THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Development and evaluation of methods for control and modelling of multiple-input multiple-output systems

Fredrik Bengtsson

Department of Electrical Engineering
Chalmers University of Technology
Gothenburg, Sweden, 2020

Development and evaluation of methods for control and modelling of multiple-input multiple-output systems

Fredrik Bengtsson

ISBN 978-91-7905-381-9

Doktorsavhandlingar vid Chalmers tekniska högskola, Ny serie nr 4848
ISSN 0346-718X
Department of Electrical Engineering

Chalmers University of Technology
SE-412 96 Gothenburg, Sweden
Phone: +46 (0)31 772 1000
www.chalmers.se

Printed by Chalmers Reproservice
Gothenburg, Sweden, September 2020

This thesis has been prepared using LaTeX.

Abstract

In control, a common type of system is the multiple-input multiple-output (MIMO) system, where the same input may affect multiple outputs, or conversely, the same output is affected by multiple inputs. In this thesis two methods for controlling MIMO systems are examined, namely linear quadratic Gaussian (LQG) control and decentralized control, and some of the difficulties associated with them.

One difficulty when implementing decentralized control is to decide which inputs should control which outputs, also called the input-output pairing problem. There are multiple ways to solve this problem, among them using gramian based measures, which include the Hankel interaction index array, the participation matrix and the Σ_2 method. These methods take into account system dynamics as opposed to many other methods which only consider the steady-state system. However, the gramian based methods have issues with input and output scaling. Generally, this is handled by scaling all inputs and outputs to have equal range. However, in this thesis it is demonstrated how this can cause an incorrect pairing. Furthermore, this thesis examines other methods of scaling the gramian based measures, using either row or column sums, or by utilizing the Sinkhorn-Knopp algorithm. It is shown that there are considerable benefits to be gained from the alternative scaling of the gramian based measures, especially when using the Sinkhorn-Knopp algorithm. The use of this method also has the advantage that the results are completely independent of the original scaling of the inputs and outputs.

An expansion to the decentralized control structure is the sparse control, in which a decentralized controller is expanded to include feed-forward or MIMO blocks. In this thesis we explore how to best use the gramian based measures to find sparse control structures, and propose a method which demonstrates considerable improvement compared to existing methods of sparse control structure design.

A prerequisite to implementing control configuration methods is an understanding of the processes in question. In this thesis we examine the pulp refining process and design both static and dynamic models for pulp and paper properties such as shives width, fiber length and tensile index, and various available inputs. We demonstrate that utilizing internal variables (primarily consistencies) estimated from temperature measurements yields improved results compared to using solely measured variables. The measurement data

from the refiners is noisy, sometimes sparse and generally irregularly sampled. This thesis discusses the challenges posed by these constraints and how they can be resolved.

An alternative way to control a MIMO system is to implement an LQG controller, which yields a single control structure for the entire system using a state based controller. It has been proposed that LQG control can be an effective control scheme to be used on networked control systems with wireless channels. These channels have a tendency to be unreliable with packet delays and packet losses. This thesis examines how to implement an LQG controller over such unreliable communication channels, and derives the optimal controller minimizing the cost function expressed in actuated controls.

When new methods of control system design and analysis are introduced in the control engineering field, it is important to compare the new results with existing methods. Often this requires application of the methods on examples, and for this purpose benchmark processes are introduced. However, in many areas of control engineering research the number of examples are relatively few, in particular when MIMO systems are considered. For a thorough assessment of a method, however, as large number of relevant models as possible should be used. As a remedy, a framework has been developed for generating linear MIMO models based on predefined system properties, such as model type, size, stability, time constants, delays etc. This MIMO generator, which is presented in this thesis, is demonstrated by using it to evaluate the previously described scaling methods for the gramian based pairing methods.

Keywords: Control configuration selection, Decentralized control, Gramian based measures, Input-output scaling, LQG control, Unreliable communication links, Delays, Hold-input, MIMO systems, TMP, Tensile Index, Modeling, Uncertain data sets, Linear regression, CTMP, Freeness, Fiber length, Shives.

List of Publications

This thesis is based on the following publications:

[A] **Fredrik Bengtsson**, Torsten Wik, Elin Svensson, "Resolving issues of scaling for gramian based input-output pairing methods". *International Journal of Control*, Taylor & Francis, 2020.

[B] **Fredrik Bengtsson**, Torsten Wik, "Finding feedforward configurations using gramian based interaction measures". *Submitted to Modeling, Identification and Control.*

[C] **Fredrik Bengtsson**, Torsten Wik, "A multiple input, multiple output model generator.". A multiple input, multiple output model generator. Technical report, Department of Signals and Systems, Chalmers University of Technology, 2017.

[D] **Fredrik Bengtsson**, Babak Hassibi, Torsten Wik, "LQG control for systems with random unbounded communication delay". In *Proceeding of the 55th Conference on Decision and Control (CDC)*, pages 1048-1055, 2016.

[E] **Fredrik Bengtsson**, Torsten Wik, "Stochastic optimal control over unreliable communication links". *To be submitted.*

[F] **Fredrik Bengtsson**, Anders Karlström, Torsten Wik, "Modeling of Tensile Index using Uncertain Data Sets". In *Nordic Pulp and Paper Research Journal*, pp 231-242 vol. 35, no. 2, pp. 2020.

[G] **Fredrik Bengtsson**, Anders Karlström, Torsten Wik, "On the modeling of pulp properties in CTMP processes". *Submitted to Nordic Pulp and Paper Research Journal.*

Other publications by the author, not included in this thesis, are:

[H] Sofie Marton, Elin Svensson, Riccardo Subiaco, **Fredrik Bengtsson**, and Simon Harvey, "A steam utility network model for the evaluation of heat integration retrofits–a case study of an oil refinery". *Journal of Sustainable Development of Energy, Water and Environment Systems*, 5(4), pp.560-578.

[I] **Fredrik Bengtsson,** Anders Karlström, Jan Hill and Lars Johansson, "Raw data for Tensile index estimations from a CD72-refiner". Technical report, Chalmers university of technology. Available at https://research.chalmers.se/publication/510615.

Acknowledgments

First I would like to thank my main supervisor Prof. Torsten Wik, for the help and guidance he has provided for the duration of my PHD. After any of our discussions I have always felt that I had a clearer view of my work, and how to proceed. I would also like to thank my co-supervisors Dr. Elin Svensson and Dr. Karin Eriksson for many interesting discussions as well as highly constructive feedback which I feel greatly improved my work.

