) + \tau_g(\bar{\varphi}) = u^+ \tag{4a}$$

$$\tau_f(-\bar{\dot{\varphi}}) + \tau_g(\bar{\varphi}) = u^-. \tag{4b}$$

In case an estimate of $\tau_g(\bar{\varphi})$ is available, it is possible to isolate the friction component in each direction using Equation (4). If such estimate is not possible (e.g. not all masses are completely known), τ_f can still be achieved in the case that τ_f is independent of the rotation direction. Subtracting the equations yields

$$\tau_f(\bar{\dot{\varphi}}) - \tau_f(-\bar{\dot{\varphi}}) = u^+ - u^-$$

and if $\tau_f(\bar{\dot{\varphi}}) = -\tau_f(-\bar{\dot{\varphi}})$, the resulting *direction independent friction* evaluated at

---

[i]It is known that using the torque reference from the servo as a measure of the joint torque might not always hold because of the temperature dependence of the torque constant of the motors. The deviations are however considered to be small and are neglected in this paper.
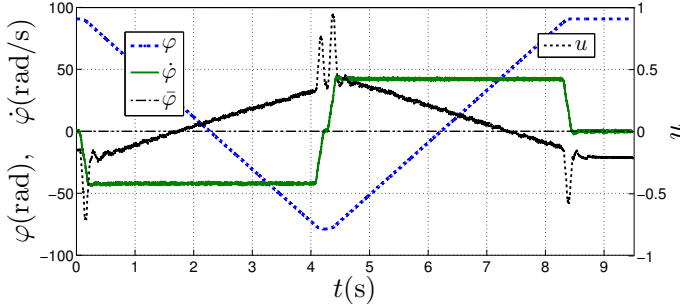
Figure 2: *Excitation signals used for the static friction estimation at $\dot{\varphi} = 42$ rad/s.*

the constant speed $\bar{\dot{\varphi}}$ is:

$$\tau_f(\bar{\dot{\varphi}}) = \frac{u^+ - u^-}{2}. \tag{5}$$

In the experiments, each joint is moved with the desired speed in both directions around a given joint angle $\bar{\varphi}$. Figure 2 shows the measured joint angle-, speed- and torque[i] signals sampled at $2\,\mathrm{kHz}$[ii] for $\bar{\dot{\varphi}} = 42$ rad/s around $\bar{\varphi} = 0$. The constant speed data is segmented around $\bar{\varphi}$ and the static friction levels can be achieved using Equation (4) or (5). With the experimental setup used, the time needed for measurements and computations of the friction level for one joint at one speed was in average 14s.

The friction values achieved over the whole joint speed range can be presented in a static friction curve, sometimes referred to as a Stribeck curve, see Figure 3. As seen in the figure, there is only a small direction dependecy of friction for the investigated joint. Therefore, in this paper, friction levels are achieved using Equation (5), which is not influenced by deviations in the gravity model of the robot.

Remark 1. Throughout the paper, friction values obtained using the method presented above are named *friction observations*.

A test cycle can be achieved by simply taking friction observations at a set of different speed levels. The choice of speed set is a design criteria, representing a compromise between the test-cycle time and amount of friction observations.

---

[i]Throughout the paper all torques are normalized to the maximum manipulation torque at low speed and are therefore displayed as dimensionless quantities.

[ii]Similar results have been experienced with sampling rates down to 220 Hz.
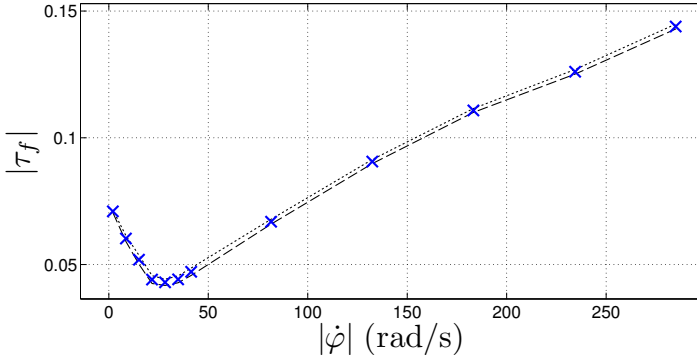
Figure 3: *Static friction curve. Crosses indicate friction levels achieved using Equation* (5), *with the assumption that friction is direction independent. Dotted/dashed lines indicate friction levels achieved using Equation* (4a) *and Equation* (4b) *respectively.*

Table 1: *Identified parameters for the model* (6).

$F_{c,0}$	F_{c,τ_l}	$F_{s,0}$	F_{s,τ_l}	$F_{s,T}$	$F_{v,0}$
$3.04_{\,10^{-2}}$	$2.32_{\,10^{-2}}$	$-2.44_{\,10^{-2}}$	$1.28_{\,10^{-1}}$	$1.69_{\,10^{-3}}$	$1.29_{\,10^{-4}}$
$F_{v,T}$	$\dot{\varphi}_{s,0}$	$\dot{\varphi}_{s,\tau_l}$	$\dot{\varphi}_{s,T}$	T_{Vo}	α
$1.31_{\,10^{-3}}$	-25.00	9.07	1.00	21.00	1.3

3   Static Friction Model

In Bittencourt et al. (2010), a static friction model that can explain the effects of temperature and load torque levels is presented. As shown, this model can be used to predict the normal behavior of static friction under broad operation conditions. The model is developed from a standard description of the static friction and empirical observations, taking the form:

$$\tau_f(\dot{\varphi}, \tau_l, T) = \{F_{c,0} + F_{c,\tau_l}|\tau_l|\} + F_{s,\tau_l}|\tau_l|e^{-\left|\frac{\dot{\varphi}}{\dot{\varphi}_{s,\tau_l}}\right|^{\alpha}} + \tag{6a}$$

$$+ \{F_{s,0} + F_{s,T}\,T\}e^{-\left|\frac{\dot{\varphi}}{\{\dot{\varphi}_{s,0}+\dot{\varphi}_{s,T}\,T\}}\right|^{\alpha}} + \tag{6b}$$

$$+ \{F_{v,0} + F_{v,T}\,e^{\frac{-T}{T_{\text{Vo}}}}\}\dot{\varphi}, \tag{6c}$$

where τ_l is the manipulated load torque and T is the joint temperature, see Figure 1b. The remaining variables are parameters used to model the friction behavior. The terms in (6a) describe the effects of τ_l, which are more significant at low speeds; the terms in (6b) and (6c) describe the effects of T at low, respectively high, speed ranges (see Bittencourt et al. (2010) for a more detailed discussion). The values for the identified parameters used in this paper are shown in Table 1.
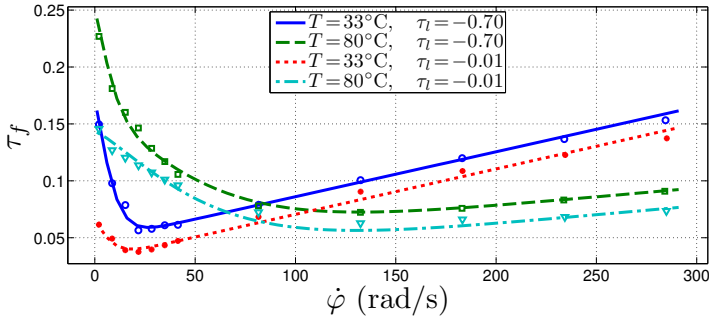
Figure 4: *Observed static friction curves (circles) and model-based predictions (lines) for low and high values of T and τ_l.*

To illustrate the model behavior, Figure 4 presents observed and model-based predictions of friction curves for high and low values of τ_l and T. Notice the effects of τ_l, which give an offset increase of the whole curve together with an exponential-like increase at speeds below 25 rad/s. The effects of T can be seen as an exponential increase at speeds below 80 rad/s and a decrease of the curve slope at higher speeds. Notice further that for such temperature changes there is a speed range where the effects are less pronounced, in this case around 80 rad/s.