Moreover I am grateful for Dr. Anders Karlström for the guidance he provided for me in my work on paper refiners. Furthermore I am grateful for Jan Hill, Crister Sandberg, and Johan Sund for helping me get initiated in the paper refining process. I would also like to thank Holmen AB and Norske Skog for providing measurement data.

I have always enjoyed going to work, and for this I would like to thank my friends and colleagues at Systems and Control. We have had many interesting conversations during lunch and fika, which I will sorely miss.

Furthermore I would like to thank Preem AB, not only for financing but also for providing useful data from the Preem plant and for many constructive meetings and discussions with the Preem employees. I am especially grateful to Peter Holmqvist for the work he did in collecting and providing data and schematics, as well as for many good discussions on the Preem plant.

I would also like to thank Process Industrial IT and Automation (PIIA) for the funding they provided for the project.

During my education I have had some truly great teachers, whom I wish to thank. I would like to especially thank Stuart Cameron and Kim Freimann from Malmö Borgarskola for the help and inspiration they provided during my studies.

Finally, I would like to thank my parents for the support and encouragement they have provided throughout my life.

Acronyms

MIMO:	Multiple-input multiple-output
SISO:	Single-input single-output

LQG:	Linear quadratic Gaussian
RGA:	Relative gain array
RIA:	Relative interaction array
PM:	Participation matrix
HIIA:	Hankel index interaction array
MPC:	Model predictive control
TFM:	Transfer function matrix
IM:	Interaction matrix
HSVs:	Hankel singular values
NI:	Niederlinski Index
IMC:	Internal model controller
TCP:	Transmission control protocol
UDP:	User datagram protocol
TMP:	Thermomechanical pulping
CTMP:	Chemical thermomechanical pulping
FZ:	Flat zone
CD:	Conical zone
PQM:	Pulp quality monitor
VIF:	Variance inflation factor

Contents

Part I

Overview

CHAPTER 1

Introduction

A common issue in industrial processes is that interaction between different parts of the plant gives rise to a multiple-input multiple-output (MIMO) system, where the same input may affect multiple outputs, or conversely, the same output is affected by multiple inputs. Such interactions make MIMO systems considerably more complex to control than single input single output (SISO) systems [1].

While there are numerous ways to control MIMO systems, the focus here is on two methods, decentralized control and linear quadratic Gaussian (LQG) control, and ways to solve some of the problems associated with these strategies. Moreover this thesis discusses tools and methods to evaluate the control of MIMO systems.

One method to control a MIMO system is to divide it into subsystems of one input and one output and implement SISO controllers for each of the subsystems. This control strategy is called decentralized control and remains widely used in industry [2]. It has several advantages compared to implementing a MIMO controller for the entire system, as it allows the use of relatively easy to design low dimensional controllers. Moreover, it is less vulnerable to sensor and actuator failures than more complex control schemes that try to control

the entire system with one overarching control scheme. However, a decentralized control scheme leads to the input-output pairing problem: which inputs should be used to control which outputs to best fulfill the control objectives?

Numerous methods have been proposed to find a suitable input-output pairing, many of which are discussed in [3]. The most widely used is the Relative Gain Array (RGA)[4] and modifications of it, such as the dynamic RGA and the Relative Interaction Array (RIA)[5]. Relatively recently a new group of input-output pairing methods have been introduced, namely the gramian based methods. This group includes the Σ_2 method [6], the participation matrix (PM)[7] and the Hankel interaction index array (HIIA)[8]. These methods use the controllability and observability gramians to create an interaction matrix, which gives a gauge of how much each input affects each output. An attractive property of these interaction matrices is that they can be used to determine both a decentralized controller structure and a sparse structure (a structure which includes feed-forward and/or MIMO blocks). Moreover, the gramian based measures take into account system dynamics and not only the steady state properties of the system.

The gramian based methods, however, differ from the RGA and its variants in that they suffer from issues of scaling, in the sense that the results of the methods vary depending on input and output scaling. There is a commonly suggested method to solve this problem, namely scaling the inputs and outputs from zero to one, presented in [9]. However, in Paper A we demonstrate that this method is insufficient in some situations. We then proceed to propose a new method of scaling, based on the Sinkhorn-Knopp algorithm [10], which removes the problems of scaling dependency.

Sometimes interactions between the different inputs on the outputs result in a decentralized control scheme yielding poor results. One possible remedy to this is to expand the decentralized control structure to include decoupling feedforward to remove the most problematic interactions. This yields what is called a sparse controller structure. However, this requires determining which interactions that are appropriate to remove with feedforward, and which ones where implementing feedforward may create interactions that result in a poorer control outcome.

The gramian based methods can also be used to determine which elements are appropriate for feed forward, and a rule of thumb of how to do so was presented in [7]. In Paper B we investigate more thoroughly how to design

sparse controls using the gramian based measures. Furthermore we investigate how sparse control design needs to be adapted when using the new method of scaling proposed in Paper A.

Another control scheme that can be used to control MIMO systems is LQG control. This is a well established method developed in the 1960s which aims to find the control scheme minimizing a quadratic cost function. While LQG control was relatively quickly adopted for the control of ships and space vehicles, the process industry was generally slow to adopt LQG control [11]. As industry has become more interested in use of networked control systems to perform remote control of factories [12], control over wireless channels is an issue that has risen into prominence. Wireless communication is prone to issues of packet losses and delays, which poses difficulties when implementing control schemes. Here LQG control is one of the proposed methods to carry out control in such situations [13] and in this thesis it is examined how to optimally implement LQG control over unreliable channels.

In Paper D we examine how to optimally implement LQG control in the case where there is such a random unbounded delay with a specified probability between the controller and actuator and in Paper E we expand this to present a solution for a more general type of unreliable communication, with any type of random delay, as well as packet losses. Moreover, we properly utilize the knowledge of which control signals have arrived to derive a solution which yields a lower cost than in Paper D.

When new methods of design and analysis are introduced in the control engineering field, it is important to compare the new results with those of existing methods, to evaluate the extent of the improvement and if it only applies to systems with certain properties. However, it is not always apparent how this can be accomplished in an unbiased and consistent way. To address this, in Paper C we propose a MIMO system generator, which allows for the creation of a large number of random MIMO systems with user defined properties. In Paper A and Paper B we demonstrate how the MIMO generator can be used to perform statistical analysis for evaluation of new methods to compare their results with those of existing methods.