4   Wear – Analyses and Modeling

It is difficult to fully comprehend the effects of wear in a robot joint. Monitoring the system until a failure takes place is a costly and time consuming task. With the objective of understanding these effects, accelerated wear tests were performed with a robot joint while friction curves were observed periodically. During an accelerated test, the robot joint under investigation is continuously run for several months or years.

The results of such an experiment is shown in Figure 5, with observed friction curves obtained at the *same load- and temperature levels*. As can be noticed, the effects of wear appear first in the low speed region, in this case, up to about 150 rad/s. In Figure 5, the dashed line indicates the samples associated with a wear level that gear experts find relevant for issuing an alarm. Up to this point, the changes appear mostly in the low speed range. If the accelerated wear tests proceed, the friction curve is also affected at higher speed levels. A direct comparison with Figure 4 reveals that the friction changes caused by increased wear are in the same magnitude as the changes caused by load/temperature but with different speed dependence. Therefore, it might be possible to obtain a selective identification of wear in a robot joint.
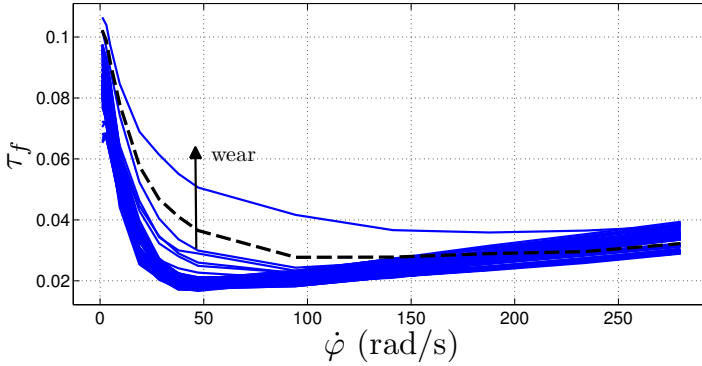
Figure 5: Effects of wear in friction curves under constant load- and temperature conditions observed at regular intervals during accelerated wear tests. The dashed line relates to the wear level at which an alarm should be generated.

4.1 Wear Modeling

Resolving for coupled effects between wear, temperature, load and other parameters would require costly long term experiments. In order to make it possible to examine and model the effects of wear, a simplifying assumption is taken that considers the effects of load/temperature independent of those caused by wear.

Under this assumption, the effects of wear in the static friction curves of Figure 5 can be isolated since temperature/load conditions are the same for these data. A *wear profile* $\tilde{\tau}_f$ is defined by subtracting a friction curve observed before the accelerated wear tests started from the remaining curves, under wear effects. The resulting wear profile from the accelerated wear test in Figure 5 can be seen in Figure 6, where the curves are presented along a time index k, indicating the length of the accelerated tests. In the figure, the dashed line at time-index $k = 80$ relates to the alarm level, as in Figure 5.

As can be noticed, the effects of wear appear as an *increase of the exponential-like* behavior of the friction curves up to 150 rad/s and a *small decrease of the velocity slope* dependecy at higher speeds. Introducing w as a wear parameter, the observations support the choice of a model structure for the wear profile as

$$\tilde{\tau}_f(\mathsf{w}) = F_{s,\mathsf{w}}\mathsf{w}e^{-\left|\frac{\dot{\varphi}}{\varphi_{s,\mathsf{w}}\mathsf{w}}\right|^{1.3}} + F_{v,\mathsf{w}}\,\mathsf{w}\dot{\varphi}. \tag{7}$$

The model represents wear effects with an exponential- and a velocity dependent terms, with 3 parameters. The model parameters cannot be directly identified since the wear quantity w is not measurable. To overcome this, w is defined with values between [0, 100], relative to a failure state, the value w = 35 is chosen as a reference for the wear effects associated with an alarm, at $k = 80$ in Figure 6. With this convention, the parameters for (7) are identified using the wear profile data
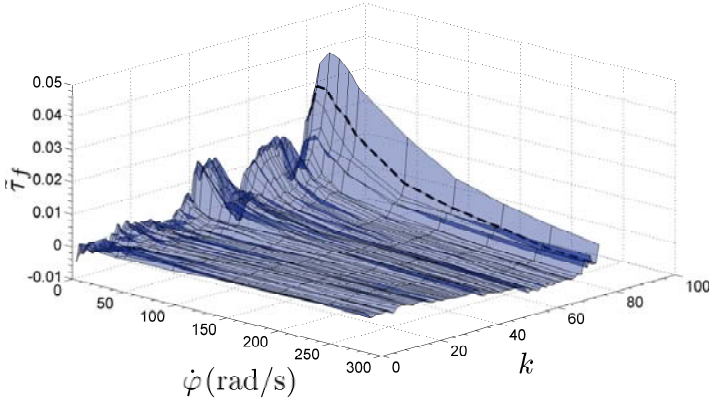
Figure 6: *Friction wear profile $\tilde{\tau}_f$ computed from the data in Figure 5. The dashed line indicates the experiment where an alarm should be given.*
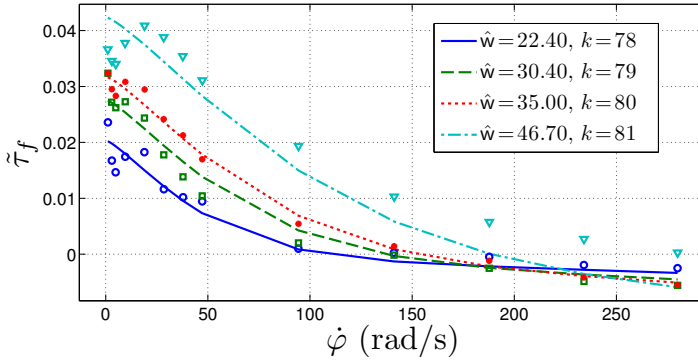
Figure 7: *Measured wear profile (circles) and model-based predictions (lines).*

$\tilde{\tau}_f$ for the curve at $k = 80$. The values achieved are

$$F_{s,w} = 9.10 \, 10^{-3}, \quad F_{v,w} = -5.3 \, 10^{-7}, \quad \dot{\varphi}_{s,w} = 2.20. \tag{8}$$

With the chosen parameters, an estimate $\hat{w}$ is obtained at each k using the data for the wear profile in Figure 6. With the identified wear values, the wear profile given by model predictions from (7) and observations are presented for the interval $k = [78, 81]$ in Figure 7. As can be noticed, the model can fairly well predict the behavior of $\tilde{\tau}_f$.

Under the assumption that the effects of load/temperature are independent of those caused by wear, it is possible to extend the model given in (6) to include the effects of wear as

$$\tau_f(\dot{\varphi}, \tau_l, T, w) = \tau_f(\dot{\varphi}, \tau_l, T) + \tilde{\tau}_f(w), \tag{9}$$

Figure 8: *Increase of wear levels given by the model* (9). *The dashed line relates to the wear level at which an alarm should be generated.*

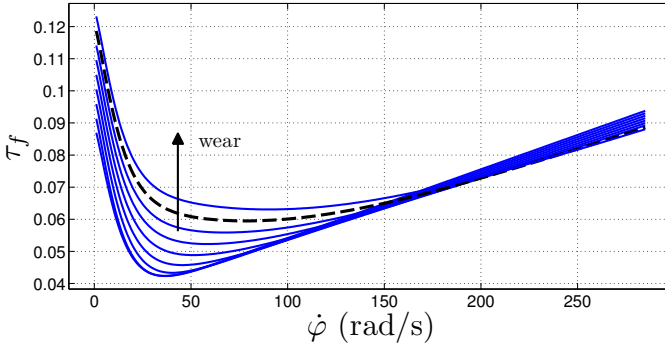where $\tau_f(\dot{\varphi}, \tau_l, T)$ is given by (6) and $\tilde{\tau}_f(w)$ is described in (7). With parameters given by Table 1 and Equation (8), Figure 8 presents the friction predictions given by the proposed model at $T = 40°C$ and $\tau_l = 0.10$ for wear values in the range w $= [0, 40]$. Notice that the effects are concentrated to the speed range of $[0, 150]$ rad/s. As previously, the dashed line in Figure 8 indicates an alarm level for the wear, it has a friction increase of 0.017 at 47 rad/s and w $= 35$.

5   A Model-based Wear Estimator

Considering that $\dot{\varphi}$, τ_l and T can be measured/estimated, it is possible to define a wear estimator in a prediction error sense as in Equation (1)

$$\hat{w} = \arg \min_{w} V\Big(\tau_f - \hat{\tau}_f(\dot{\varphi}, \tau_l, T, w)\Big), \qquad (10)$$

where τ_f is observed through the procedure presented in Section 2 and the predictions $\hat{\tau}_f$ are given by the model in (9). Since the model is not invertible with respect to w, the problem should be solved by nonlinear identification methods.

In industrial applications, joint temperature measurements are seldom available. In order to solve this problem it is proposed to consider joint temperature to be a random variable, with a certain probability distribution function $p(T)$. The distribution function $p(T)$ is assumed to be unknown, but with known lower and upper limits $\underline{T}$, $\overline{T}$. For a robot operating in a controlled indoor environment, $\underline{T}$ would be minimum room temperature while $\overline{T}$ is given by the maximum room temperature and self heating of the joint due to actuator losses.

To include the effects of temperature in the estimator, Equation (1) is solved for N realizations of the considered random T. The estimator assumes that $p(T)$ is constant in the given temperature range and samples are drawn with equal probabilities over $[\underline{T}, \overline{T}]$. The expected value of the resulting N estimates is then

taken as the wear estimate,

$$\hat{w}_i = \arg\min_{w} V\left(\tau_f - \hat{\tau}_f(\dot{\varphi}, \tau_l, T_i, w)\right) \tag{11a}$$

$$\hat{w} = \mathbb{E}\left[\{\hat{w}_N\}\right], \quad T_i \sim \mathcal{U}(\underline{T}, \overline{T}), \quad i = 1, \ldots, N \tag{11b}$$

here $\hat{\tau}_f(\dot{\varphi}, \tau_l, T_i, w)$ is given by (9) and $\{\hat{w}_N\}$ denotes the sequence $\{\hat{w}_1, \ldots, \hat{w}_N\}$.

To solve the minimization problem in (11a), grid search is used with a square error cost function $V(\cdot)$. To evaluate (11b), a Monte Carlo simulation is carried out. For a given observation of τ_f, it takes N samples T_i from $\mathcal{U}(\underline{T}, \overline{T})$, yielding N estimates $\{\hat{w}_N\}$. The expected value of the sequence $\{\hat{w}_N\}$ is taken as the wear estimate $\hat{w}$.

5.1   Statistical Properties

To evaluate the behavior of the proposed estimator, the static friction observations are assumed to follow the model

$$\tau_f = \tau_f(\dot{\varphi}, \tau_l, T, w^0) + e, \tag{12a}$$

$$e \sim \mathcal{N}(0, \sigma_e^2), \quad T \sim p(T) = \mathcal{N}(\mu_T, \sigma_T^2) \tag{12b}$$

where $\tau_f(\cdot)$ is generated according to the model (9) with parameters given in Table 1 and Equation (8), and where e is additive measurement noise. The stochastic properties of e are motivated by experimental studies of the estimation method presented in Section 2. For the considered joint, the estimated standard deviation is $\sigma_e = 0.0015$. The wear level is defined as $w^0 = 35$, which means 35% wear in relation to a failure state, and is related to the alarm level for the wear as discussed in Section 4. Temperature is considered to follow a normal distribution with mean and standard deviation $[\mu_T, \sigma_T] = [40, 3]°C$. The chosen distribution illustrates the scenario where friction observations are always taken after the same sequence of events during the day, for instance between two production shifts in a room with controlled environment temperature. The large standard deviation is used to cope with, amongst others, variations of room temperature and variations of self-heating caused by the loses for the different operations of the robot. The observations of friction torques τ_f obtained from the model in (12) are used as input to the estimator defined in (11), with $N = 200$, $[\underline{T}, \overline{T}] = [30, 50]°C$ and parameters to $\hat{\tau}_f(\dot{\varphi}, \tau_l, T_i, w)$ given in Table 1 and Equation (8).

To evaluate the performance of the method, 10.000 simulations were made at 56 different speeds in the range $\dot{\varphi} = [0, 280]$ rad/s. The estimation is performed separately for each speed point to assess the method performance along the achievable speed range. Figure 9a presents the estimated probability distribution function of the estimates $\hat{w}$ with respect to speed in a contour plot, together with their expected values. As can be noticed, the variability of the estimates is smaller in the speed range $[30, 55]$ rad/s. Figure 9b presents the estimated bias over speed. Notice the small bias at speeds up to 150 rad/s and the monotonic increase at higher speeds. The results are directly related to the behavior of the modeled wear and temperature effects for the $p(T)$ considered (recall Figures 4 and 8).
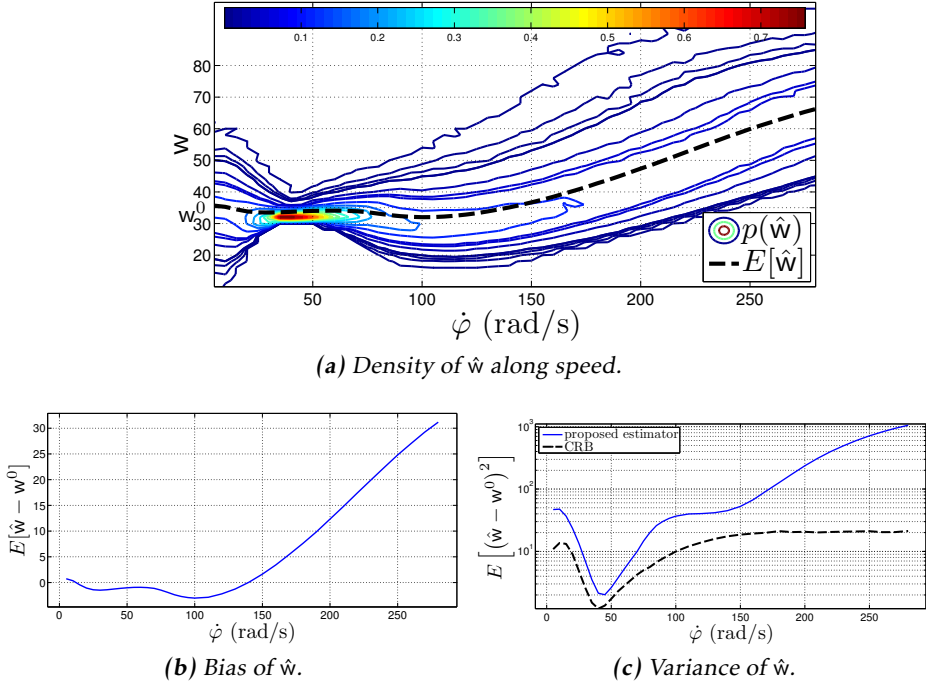
(a) Density of ŵ along speed.

(b) Bias of ŵ.

(c) Variance of ŵ.