To implement and design control structures for MIMO systems, an understanding of the system in question is necessary. Full MIMO controllers such as LQG control generally require a full system model. Decentralized controllers generally do not require full system knowledge. However, they do require some

understanding of how the inputs affect the outputs as decentralised control implementation includes selecting which inputs are appropriate to use. As there are often more available inputs than outputs to control, an assessment of how the various inputs affect the outputs is often necessary as a first step when designing a decentralised control structure. In this thesis we will derive models for a paper refiner, assessing how the available inputs of the refiner affect the resulting pulp or paper.

A important measure of paper quality is tensile index. However, it requires time and resource consuming manual measurements, which in turn are quite unreliable [14]–[16]. Hence, considerable benefits can be gained from modelling and estimating tensile index. The modeling approach in mechanical pulping processes, has been that external variables, such as specific energy (i.e. the ratio between motor load and production), dilution water added to the refiners, plate gaps (disc clearance) etc., should be used for process follow up of pulp and handsheet properties [17]–[22]. However, when using external variables as predictors, the process non-linearities tend to negatively affect the result. To cope with that soft sensors, describing physical phenomena in the refining zone, have been developed during the last decade [23]–[26]. The soft sensor's outputs can be seen as estimates of internal variables (such as fiber residence time, consistency profile, forces on bars, distributed defibration, thermodynamic work etc.), which are difficult to measure directly in the process. Typically, such soft sensors are non-linear but have become important for advanced process optimization. Specifically, consistency and fiber residence time have been candidates for such activities for some years, as they provide a link to e.g. tensile index, mean fiber length and Somerville shives, [14]–[16], [27]–[29]. In Paper F modelling tensile index using both external and internal variables is explored. It is shown that internal variables generally outperform external variables when deriving models for tensile index.

Due to the cost of measurements the available data set of tensile index is quite limited. As the measurements are also quite unreliable Paper F also explores how to best derive models subject to these limitations.

In Paper G our modelling of the refining process is expanded to include pulp properties, such as freeness, shives width and fiber length. Unlike tensile index these can be measured automatically, so considerably larger data sets were available. However, just as in the case of tensile index, the measurements were noisy and irregularly sampled. As in the case of tensile index, models

were derived using a combination of internal and external variables, and it is shown that reasonable models can be found for shives width and fiber length estimation. The greater size of the data sets also allowed for the use of dynamic models, even though the generally low sampling rate resulted in these models performing similarly to static models.

The thesis is organized as follows: In Chapter 2 the control configuration problem is presented, and common methods to find an input-output pairing are discussed with special focus on the gramian based pairing measures. In Chapter 3 the difficulties the gramian based pairing measures have with input and output scaling is discussed, along with possible methods to resolve this issue. In Chapter 4 methods to evaluate and compare new methods are discussed. In Chapter 5 the methods described in Chapters 2-4 are demonstrated. In Chapter 6 LQG control is described and in Chapter 7 control of systems with delay is discussed. A short introduction to pulp refining is presented in Chapter 8, and in Chapter 9 a discussion of modelling techniques is presented. In Chapter 10 the papers included in this thesis are summarized and in Chapter 11 possible future work is discussed.

1.1 Main contributions

The main contributions in this thesis are as follows:

1. A new method for scaling the gramian based input output pairing methods is proposed, which removes the scaling dependency of the gramian based measures.
   a) The method is shown to outperform existing measures on a large number of systems.
   b) It is also shown how the method should be adapted in order to identify feedforward structures.
2. The construction of a multiple-input multiple-output (MIMO) system generator, which can be used to evaluate and compare different control methods.
3. The derivation of optimal LQG control in the case where there are unbounded delays and packet losses in the communication channel between the actuators and controller, and between the sensors and controller.

   a) Two solutions are presented, one less computationally intense solution which does not fully utilize all available information, and one optimal version in which all available information is fully utilized.

   b) In simulation it is demonstrated that the derived controllers significantly outperform traditional LQG control in cases of delays.

4. The derivation of models for predicting tensile index, shives width and fiber length from process measurements.

   a) Both static and dynamic models are evaluated.

   b) It is further explored how to best deal with issues such as unreliable, sparse and irregularly sampled data, which are typical when modelling paper refiners.

CHAPTER 2

Control configuration selection

A key property of integrated plants is that they tend to have numerous outputs (controlled variables) and numerous possible inputs (manipulated variables). There are two basic strategies that can be implemented here. One alternative is to treat the control of the entire system as one control problem and design a control scheme for the entire system, using a multiple-input multiple-output control strategy such as, for example, model predictive control (MPC). Alternatively, one can divide the system into subsystems and design a separate control scheme for each subsystem. While designing a control scheme for the entire system may yield the best solution in theory, this solution also tends to be complex to implement and maintain as it generally requires a good model of the entire system. Furthermore, a single actuator or sensor failure may jeopardize the entire control scheme.

Splitting the system into subsystems can alleviate this problem as each subsystem has a control scheme designed independently of the other subsystems. An extreme case of this is the decentralized control structure, where the system is divided into subsystems of one input and one output. This method is commonly used in industrial processes, as it is straightforward to implement using simple PI or PID controllers.

However, to implement a decentralized control structure two problems need to be resolved. Firstly, if there are more inputs than outputs available, decisions have to be made regarding which inputs will not be used (as each output here is controlled only by one input). When this is done one needs to decide which input is to control which output. This is known as the input-output pairing problem, which is one of the focus areas in this thesis (it is assumed that the decision of which inputs to use has already been made).

2.1 The Input-output pairing problem

As previously stated, the input-output pairing problem consists of choosing which input should control which output using a decentralized control scheme. While in industry this is still sometimes done using rules of thumb and experience [30], there are pairing methods which give systematic ways to determine the input-output pairings. These pairing method analyze some properties of the system and from there find a recommended pairing. While these methods often find pairings that allow for good control, there are no guarantees of optimality from any of the methods as there is no definition of what an optimal pairing may be. Moreover, different pairing methods may give different recommended pairings, which presents additional difficulties when determining which pairing to use.