Figure 9: *Contour plot of the estimated probability distribution function of ŵ as a function of speed. Notice the sharper densities in the interval [30, 55] rad/s. The variance in (c) is shown together with its CRB in a semi-log scale.*

Variance.   From the simulation studies, it is possible to estimate the resulting variance of the estimator. A lower bound for the achievable variance allows for an assessment of the method performance. For the observation model defined in (12), introduce

$$\bar{p}(\tau_f|\mathsf{w}) \triangleq \iint p(\tau_f|\mathsf{w}, e, T)\, p(e)p(T)\, \mathrm{d}e\, \mathrm{d}T, \tag{13}$$

where $p(\,\cdot\,)$ denotes a probability distribution function. Equation (13) is the marginalization of the effects of measurement noise and random temperature. Using $\bar{p}(\tau_f|\mathsf{w})$, the Cramer-Rao bound (CRB) for *any unbiased* estimator can be defined as (Van Trees (2001))

$$\mathbb{E}\left[\left(\hat{\mathsf{w}} - \mathsf{w}^0\right)^2\right] \geq M^{-1} \tag{14a}$$

$$M = \mathbb{E}\left[\left(\frac{\partial}{\partial \mathsf{w}} \log \bar{p}(\tau_f|\mathsf{w})\right)^2\right] \tag{14b}$$

Since there is no analytical solution for (13), Monte Carlo Integration (MCI) is

used to compute it numerically. The derivative in (14b) is approximated numerically with the central difference,

$$\frac{\log \bar{p}(\tau_f|w^0 + h) - \log \bar{p}(\tau_f|w^0 - h)}{2h}$$

for $h = 0.1$.

The CRB is shown together with the variance for the proposed estimator in Figure 9c. As can be noticed, the variance is high at low and high speeds. The estimator can however approximate the CRB at the speed region [30, 55] rad/s. This illustrates the relevance of a correct choice of speed points to observe friction for wear identification.

6 Case Study

Gathering enough informative data related to wear from the field is inviable since robots will rarely have wear related problems. The alternative of running accelerated wear tests is also difficult since it might take several months before any wear effects can be seen and it is difficult to obtain reliable statistics without extremely high cost of running several robots. Moreover, temperature studies are also challenging since the thermal time constant of a large robot is several hours. An alternative is proposed that considers two data sets:

Wear profile The first data set is used to relate to the evolution of the wear effects. A wear profile $\tilde{\tau}_f(k)$, as defined in Section 4.1 is used with this purpose. In this study, the data from Figure 6 is considered.

Temperature effects The second data set is used to relate to effects caused by temperature. Several friction curves are observed on another robot of same type over broad temperature conditions and no influence of wear. Joint temperature measurements were registered during the experiments and each friction curve in the set is associated with a joint temperature T. The notation $\tau_f^0(T)$ is then used to denote the friction curve in this data set related to the temperature T.

Neglecting any possible combined effects of temperature and wear, the static friction observation model is defined as

$$\tau_f^*(k) = \tilde{\tau}_f(k) + \tau_f^0(T). \tag{15}$$

Using this framework, it is possible to generate faulty friction observations under different temperature conditions. Notice that these data are not analytically generated, but actually based on static friction observations.

To achieve a desired temperature distribution for the observations $\tau_f^*(k)$, the friction data $\tau_f^0(T)$ are chosen according to the associated T. The reference distribution considered for this case study is the same used in Section 5.1, $\mathcal{N}(\mu_T, \sigma_T)$, with mean and standard deviation $[\mu_T, \sigma_T] = [40, 3]°C$.
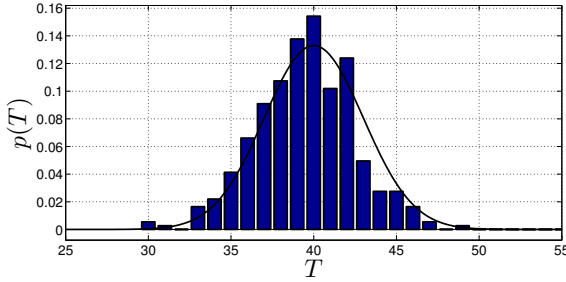
Figure 10: *Reference (line) and achieved (histogram) temperature distributions for the data used in the case study.*

The achieved and reference temperature distributions for the data used in this case study can be seen in Figure 10. With the generated data, wear is estimated separately for each speed, using the method given in Equation (11), with $[\underline{T},\ \overline{T}] = [30,\ 50]°C$, $N = 200$ and parameters given in Table 1 and Equation (8). The resulting estimates can be seen in Figure 11 for $\dot{\varphi} = 35.98$ and $\dot{\varphi} = 84.23$.

An evaluation of the estimates against the actual wear value is not possible since the quantity w is not measurable. As an alternative, a wear estimate is taken directly from the fault profile data $\tilde{\tau}_f(k)$. This estimate is found as the value of w minimizing the prediction error between $\tilde{\tau}_f(k)$ and the fault profile model in Equation (7),

$$\hat{w}^*(k) = \arg\min_{w} V(\tilde{\tau}_f(k) - \hat{\tilde{\tau}}_f(w)), \tag{16}$$

with parameters given in Equation (8). The minimization at index k is performed using data from all speeds. Since $\hat{w}^*(k)$ is achieved directly from $\tilde{\tau}_f(k)$ (under no temperature disturbances), using all data available at k, its value is considered as the best possible wear estimate given the available information. The estimate $\hat{w}^*$ is therefore used as a reference for a comparison. The resulting values are shown as the dashed line in Figure 11, notice that the estimates achieved at $\dot{\varphi} = 35.98$ are quite close to $\hat{w}^*$.

Using $\hat{w}^*$ as a reference, the estimates $\hat{w}$ can be evaluated through $\Delta\hat{w} = \hat{w} - \hat{w}^*$. Its mean, $\mu_{\Delta\hat{w}}$, standard deviation $\sigma_{\Delta\hat{w}}$ and worst case deviation, $|\Delta\hat{w}|_\infty$, are computed over the indexes k at each speed. The results, shown in Figure 12, relate qualitatively to the simulation studies presented in Section 5.1, with larger variance at low and high speeds and with increasing bias at high speeds.

The estimation performance was found to be good at speeds close to 35 rad/s. In this speed region, $\hat{w}$ increases according to $\hat{w}^*$ but with slight biases. The best result was found at $\dot{\varphi} = 35.98$ rad/s, with $[\mu_{\Delta\hat{w}}, \sigma_{\Delta\hat{w}}, |\Delta\hat{w}|_\infty] = [-0.18, 4.02, 8.11]$. At this speed region, even a simple threshold, set at 35 could be used to detect the critical wear increase w = 35.

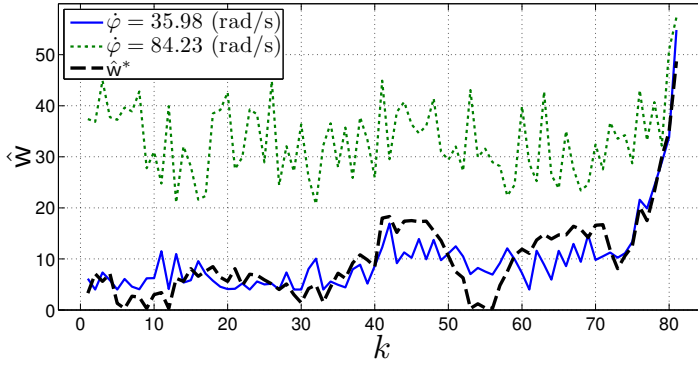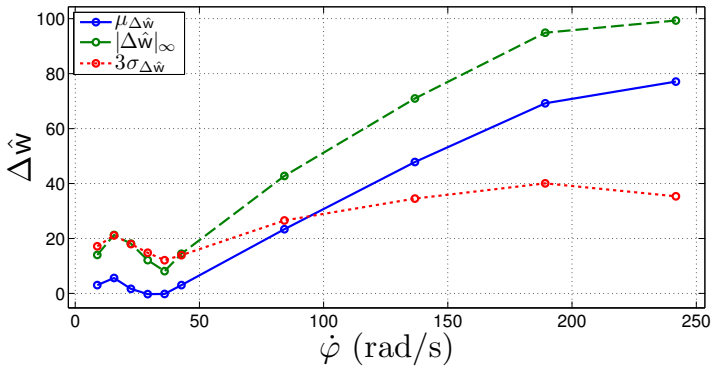**Figure 11:** *Estimated wear levels as function of k at selected speed points and ŵ*.*

Figure 12: *Mean, worst case deviation and 3 standard deviation of* $\Delta\hat{w} = \hat{w} - \hat{w}^*$ *along* $\dot{\varphi}$.