2.2 The transfer function matrix

To use most pairing methods the MIMO system is defined using its transfer function matrix (TFM) which describes the interactions between the outputs and inputs of a MIMO system as:

$$Y = G(s)U \tag{2.1}$$

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$$

$$U = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{bmatrix}$$

$$G(s) = \begin{bmatrix} g_{11}(s) & g_{12}(s) & \cdots & g_{1N}(s) \\ g_{21}(s) & g_{22}(s) & & \\ \vdots & & \ddots & \\ g_{N1}(s) & & & g_{NN}(s) \end{bmatrix} \tag{2.2}$$

with y_1,...,y_N being the system's outputs, u_1,...,u_N being the system's inputs and $G(s)$ is the TFM of the system. Note that $G(s)$ is a square matrix as the goal is to match each input with one output (and we have already determined which inputs to use).

2.3 RGA

The most common pairing method is the RGA [4], which determines a pairing by comparing the open loop and closed loop properties of the system. It is traditionally calculated from the static gain of the system's TFM as

$$RGA = G(0) \circ G(0)^{-T},$$

with superscript $-T$ denoting the inverse transpose of the matrix and $\circ$ denoting element-wise multiplication. To find a pairing from the RGA matrix one selects the pairing with elements closest to 1, while avoiding negative elements. Explicitly, if the element of row i and column j in the RGA is close to 1, then u_j should be used to control output y_i.

An important property of the RGA is that it is scaling independent, which means that it gives the same results regardless of the scaling of the outputs and inputs. However, it has a few limitations, one of which is that it only takes into account two way interaction. As a consequence interactions from a triangular TFM would not appear in the RGA. Moreover, the static RGA does not take into account system dynamics, including delays. However, it can be expanded with the dynamic RGA [31], which examines a frequency range rather than the zero frequency. The dynamic RGA, though, is based

on the assumption of perfect closed loop control for all frequencies it covers, which is unrealistic for high frequencies [30].

2.4 Gramian based measures

Another group of input-output pairing methods which will now be examined and henceforth be referred to as the gramian based measures are the Σ_2 method, the participation matrix (PM) and the Hankel interaction index array (HIIA). These methods examine each of the transfer functions of the TFM separately to gauge the impact of each input on each output. Unlike the static RGA they take into account the system's dynamics and not only its steady-state properties. The gramian based measures (PM, HIIA and Σ_2) can be calculated from a system's TFM [6]–[8]. Given a TFM as described in (2.2) each measure generates an interaction matrix (IM). For the HIIA and Σ_2 it is generated by

$$[\Gamma]_{ij} = \frac{||g_{ij}(s)||}{\sum_{k,l} ||g_{k,l}(s)||}, \; i, \; j = 1, 2, ..., N,$$

using the Hankel norm and 2-norm for the HIIA and Σ_2, respectively. The PM is derived in a similar fashion, but it uses the squared Hilbert-Schmidt norm, i.e.

$$[\Gamma]_{ij} = \frac{||g_{ij}(s)||^2_{HS}}{\sum_{k,l} ||g_{kl}(s)||^2_{HS}}.$$

Hilbert-Schmidt norm and Hankel norm

The Hilbert-Schmidt norm and Hankel norm both utilize the Hankel singular values (HSV) of the system. These are defined as

$$\sigma_H^{(i)} = \sqrt{\lambda_i},$$

where $\lambda_1, ..., \lambda_n$ are the eigenvalues of PQ, with P being the controllability gramian and Q being the observability gramian. Thus this is a gauge of a combined controllability and observability of the system. The Hilbert-Schmidt norm is the square root of the sum of the squared HSVs of the system, while the Hankel norm is the maximum HSV.

H_2 norm

The H_2 norm, which is used for the Σ_2 method can be written as

$$||g_{ij}(s)||_2 = \sqrt{\frac{1}{2\pi}\int_{-\infty}^{\infty}|g_{ij}(j\omega)|^2 d\omega}.$$

It is proportional to the integral of the squared magnitude of the bode plot, and can be seen as a measure of the energy in the impulse response.

Determination of the pairing

For the gramian based measures the generated IM is used to determine the pairing. With an interaction matrix Γ of

$$\Gamma = \begin{bmatrix} \gamma_{11} & \cdots & \gamma_{1N} \\ \vdots & \ddots & \vdots \\ \gamma_{N1} & \cdots & \gamma_{NN} \end{bmatrix} \tag{2.3}$$

the pairing that has the largest total interaction from the IM is preferred. For instance, a diagonal pairing matching u_1 with y_1, u_2 with y_2 etc, would have a total interaction of

$$\sum_{i=1}^{N}\gamma_{ii}$$

while an anti-diagonal pairing would have a total interaction of

$$\sum_{i=1}^{N}\gamma_{(N+1-i)i}.$$

When an initial pairing has been determined, the control structure can be expanded to include feedforward by selecting additional elements from the IM. If γ_{1N} is large but not included in the original pairing one can still include the interaction by using feedforward on the control of y_1 to compensate for the impact of u_N.

Design of feedforward control structures

Once a decentralized control structure has been found it can be expanded to include feedforward blocks. To understand why one would wish to do this, we begin by examining a 3 by 3 system, i.e.

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} G_{11}(s) & G_{12}(s) & G_{13}(s) \\ G_{21}(s) & G_{22}(s) & G_{23}(s) \\ G_{31}(s) & G_{32}(s) & G_{33}(s) \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}.$$

Let us assume that the inputs and outputs have been ordered such that our decentralized controller design has a diagonal pairing where y_i is controlled by u_i $\forall i$. Now, u_1 will also affect y_2 and y_3 by $G_{21}(s)$ and $G_{31}(s)$, respectively. If u_1 affects y_3 to such an extent that it poses a problem, this can ideally be resolved by using the feedforward

$$u_3 = u_3^* - \frac{G_{31}(s)}{G_{33}(s)} u_1, \tag{2.4}$$

where u_3^* is the control signal from the decentralized controller and we assume $\frac{G_{31}(s)}{G_{33}(s)}$ is stable and proper. If we implement this feed-forward loop we will have removed the direct effect of u_1 on y_3. However, there are other consequences of this implementation since the change of u_3 will also affect y_1 and y_2. If these interactions are significant the feed-forward loop might do more harm than good. Having this in mind, we examine how the IM can be used to determine when feed-forward might be appropriate.