7 Conclusions

In this paper, a model-based wear estimator was proposed that provides a wear estimate given static friction observations at pre-defined speed levels. A detailed simulation study of the best speed range for wear estimation under joint temperature was taken. The simulation results were supported by a case study based on real data. As it was shown, the wear estimates in the optimal speed region could be used to perform fault detection, allowing for robust condition based maintenance of industrial robots. This optimal speed region is however narrow, emphasizing the relevance of a correct choice of speed values.

If the friction observations are made on a robot installed in a manufacturing line, there is the tradeoff between making the test cycle both accurate and short. In case more time can be spent for the diagnosis routine, the accuracy of the estimation could be improved by observing friction at various speeds around the optimal one.

The work will proceed with studies of the method when also variations of the torque levels and the lubricant properties take place. Later studies will moreover be made of the accuracy of the developed method for different types of robot joints. Especially interesting is to find out the friction- and wear behavior of different gear types. Investigations will also be made to see if it is possible to perform reliable wear estimations without using custom designed experiments.

Bibliography

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

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

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

F. Caccavale and L. Villani. Fault diagnosis for industrial robots. In *Fault Diagnosis and Fault Tolerance for Mechatronic Systems:Recent Advances*, volume 1 of *Springer Tracts in Advanced Robotics*, pages 85–108. Springer Berlin / Heidelberg, 2003.

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.

A. De Luca and R. Mattone. An adapt-and-detect actuator fdi scheme for robot manipulators. In *Proc. of the 2004 IEEE International Conference on Robotics and Automation*, volume 5, pages 4975 – 4980 Vol.5, apr. 2004.

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

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

R. Isermann. Model-based fault-detection and diagnosis - status and applications. *Annual Reviews in Control*, 29(1):71 – 85, 2005.

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

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

B. Roylance. Ferrography–then and now. *Tribology International*, 38(10):857 – 862, 2005.

W. Seifert and V. Westcott. A method for the study of wear particles in lubricating oil. *Wear*, 21(1):27 – 42, 1972.

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

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

Paper C

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

Authors: André Carvalho Bittencourt, Kari Saarinen and Shiva Sander-Tavallaey

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

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

*Dept. of Electrical Engineering,
Linköping University,
SE–581 83 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.

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.

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 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. (2011a), this depen-
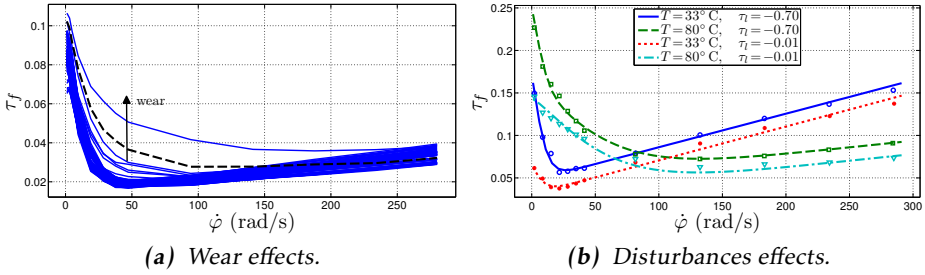
(a) Wear effects.

(b) Disturbances effects.

Figure 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_l 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.*

dency in an industrial robot joint is studied and modeled. A possible diagnostic 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 Figure 1. In Bittencourt et al. (2011a), 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. Furthermore, test-cycles reduce the robot availability which is not desirable from a robot user perspective.

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 Section 2 with an experimental study of wear monitoring in a robot joint. In Section 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 Section 4 with detailed simulation studies of the effects of temperature for the robotics application. Finally, conclusions and possible extensions are given in Section 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[i]), 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 C.1 (Faults are observable). Changes on $\mathbf{f}^j$ affect the measured data $\mathbf{y}^j$.

Assumption C.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 C.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 C.2.

Assumption C.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 C.1, C.2 and C.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 C.4) against the remaining sequences $\mathbf{y}^j$. While Assumption C.1 is necessary, Assumption C.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

---

[i]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.

- How to characterize $\mathbf{y}^j$?

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

Furthermore, Assumptions C.2, C.3 and C.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 multibody dynamic mechanism by

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

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_l, the temperature inside the joint T, and the wear levels w.

Using the introduced notation, the deterministic unknown 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[i]. Since, from (2), it is clear that τ is affected directly by w (satisfying Assumption C.1), only τ is considered of interest and included in $\mathbf{y}$. The remaining variables, φ and its derivatives, load torque τ_l and joint temperature T are considered as disturbances and included in $\mathbf{d}$.

Notice that the effects of φ, its derivatives, and τ_l are defined by the trajectory $\mho$, 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 C.3, notice that they considerably vary along i. If this behavior is also valid for T, then $\mathbf{y}^j$ satisfies Assumption C.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 Figure 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.

---

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

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)),

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

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 estimate is normalized to 1. The quantity $\hat{p}^j(y)$ is a nonparametric smooth empirical distribution estimate (NSEDE) of $\mathbf{y}^j$.

Example 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 C.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 Figure 2a, together with their estimated NSEDEs in Figure 2b. 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 Example 1 and Figure 2, it is possible to see that Assumptions C.2 and C.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.
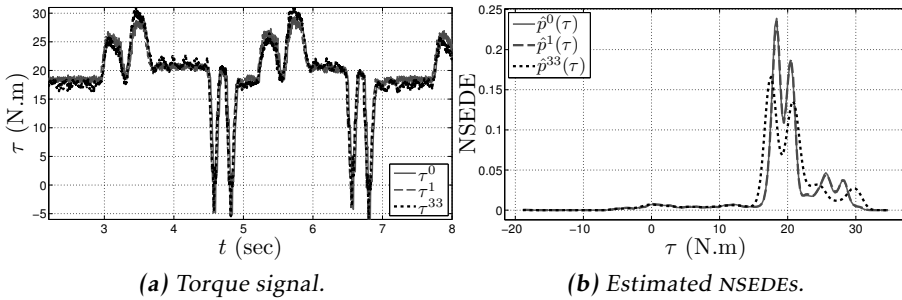
(a) Torque signal.   *(b) Estimated* NSEDE*s.*

Figure 2: (a), torque signals at a joint under accelerated wear tests and their NSEDE*s, (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* NSEDE*s.*

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

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

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 distance, 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 (KL), is however symmetric and also a metric. 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 C.4, fault free data are always available, so that $\mathbf{y}^0$ is known and $\hat{p}^0$ can be evaluated. The quantities $\mathrm{KL}\left(\hat{p}^0\|\hat{p}^j\right)$ can therefore be used as a fault indicator.

----- **Example 2** -----

Application of $\mathrm{KL}\left(\hat{p}^0\|\hat{p}^j\right)$ **for Experimental Wear Monitoring in a Robot**
The same sequence $[\boldsymbol{\tau}^0, \cdots, \boldsymbol{\tau}^{M-1}]$ used in Example 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 $\mathrm{KL}\left(\hat{p}^0\|\hat{p}^j\right)$ are computed for $j = 1, \ldots, M-1$. As shown in Figure 3b, these quantities show a clear response to how the wear

(a) *Friction curves. The colormap relates to j.*

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

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

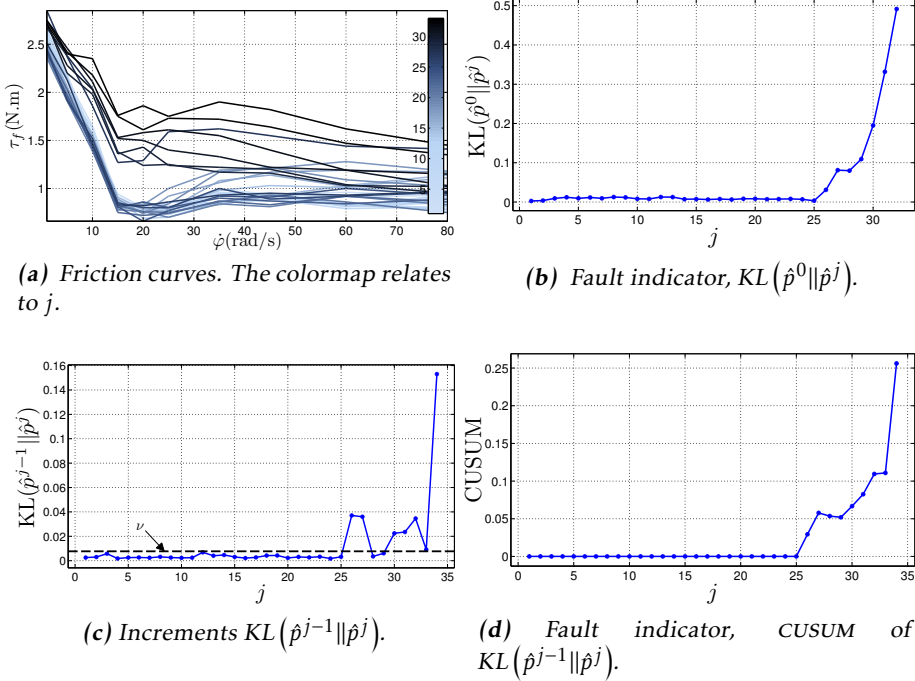(d) *Fault indicator, CUSUM of $KL\left(\hat{p}^{j-1}\|\hat{p}^j\right)$.*

Figure 3: *Monitoring of a wear fault in an industrial robot joint under accelerated wear tests. The friction changes caused by the wear 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 Example 2 is shown in (b). The lower row presents the resulting quantities when monitoring the accumulated changes from Example 3. 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).*

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 Figure 3a.

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 C.2 and C.3 are however limiting in many practical applications. The next sections discuss approaches to relax these assumptions.

3   Monitoring the Accumulated Changes

Because the KL $(\cdot\|\cdot)$ is a metric, it satisfies the triangle inequality, which gives

$$\mathrm{KL}\left(\hat{p}^0\|\hat{p}^j\right) \le \sum_{k=1}^{j} \mathrm{KL}\left(\hat{p}^{k-1}\|\hat{p}^k\right). \tag{6}$$

Since $\mathrm{KL}\left(\hat{p}^{k-1}\|\hat{p}^k\right)$ 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 $\mathrm{KL}\left(\hat{p}^{j-1}\|\hat{p}^j\right)$ 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 3 CUSUM

$$g(j) = g(j-1) + s(j) - \nu \tag{7}$$
$$g(j) = 0 \text{ if } g(j) < 0. \tag{8}$$

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 (7). 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 C.4 is relaxed. The drift parameter can be chosen as

$$\nu = \mu + \kappa\sigma, \tag{9}$$

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

┌─── **Example 3** ───
Application of the CUSUM to $\mathrm{KL}\left(\hat{p}^{j-1}\|\hat{p}^j\right)$ for Experimental Wear Monitoring in a Robot Joint
The real fault case in Example 2 is considered again. Instead of using $\mathrm{KL}\left(\hat{p}^0\|\hat{p}^j\right)$ as a fault indicator, the increments $\mathrm{KL}\left(\hat{p}^{j-1}\|\hat{p}^j\right)$ are computed and the CUSUM filter is used. The drift parameter is chosen as in (9), with $\kappa = 3$ and $[\mu, \sigma]$ are estimated from the first 5 sequences. The resulting quantities are shown in Figures 3c and 3d, with a clear response to the wear increases.

3.1   Monitoring Irregular Data

Let $\mho^j$ denote the conditions not related to faults under which a sequence $\mathbf{y}^j$ was generated. Assumption C.2 requires the whole sequence $\mathbf{Y}^M$ to have been generated under the same $\mho$, so that they are comparable. In principle, the alternative solution of monitoring the accumulated consecutive increments $\mathrm{KL}\left(\hat{p}^{j-1}\|\hat{p}^j\right)$ requires only that $\mho^{j-1}$ and $\mho^j$ are the same, thus relaxing Assumption C.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 $\mho^j$.

Example 4

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 on 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

$$w(j) = w^0 + \frac{w^f - w^0}{2}\xi(j) \tag{10a}$$

$$\xi(j) = 1 + \frac{j - j_m}{\left(1 + |j - j_m|^b\right)^{-b}} \tag{10b}$$

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 $w^0 = 0$ and $w^f = 50$ 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 Figure 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 $\nu(j) = \mu^t + \sigma^t$, where $[\mu^t, \sigma^t]$ 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
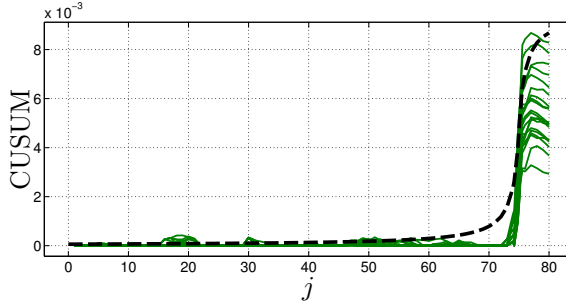
Figure 4: *CUSUM taken over sequential increments $KL\left(\hat{p}^{j-1}\|\hat{p}^{j}\right)$ resulting from 3 different tasks. The increments are mixed at random, 10 cases are presented (solid lines). Also shown are the scaled values of w (dashed) for a comparison.*

a weighting vector $\mathbf{w} \in \mathbb{R}^N$, the weighted data are written as

$$\bar{\mathbf{y}}^j = \mathbf{w} \circ \mathbf{y}^j, \tag{11}$$

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 Section 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 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 w – Linear Discriminant Analyses

Consider that the data set $\mathbf{Y}^M$ is available, 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

$$\bar{\mathbf{Y}}^M \triangleq \left[\bar{\mathbf{y}}^0, \ldots, \bar{\mathbf{y}}^{M_0}, \bar{\mathbf{y}}^{M_0+1}, \ldots, \bar{\mathbf{y}}^{M_1+M_0}\right], \tag{12}$$

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

$$\bar{m}^c \triangleq N^{-1} \sum_{i=0}^{N-1} \left[M_c^{-1} \sum_{j \in \mathcal{C}_c} w_i y_i^j \right] \tag{13}$$

$$= N^{-1} \sum_{i=0}^{N-1} w_i \underbrace{\left[M_c^{-1} \sum_{j \in \mathcal{C}_c} y_i^j \right]}_{\triangleq m_i^c} = N^{-1} \mathbf{w}^T \mathbf{m}^c. \tag{14}$$

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

$$\bar{m}^1 - \bar{m}^0 \propto \mathbf{w}^T(\mathbf{m}^1 - \mathbf{m}^0). \tag{15}$$

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 show 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,

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

$$= N^{-1} \sum_{i=0}^{N-1} w_i^2 \underbrace{\left[M_c^{-1} \sum_{j \in \mathcal{C}_c} (y_i^j - m_i^c)^2 \right]}_{\triangleq s_i^c} \tag{17}$$

$$= N^{-1} \mathbf{w}^T \mathbf{S}^c \mathbf{w}, \tag{18}$$

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

$$\frac{(\bar{m}^1 - \bar{m}^0)^2}{\bar{s}_1 + \bar{s}_0} \propto \frac{\mathbf{w}^T(\mathbf{m}^1 - \mathbf{m}^0)(\mathbf{m}^1 - \mathbf{m}^0)^T \mathbf{w}}{\mathbf{w}^T(\mathbf{S}^1 + \mathbf{S}^0)\mathbf{w}}, \tag{19}$$