Consider an interaction matrix

$$\Gamma = \begin{bmatrix} \gamma_{11} & \cdots & \gamma_{1N} \\ \vdots & \ddots & \vdots \\ \gamma_{N1} & \cdots & \gamma_{NN} \end{bmatrix}.$$

First we choose the elements for the decentralized pairing as described previously and assume, without loss of generality, that the pairing elements are on the diagonal. After this, we look in the interaction matrix for large elements not yet selected for pairing. The current method for determining feedforward is simply to use the largest elements not selected for pairing [7]. However, doing this means that other potential interactions are not taken into account. For example, assume that γ_{N1} is large, making u_1 a potential candidate for

feed-forward. However, as described in the example, this will impact u_N, which will not only impact y_N, but also the other outputs. A gauge of the size of this impact is $\sum_{i=1}^{N-1} \gamma_{iN}$. If these values are very large then the IM indicates that adding the described feed-forward of u_1 is unwise. To determine the use of feed-forward in the general case we therefore create a new matrix Γ^*, whose elements are defined by

$$\gamma_{ij}^* = \gamma_{ij} - \rho \sum_{\substack{k=1 \\ k \neq i}}^{N} \gamma_{ki},$$

where ρ is a tuning parameter. With this new IM, the largest elements where $i \neq j$ are chosen for feed-forward until the sum of elements chosen (both for control and feedforward) is larger than 0.7, a rule of thumb for gramian based measures [9].

Delays

Continuous time gramian based measures struggle to appropriately deal with delays. This as the Σ_2 method is completly unaffected by delays, and the Hankel singular values of systems with delays are problematic, as continuous time systems with delays are of infinite order. One solution for this is to discreatize the system and implement the methods on the discrete time system, as discussed in [9]. A pairing found on the discrete time system can then be implemented on the continuous time system. Note that when implementing decoupling feedforward on systems with delays some decouplings may not be possible as they would be non-causal.

2.5 Niederlinski Index

The Niederlinski Index (NI) can be used to determine a necessary condition for a decentralized closed loop system to be stable [32]. Consider a system described by a TFM $G(s)$ controlled by a decentralized and diagonal controller $C(s)$ with integral action. If $G(s)$ is stable, $G(s)C(s)$ is proper, and all SISO control loops (created by opening the other loops) are stable, a necessary condition for the existence of a stable control scheme with integral action is

$$NI = \frac{\det[G(0)]}{\prod_{i=1}^{N} g_{ii}(0)} \geq 0,$$

where $g_{ii}(0)$ refers to the diagonal elements of $G(0)$. The NI can be used in combination with other pairing methods when determining pairing. That is to say that one discards solutions which have a negative NI, even if they are recommended by the control configuration method (instead choosing the best solution which has a positive NI). Note that the NI is a necessary but not sufficient condition for closed loop stability, so there is still a risk of unstable pairings even when using the NI.

CHAPTER 3

Scaling the gramian based measures

The gramian based measures are based on various norms. Norms have the property that

$$||\alpha g_{ij}(s)|| = |\alpha|||g_{ij}(s)||,$$

where α is a scalar constant.

This means that for the gramian based measures input and output scaling will affect the recommended pairing. For example, if one would have a system as in (2.2), which would yield the IM (2.3), and one was to change the scaling of the first input such that $\tilde{u}_1 = u_1/\alpha$ is the new input, the scaled TFM would become

$$G^*(s) = \begin{bmatrix} \alpha g_{11}(s) & g_{12}(s) & \cdots & g_{1n}(s) \\ \alpha g_{21}(s) & g_{22}(s) & & \\ \vdots & & \ddots & \\ \alpha g_{n1}(s) & & & g_{nn}(s) \end{bmatrix}, \tag{3.1}$$

which in turn would yield an IM

$$\Gamma^* \propto \begin{bmatrix} |\alpha|\gamma_{11} & \cdots & \gamma_{1N} \\ \vdots & \ddots & \vdots \\ |\alpha|\gamma_{N1} & \cdots & \gamma_{NN} \end{bmatrix}. \tag{3.2}$$

This IM may yield a different recommended pairing than (2.3) depending on α. Consequently, as different scalings of the system may yield different results, emphasis needs to be placed on how to best scale the system when using one of the gramian based measures to find a pairing. Generally, this is resolved by scaling the inputs and outputs from 0 to 1, setting zero to the lowest value they are likely to reach and 1 to the highest value [9]. However, there are other methods to scale the system, which will be discussed in the following sections.

3.1 Row or column scaling

Each column in the IM corresponds to the interactions from one input, while each row corresponds to the interactions affecting one output. One way to scale the system prior to pairing is to divide the elements in each column of the IM by the corresponding column sum. This was presented in [33] for the Σ_2 method, and ensures that when conducting the pairing algorithm, equal importance is given to each input. In the new IM (Γ_c) the scaled elements would become:

$$[\Gamma_c]_{ij} = \frac{[\Gamma]_{ij}}{\sum_{k=1}^{N}[\Gamma]_{kj}},$$

where Γ_c is an interaction matrix with normalized columns. If one instead wishes to ensure that equal importance is given to each output, one could instead chose to normalize the rows, which gives an interaction measure (Γ_r) defined by

$$[\Gamma_r]_{ij} = \frac{[\Gamma]_{ij}}{\sum_{k=1}^{N}[\Gamma]_{ik}}.$$

3.2 Sinkhorn-Knopp algorithm

By scaling the columns or rows one can guarantee that equal importance is given to either each input or each output when determining the pairing. However, if one wishes to have both the columns and rows scaled, the Sinkhorn-Knopp algorithm can be used. This algorithm combines row and column scaling by alternating between normalizing the rows and normalizing the columns. In cases where the matrix can be made to have positive elements on the diagonal (as is always the case with gramian based measures) this algorithm is guaranteed to converge to a matrix that will have both rows and columns normalized [10]. While the Sinkhorn-Knopp algorithm can be implemented by simply alternating between dividing the elements in each column of the IM by the corresponding column sum and dividing the elements in each row by the corresponding row sum, it can also be implemented as described in [34], i.e.,

$$\begin{aligned} r_0 &= e \\ c_{k+1} &= \mathcal{D}(\Gamma^T r_k)^{-1} e \\ r_{k+1} &= \mathcal{D}(\Gamma c_{k+1})^{-1} e, \end{aligned}$$

where e is a vector of ones, and $\mathcal{D}(x)$ turns a vector into a diagonal matrix by creating a matrix with the elements of the vector on its diagonal and zeros in all remaining positions. The scaled IM then becomes

$$\Gamma_{SK} = \mathcal{D}(r)\Gamma\mathcal{D}(c).$$

To calculate how far the solution is from being perfectly scaled (that is having both column and row sums equal to one), one can use the following formula [34]:

$$err_k = ||c_k \circ \mathcal{D}(c_{k+1})^{-1} - e||_1,$$

where $\circ$ denotes element-wise multiplication. This expression can be used as a stopping criterion.