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

$$\mathbf{w} \propto (\mathbf{S}^1 + \mathbf{S}^0)^{-1}(\mathbf{m}^1 - \mathbf{m}^0). \tag{20}$$

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 (15) and (20) 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 5: 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,

$$\tau^k : \text{w} = 0, \; T \sim \mathcal{U}[\underline{T}, \underline{T} + \Delta_T], \; k \in \mathcal{C}_0 \tag{21a}$$

$$\tau^l : \text{w} = \text{w}_c, \; T \sim \mathcal{U}[\underline{T}, \underline{T} + \Delta_T], \; l \in \mathcal{C}_1 \tag{21b}$$

where $k \in \mathcal{C}_0$ corresponds to the first M_0 sequences and $l \in \mathcal{C}_1$ are the remaining ones, $\text{w}_c = 35$ is a wear level considered critical to generate an alarm (see Bittencourt et al. (2011a)). Here, T is considered random, with uniform distribution given by $\underline{T} = 30°C$ and $\Delta_T = 40°C$. 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 Figure 5a. 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). Figure 5b presents the ratio for such worst case estimate, which is considered as the optimal weights according to (20). 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 Figure 1. The solid line in Figure 5b is a function approximation of the optimal weights given by

$$w(\dot{\varphi}) = \text{sech}(\beta\dot{\varphi})\tanh(\alpha\dot{\varphi}) \tag{22}$$

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. (2011a), 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 (21), a pair (τ^m, τ^n) is given and the objective is to decide whether the pair is from the same class or not, that is, the two hypotheses are considered

$$\mathcal{H}_0 : m, n \in \mathcal{C}_0 \text{ or } m, n \in \mathcal{C}_1 \tag{23a}$$

$$\mathcal{H}_1 : m \in \mathcal{C}_0, n \in \mathcal{C}_1 \text{ or } m \in \mathcal{C}_1, n \in \mathcal{C}_0. \tag{23b}$$

In view of the framework presented in Section 2, this problem is analyzed by computing the distribution of 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
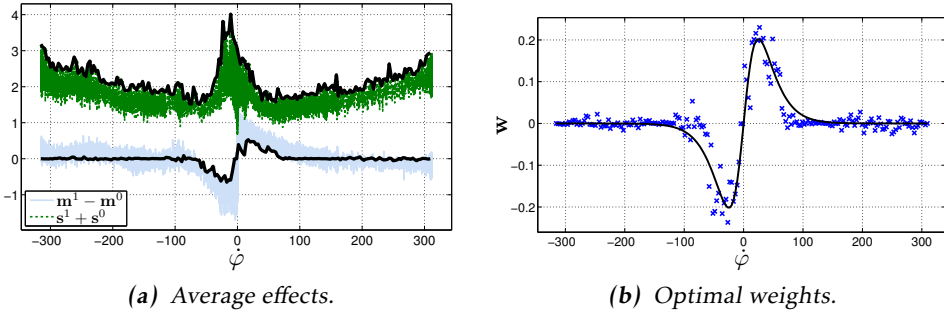
(a) Average effects.

(b) Optimal weights.

Figure 5: *Choice of optimal weights* **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 Figure 5b together with a function approximation (solid). Notice how the optimal region for wear monitoring is concentrated in a narrow speed range.*

the use of the weighting function. For the fixed $P_f = 0.01$, Figure 6a 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 $\Delta_T = 25°C$, data are generated according to (21) for different values of w_c. Similarly, the hypotheses distributions are computed. Figure 6b 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
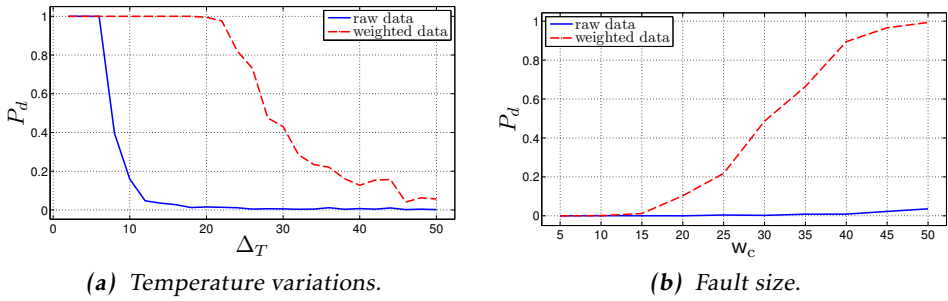
(a) Temperature variations.

(b) Fault size.

Figure 6: *Probability of detection P_d when $P_f = 0.01$ for an abrupt fault with $w_c = 35$ as a function of temperature variations Δ_T (a), and as function of the fault size w_c for $\Delta_T = 25°C$ (b). Notice the considerable improvements when using the weighted data.*

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. Therefore 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 isolation were not addressed.

A Appendix

A.1 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. (2011a) 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_l, temperature T, and wear w.

In the simulation setup, a task $\mho$ is described by a set of reference joint positions through time to the robot, which is controlled with feedforward and feedback control actions, guaranteeing the motion performance. If no variations of w and T are allowed, the torque sequence required for the execution of a task $\mho$ 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.

Bibliography

C. M. Bishop. *Pattern Recognition and Machine Learning.* Springer, 1st ed edition, 2007.

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

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

A. C. Bittencourt, K. Saarinen, S. Sander-Tavallaey, and H. A. Andersson. A method for monitoring of systems that operate in a repetitive manner - application to wear monitoring of an industrial robot joint. In *Proc. of the 2011 PAPYRUS Workshop*, Corsica, France, Oct 2011b.

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

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

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.

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

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

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.

A. De Luca and R. Mattone. An adapt-and-detect actuator fdi scheme for robot manipulators. In *Proc. of the 2004 IEEE International Conference on Robotics and Automation*, volume 5, pages 4975 – 4980 Vol.5, apr. 2004.

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

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

F. Gustafsson. *Adaptive Filtering and Change Detection.* Wiley, Oct. 2000.

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

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

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.

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

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

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

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

Licentiate Theses
Division of Automatic Control
Linköping University

P. Andersson: Adaptive Forgetting through Multiple Models and Adaptive Control of Car Dynamics. Thesis No. 15, 1983.

B. Wahlberg: On Model Simplification in System Identification. Thesis No. 47, 1985.

A. Isaksson: Identification of Time Varying Systems and Applications of System Identification to Signal Processing. Thesis No. 75, 1986.

G. Malmberg: A Study of Adaptive Control Missiles. Thesis No. 76, 1986.

S. Gunnarsson: On the Mean Square Error of Transfer Function Estimates with Applications to Control. Thesis No. 90, 1986.

M. Viberg: On the Adaptive Array Problem. Thesis No. 117, 1987.

K. Ståhl: On the Frequency Domain Analysis of Nonlinear Systems. Thesis No. 137, 1988.

A. Skeppstedt: Construction of Composite Models from Large Data-Sets. Thesis No. 149, 1988.

P. A. J. Nagy: MaMiS: A Programming Environment for Numeric/Symbolic Data Processing. Thesis No. 153, 1988.

K. Forsman: Applications of Constructive Algebra to Control Problems. Thesis No. 231, 1990.

I. Klein: Planning for a Class of Sequential Control Problems. Thesis No. 234, 1990.

F. Gustafsson: Optimal Segmentation of Linear Regression Parameters. Thesis No. 246, 1990.

H. Hjalmarsson: On Estimation of Model Quality in System Identification. Thesis No. 251, 1990.

S. Andersson: Sensor Array Processing; Application to Mobile Communication Systems and Dimension Reduction. Thesis No. 255, 1990.