Scaling the IMs with the Sinkhorn-Knopp algorithm has the additional benefit of removing the impact of input and output scaling on the IMs. Using the Sinkhorn-Knopp algorithm to scale the system will yield the same IM,

regardless of what the original scaling of the system was.

In Paper A, we compare the Sinkhorn-Knopp scaling with alternative scalings on a large number of randomly generated MIMO systems and find that it performs significantly better than the alternative scaling methods.

3.3 A demonstrative example

To demonstrate the importance of scaling the IM we will implement input-output pairing on the FS configuration of a heat-integrated distillation column [35],

$$G(s) = \begin{bmatrix} \frac{4.45}{(14s+1)(4s+1)} & \frac{-7.4}{(16s+1)(4s+1)} & 0 & \frac{0.35}{(25.7s+1)(2s+1)} \\ \frac{17-3e^{-0.9s}}{(17s+1)(0.5s+1)} & \frac{-41}{(21s+1)(s+1)} & 0 & \frac{9.2e^{-0.3s}}{20s+1} \\ \frac{0.22e^{-1.2s}}{(17.5s+1)(4s+1)} & \frac{-4.66}{(13s+1)(4s+1)} & \frac{3.6}{(13s+1)(4s+1)} & \frac{0.042(78s+1)}{(21s+1)(11.6s+1)(3s+1)} \\ \frac{1.82e^{-s}}{(21s+1)(s+1)} & \frac{-34.5}{(20s+1)(s+1)} & \frac{12.2e^{-0.9s}}{(18.85s+1)(s+1)} & \frac{-6.92e^{-0.6s}}{(15s+1)(4s+1)} \end{bmatrix}$$

Using the HIIA we can derive an IM for the system,

$$\Gamma = \begin{bmatrix} \mathbf{0.034} & 0.056 & 0 & 0.0025 \\ 0.118 & \mathbf{0.28} & 0 & 0.0593 \\ 0.013 & 0.24 & \mathbf{0.0845} & 0.0452 \\ 0.0016 & 0.036 & 0.0276 & \mathbf{0.0008} \end{bmatrix},$$

where the bold numbers indicate the pairing which yields the largest sum. As can be seen the values in the second and third rows are generally considerably larger then those of the other rows. This means that greater importance is given to finding an optimal match for the outputs of y_2 and y_3, while those outputs which correspond to rows with less interaction are given lesser importance. This leads to that the input that can be seen to have the least affect on y_4 (i.e. u_4) is chosen to control it.

If we rescale the IM using the different scaling methods we get,

$$\Gamma_c = \begin{bmatrix} 0.051 & 0.023 & 0 & \mathbf{0.0057} \\ \mathbf{0.18} & 0.12 & 0 & 0.14 \\ 0.019 & 0.097 & \mathbf{0.19} & 0.11 \\ 0.0024 & \mathbf{0.015} & 0.062 & 0.0019 \end{bmatrix}$$

$$\Gamma_r = \begin{bmatrix} \mathbf{0.092} & 0.1514 & 0 & 0.0067 \\ 0.064 & 0.1538 & 0 & \mathbf{0.032} \\ 0.0082 & \mathbf{0.1567} & 0.055 & 0.030 \\ 0.0061 & 0.1359 & \mathbf{0.105} & 0.0031 \end{bmatrix}$$

$$\Gamma_{SK} = \begin{bmatrix} \mathbf{0.15} & 0.073 & 0 & 0.029 \\ 0.082 & 0.059 & 0 & \mathbf{0.11} \\ 0.011 & \mathbf{0.060} & 0.079 & 0.10 \\ 0.0089 & 0.059 & \mathbf{0.17} & 0.012 \end{bmatrix}$$

As can be seen, with the new scaling methods, using row or SK scaling we find a pairing in which, while some compromises are needed, each output is controlled by an input that seems to have at least a moderate amount of interaction. With column scaling, however, we see that y_1 is controlled using u_4, which has very little affect on y_1.

If we simulate the different control configurations, we find that the configuration from using row or SK scaling performs very well, while the configuration recommended by the unscaled IM has a quadratic cost of several magnitudes higher, and the configuration from the column scaled IM is unstable. In Papers A and B we will further explore the methods of control configuration selection, investigating many different cases and systems.

CHAPTER 4

Evaluation of control methods

Whenever a new method, or a change to an existing method is proposed, it needs to be evaluated to determine if it offers a significant improvement. Moreover, in cases where there are numerous competing methods to solve the same problem (such as the input-output pairing problem) there is a need of a way to compare the methods and determine for which types of system each method is preferable.

4.1 Generation of system models for evaluation

When analyzing new methods for control system design, it is common to demonstrate their benefits on one or a few example systems. While this is a useful way to demonstrate a new method, it does not easily allow for general conclusions of the strengths and limitations of the new method. To do this it would be beneficial to implement the method on a large number of systems with varying properties. For single-input single-output system a large batch of process models have been collected for such evaluations [36], but there is no similar batch for MIMO systems.

In Paper C we present a MIMO model generator which allows for the gen-

eration of a large number of MIMO systems to enable comprehensive testing on MIMO systems. The MIMO model generator generates TFMs with predefined properties such as system size, stability, time constants, delays etc. It is implemented in MATLAB and the code is freely available [37].

4.2 Determination of a cost

To compare different methods there needs to be a method to evaluate how well the controller performs on a given system. A well established method to assess the performance of control systems is to evaluate its response to reference steps and to various types of disturbances by integrating the squared deviation from the reference over a specified time h, i.e.

$$c = \int_0^h (R(t) - Y(t))^T (R(t) - Y(t)) dt, \tag{4.1}$$

where c is the derived cost, $R(t)$ is a vector containing the reference signals, and $Y(t)$ is a vector of the outputs. Typical disturbances one may test are step disturbances on the inputs and high frequency noise on the outputs. This cost can be expanded to include a cost on the control inputs, for example

$$c = \int_0^h (R(t) - Y(t))^T Q_1 (R(t) - Y(t)) + U(t)^T Q_2 U(t) dt,$$

where Q_1 and Q_2 are user defined matrices, used to weight the different parts of the cost.

4.3 Comparison of costs

While the above cost works well to evaluate different controllers on one single system, for a thorough comparison of control methods one would need to evaluate more than one system. However, the costs are not immediately comparable between different systems, as the systems may be of different scale and of varying difficulty to control. To allow comparison between different controlled systems the costs can be normalized for each control configuration on the system using the following equation to produce a score for each configuration:

$$S = \frac{c_{min}}{c}, \tag{4.2}$$

where S is the score of the configuration, c is the configuration's cost, and c_{min} is the lowest cost of all tested configurations for the system. This ensures that each configuration has a score from 0 to 1 for each system, which allows comparisons to be made for the result on different systems. In the comparisons presented here the score is set to zero for a control scheme that does not yield stable results.

4.4 Controller Tuning

When performing evaluations and comparisons in cases where controller design is not the focus of the evaluation, for instance in cases of input output pairing, controllers needs to be implemented in a generalized and consistent way that yields reasonable results without favoring one method over the other. There are numerous methods to design PID controllers automatically, some of which are discussed in [38]. The methods that will be discussed here are lambda tuning and internal model controller (IMC) tuning. They are among the most common methods for commercial auto-tuners [39], making them reasonable methods to be used for comparison purposes (as they are fairly likely to be applied when the control system is implemented in the industry).

Lambda controller tuning

The lambda method [40]–[42], is a two step procedure where for non-integrating systems, the first step is to approximate the SISO transfer function by a first order system with dead time, i.e

$$\hat{G}(s) = \frac{K}{1+Ts}e^{-Ls}.$$

The next step is to determine a PI controller

$$C(s) = K_p(1 + \frac{1}{T_i s}),$$

where the controller parameters are derived from $\hat{G}(s)$ according to

$$\begin{aligned} K_p &= \frac{1}{K}\frac{T}{L+\lambda} \\ T_i &= T \end{aligned}$$

where λ is a tuning parameter. This implementation yields,

$$\hat{G}(s)C(s) = \frac{e^{-Ls}}{(L+\lambda)s},$$

For the case when $L = 0$, the closed loop system becomes,

$$\frac{\hat{G}(s)C(s)}{1+\hat{G}(s)C(s)} = \frac{1}{1+\lambda s},$$

and thus λ can be seen as the targeted time constant of the closed loop system. A common choice of λ is to set it to T, giving the closed loop system the same time constant as the first order plus dead time approximation of the process.

Other than choice of λ, every step in implementing this control scheme can be done automatically, and hence do not require any user input that may add bias to the results.

IMC controller

An alternative to lambda tuned controllers, is to use IMC tuning, which uses a model of the system to cancel out as much of the system dynamics as possible. An IMC tuned controller can be implemented as described in [43], i.e., given a stable transfer function model $G(s)$ of the system, one starts by factorizing the model into two parts:

$$G(s) = \tilde{g}_+(s)\tilde{g}_-(s)$$

such that $\tilde{g}_+(s)$ contains the delays and the non minimum phase zeros of $G(s)$, while $\tilde{g}_-(s)$ contains the remaining dynamics. This ensures that $\tilde{g}_-^{-1}(s)$ is stable. A controller can then be implemented as

$$C = \frac{f(s)\tilde{g}_-^{-1}(s)}{1-f(s)\tilde{g}_+(s)},$$

where

$$f(s) = \frac{1}{(1+\epsilon s)^q}$$

is a user designed filter, ϵ is a tuning parameter and q is chosen such that the controller is proper. This results in a closed loop structure of:

$$\frac{G^*(s)C(s)}{1+G^*(s)C(s)} = f(s)\tilde{g}_+(s).$$

Note that for minimum phase systems (i.e. $g_+ = 1$), we have the same closed loop dynamics structure as the Lambda tuned system without delays. Thus, ϵ can be chosen using the same reasoning as for Lambda tuning for minimum phase systems. For non-minimum phase system one can use

$$\epsilon = \eta Z,$$

where Z is the largest time constant of the model's non-minimum phase zeros and η is a tuning parameter, typically around 1.

CHAPTER 5

An illustrative example

To illustrate some of the methods discussed in Chapters 2-4 we evaluate the scaling methods on first order plus dead time systems. These types of system models are common in the process industry, where they are often derived experimentally. To begin we use our MIMO generator to generate a single first order plus dead time system,

$$G(s) = \begin{bmatrix} \frac{39.7}{6.5s+1}e^{-0.32s} & \frac{242}{1.14s+1}e^{-1.58s} & \frac{7.15}{5.7s+1}e^{-0.12s} & \frac{2.59}{1.36s+1}e^{-1.03s} \\ \frac{516}{3.3s+1}e^{-0.11s} & \frac{8.2}{3.4s+1}e^{-0.59s} & \frac{608}{2.4s+1}e^{-1.33s} & \frac{7.2}{2.9s+1}e^{-1.46s} \\ \frac{7.05}{1.5s+1}e^{-1.53s} & \frac{1.1}{4.3s+1}e^{-1.73s} & \frac{1}{7s+1}e^{-0.15s} & \frac{41.2}{1.3s+1}e^{-0.57s} \\ \frac{38.3}{7.6s+1}e^{-1.72s} & \frac{172}{7.6s+1}e^{-1.46s} & \frac{26}{3.1s+1}e^{-0.19s} & \frac{10.3}{1.9s+1}e^{-1.97s} \end{bmatrix}$$

Now, if we use the HIIA to generate an IM, we get the following IM

$$\Gamma = \begin{bmatrix} 0.024 & \mathbf{0.1481} & 0.0042 & 0.0011 \\ 0.28 & 0.0052 & \mathbf{0.3458} & 0.0044 \\ \mathbf{0.0043} & 0.0007 & 0.0006 & 0.025 \\ 0.0245 & 0.1064 & 0.0145 & \mathbf{0.0079} \end{bmatrix}$$

Table 5.1: The recommended control configurations with different scaling methods.

	No scaling	Row scaling	Column Scaling	SK-scaling
y_1	u_2	u_2	u_2	u_1
y_2	u_3	u_1	u_1	u_3
y_3	u_1	u_4	u_4	u_4
y_4	u_4	u_3	u_3	u_2

Table 5.2: The cost for the different control configurations.