K. Wang Chen: Observability and Invertibility of Nonlinear Systems: A Differential Algebraic Approach. Thesis No. 282, 1991.

J. Sjöberg: Regularization Issues in Neural Network Models of Dynamical Systems. Thesis No. 366, 1993.

P. Pucar: Segmentation of Laser Range Radar Images Using Hidden Markov Field Models. Thesis No. 403, 1993.

H. Fortell: Volterra and Algebraic Approaches to the Zero Dynamics. Thesis No. 438, 1994.

T. McKelvey: On State-Space Models in System Identification. Thesis No. 447, 1994.

T. Andersson: Concepts and Algorithms for Non-Linear System Identifiability. Thesis No. 448, 1994.

P. Lindskog: Algorithms and Tools for System Identification Using Prior Knowledge. Thesis No. 456, 1994.

J. Plantin: Algebraic Methods for Verification and Control of Discrete Event Dynamic Systems. Thesis No. 501, 1995.

J. Gunnarsson: On Modeling of Discrete Event Dynamic Systems, Using Symbolic Algebraic Methods. Thesis No. 502, 1995.

A. Ericsson: Fast Power Control to Counteract Rayleigh Fading in Cellular Radio Systems. Thesis No. 527, 1995.

M. Jirstrand: Algebraic Methods for Modeling and Design in Control. Thesis No. 540, 1996.

K. Edström: Simulation of Mode Switching Systems Using Switched Bond Graphs. Thesis No. 586, 1996.

J. Palmqvist: On Integrity Monitoring of Integrated Navigation Systems. Thesis No. 600, 1997.

A. Stenman: Just-in-Time Models with Applications to Dynamical Systems. Thesis No. 601, 1997.

M. Andersson: Experimental Design and Updating of Finite Element Models. Thesis No. 611, 1997.

U. Forssell: Properties and Usage of Closed-Loop Identification Methods. Thesis No. 641, 1997.

M. Larsson: On Modeling and Diagnosis of Discrete Event Dynamic systems. Thesis No. 648, 1997.

N. Bergman: Bayesian Inference in Terrain Navigation. Thesis No. 649, 1997.

V. Einarsson: On Verification of Switched Systems Using Abstractions. Thesis No. 705, 1998.

J. Blom, F. Gunnarsson: Power Control in Cellular Radio Systems. Thesis No. 706, 1998.

P. Spångéus: Hybrid Control using LP and LMI methods – Some Applications. Thesis No. 724, 1998.

M. Norrlöf: On Analysis and Implementation of Iterative Learning Control. Thesis No. 727, 1998.

A. Hagenblad: Aspects of the Identification of Wiener Models. Thesis No. 793, 1999.

F. Tjärnström: Quality Estimation of Approximate Models. Thesis No. 810, 2000.

C. Carlsson: Vehicle Size and Orientation Estimation Using Geometric Fitting. Thesis No. 840, 2000.

J. Löfberg: Linear Model Predictive Control: Stability and Robustness. Thesis No. 866, 2001.

O. Härkegård: Flight Control Design Using Backstepping. Thesis No. 875, 2001.

J. Elbornsson: Equalization of Distortion in A/D Converters. Thesis No. 883, 2001.

J. Roll: Robust Verification and Identification of Piecewise Affine Systems. Thesis No. 899, 2001.

I. Lind: Regressor Selection in System Identification using ANOVA. Thesis No. 921, 2001.

R. Karlsson: Simulation Based Methods for Target Tracking. Thesis No. 930, 2002.

P.-J. Nordlund: Sequential Monte Carlo Filters and Integrated Navigation. Thesis No. 945, 2002.

M. Östring: Identification, Diagnosis, and Control of a Flexible Robot Arm. Thesis No. 948, 2002.

C. Olsson: Active Engine Vibration Isolation using Feedback Control. Thesis No. 968, 2002.

J. Jansson: Tracking and Decision Making for Automotive Collision Avoidance. Thesis No. 965, 2002.

N. Persson: Event Based Sampling with Application to Spectral Estimation. Thesis No. 981, 2002.

D. Lindgren: Subspace Selection Techniques for Classification Problems. Thesis No. 995, 2002.

E. Geijer Lundin: Uplink Load in CDMA Cellular Systems. Thesis No. 1045, 2003.

M. Enqvist: Some Results on Linear Models of Nonlinear Systems. Thesis No. 1046, 2003.

T. Schön: On Computational Methods for Nonlinear Estimation. Thesis No. 1047, 2003.

F. Gunnarsson: On Modeling and Control of Network Queue Dynamics. Thesis No. 1048, 2003.

S. Björklund: A Survey and Comparison of Time-Delay Estimation Methods in Linear Systems. Thesis No. 1061, 2003.

M. Gerdin: Parameter Estimation in Linear Descriptor Systems. Thesis No. 1085, 2004.

A. Eidehall: An Automotive Lane Guidance System. Thesis No. 1122, 2004.

E. Wernholt: On Multivariable and Nonlinear Identification of Industrial Robots. Thesis No. 1131, 2004.

J. Gillberg: Methods for Frequency Domain Estimation of Continuous-Time Models. Thesis No. 1133, 2004.

G. Hendeby: Fundamental Estimation and Detection Limits in Linear Non-Gaussian Systems. Thesis No. 1199, 2005.

D. Axehill: Applications of Integer Quadratic Programming in Control and Communication. Thesis No. 1218, 2005.

J. Sjöberg: Some Results On Optimal Control for Nonlinear Descriptor Systems. Thesis No. 1227, 2006.

D. Törnqvist: Statistical Fault Detection with Applications to IMU Disturbances. Thesis No. 1258, 2006.

H. Tidefelt: Structural algorithms and perturbations in differential-algebraic equations. Thesis No. 1318, 2007.

S. Moberg: On Modeling and Control of Flexible Manipulators. Thesis No. 1336, 2007.

J. Wallén: On Kinematic Modelling and Iterative Learning Control of Industrial Robots. Thesis No. 1343, 2008.

J. Harju Johansson: A Structure Utilizing Inexact Primal-Dual Interior-Point Method for Analysis of Linear Differential Inclusions. Thesis No. 1367, 2008.

J. D. Hol: Pose Estimation and Calibration Algorithms for Vision and Inertial Sensors. Thesis No. 1370, 2008.

H. Ohlsson: Regression on Manifolds with Implications for System Identification. Thesis No. 1382, 2008.

D. Ankelhed: On low order controller synthesis using rational constraints. Thesis No. 1398, 2009.

P. Skoglar: Planning Methods for Aerial Exploration and Ground Target Tracking. Thesis No. 1420, 2009.

C. Lundquist: Automotive Sensor Fusion for Situation Awareness. Thesis No. 1422, 2009.

C. Lyzell: Initialization Methods for System Identification. Thesis No. 1426, 2009.

R. Falkeborn: Structure exploitation in semidefinite programming for control. Thesis No. 1430, 2010.

D. Petersson: Nonlinear Optimization Approaches to $\mathcal{H}_2$-Norm Based LPV Modelling and Control. Thesis No. 1453, 2010.

Z. Sjanic: Navigation and SAR Auto-focusing in a Sensor Fusion Framework. Thesis No. 1464, 2011.

K. Granström: Loop detection and extended target tracking using laser data. Thesis No. 1465, 2011.

J. Callmer: Topics in Localization and Mapping. Thesis No. 1489, 2011.

F. Lindsten: Rao-Blackwellised particle methods for inference and identification. Thesis No. 1480, 2011.

M. Skoglund: Visual Inertial Navigation and Calibration. Thesis No. 1500, 2011.

S. Khoshfetrat Pakazad: Topics in Robustness Analysis. Thesis No. 1512, 2011.

P. Axelsson: On Sensor Fusion Applied to Industrial Manipulators. Thesis No. 1511, 2011.