No scaling	Row scaling	Column Scaling	SK-scaling
1.34×10^7	4631	4631	11468

In bold are the values that correspond to the largest sum of interactions with a positive Niederlinski Index. From this we can see that the HIIA interaction matrix recommends the pairings $y_1 - u_2$, $y_2 - u_3$, $y_3 - u_1$ and $y_4 - u_4$. If we rescale the IM using the methods listed in Chapter 3 we get the pairings described in Table 5.1.

As can be seen from Table 5.1, we get three different recommend pairings depending on what scaling we use. For each configurations, we design a decentralized control scheme using the lambda method for varying values of λ. The resulting feedback systems are then tested for a reference step. For the comparison we define a cost using (4.1) with a simulation time of 2000 time units after the reference step. For each configuration the cost is calculated for values of λ ranging from $0.1T$ to $10T$ (with T being the time constant of the SISO-system) and the lowest cost was then saved. Table 5.2 shows the results for each configuration.

From this it can be seen that the pairing recommended by both row and column scaling yields the best result, with the pairing recommended by the unscaled IM being very poor.

If we wish to further investigate this we can use the MIMO generator to generate 100 similar first order plus dead time systems (with the settings described in Table 5.3) and repeat the same investigation. To compare the results on different systems we normalise the costs using Equation (4.2) and the resulting average score for each scaling method is listed in Table 5.4. From here we can see that SK scaling generally yields the best result for reference following for these types of first order plus dead time systems. In Paper A

and Paper B we will utilize this method to evaluate many different types of methods and systems, including sparse control structures.

Table 5.3: Table showing the MIMO model generator settings

Parameter	Value
Size	
Number of inputs	4
Number of outputs	4
Minimum number of inputs affecting each output	4
Maximum number of inputs affecting each output	4
Dynamics	
Maximum static gain	1000
Minimum pole time constant	1
Maximum pole time constant	10
Percentage of transfer functions with delay	100
Minimum Delay	0
Maximum Delay	2

So from this we can see while for a single system there is no guarantee that SK scaling gives the best result, but on average it seems to outperform the other scaling methods for these types of systems. Furthermore none of the configuration found with SK scaling were unstable, demonstrating its consistency. Moreover we can note that the unscaled IMs yield comparably poor results, which illustrates the importance of scaling the IMs.

Table 5.4: The score and number of unstable configurations (U) for the different scaling methods tested on 100 systems.

No scaling		Row scaling		Column scaling		SK-scaling	
Score	U	Score	U	Score	U	Score	U
59.3	25	91.1	5	90.4	6	97.2	0

CHAPTER 6

LQG control

In chapters 2-5 we examined methods of decentralised or sparse control of MIMO systems, in which the MIMO control problem is divided into a number of SISO control problems. However while these methods are sufficient for many industrial practices, the fact that the entire system is not taken into account can lead to sub-optimal control behaviour. An alternative to these methods is full MIMO control in which the entire system is treated as a single control problem. The perhaps most common such method is LQG control, which is based on control of linear systems, for sampled systems defined as

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k + w_k \\ y_{k+1} &= Cx_k + Du_k + e_k, \end{aligned} \tag{6.1}$$

were w_k and e_k are Gaussian model noise and measurement noise, respectively. LQG control yields a full multi-variable controller where all the inputs are used to control all the states. It is based on finding the control input that minimizes a cost function similar to that described in Section 4.2, namely

$$J_N = \min_u E\left[\sum_{i=0}^{N} u_i^T Q u_i + \sum_{i=0}^{N} x_i^T R x_i + x_{N+1}^T S_{N+1} x_{N+1}\right], \tag{6.2}$$

where Q is a positive definite symmetric matrix and R and S_{N+1} are positive semi-definite symmetric matrices.

To derive the optimal solution dynamic programming is used. This means that first the u_N that minimizes the cost function (6.2) is found, then for the remaining cost a control signal u_{N-1} that minimizes this cost is found. After this, u_{N-2} is found to minimize the now remaining cost. This is repeated until all u_k have been found. It can be shown that the optimal control signal is [44]

$$u_i = -L_i x_i,$$

where

$$\begin{aligned} L_i &= (B^{\mathrm{T}} S_{i+1} B + Q)^{-1} B^{\mathrm{T}} S_{i+1} A \\ S_i &= A^{\mathrm{T}} \left(S_{i+1} - S_{i+1} B \left(B^{\mathrm{T}} S_{i+1} B + Q \right)^{-1} B_i^{\mathrm{T}} S_{i+1} \right) A + R. \end{aligned}$$

In this case, where N is a finite number, this is the solution to what is called the finite horizon problem. If $N \to \infty$ this becomes what is known as the infinite horizon or stationary problem, which has the solution

$$u_i = -L x_i$$

$$L = (B^{\mathrm{T}} S B + Q)^{-1} B^{\mathrm{T}} S A,$$

where S is found by solving

$$R + A^T S A - S - A^T S B (Q + B^T S B)^{-1} B^T S A = 0.$$

It can also be shown that for sufficiently large N the finite horizon solution L_i will tend towards the infinite horizon solution L.

6.1 State observers

LQG control is a state based control scheme, that is to say the control input is calculated based on the states of the system. However, this assumes that the state is known. This is not always the case as many of the states may not be measured, and the measurements which are available are subject to measurement noise. This creates the need to estimate the states using what is called an observer, which combines measurements and model based estimates to derive an estimate of the states. The most common observers using state space models can be written on the so-called innovation form [45]

$$\begin{aligned} \hat{x}_{k+1|k} &= A\hat{x}_{k|k} + Bu_k \\ \hat{x}_{k|k} &= \hat{x}_{k|k-1} + K_k(y_k - C\hat{x}_{k|k-1}) \end{aligned}$$

where $\hat{x}_{k+1|k}$ is the estimate of $\hat{x}_{k+1}$ at time k. K_k is the observer constant gain specified by the user. To calculate the K_k that minimizes the variance of the estimate, a Kalman filter can be used, which is given by:

$$\begin{aligned} \hat{x}_{k|k} &= \hat{x}_{k|k-1} + K_k(y_k - C\hat{x}_{k|k-1}) \\ \hat{x}_{k+1|k} &= A\hat{x}_{k|k} + Bu_k \\ K_k &= P_{k|k-1}C^T(CP_{k|k-1}C^T + R_e)^{-1} \\ P_{k|k} &= P_{k|k-1} - K_kCP_{k|k-1} \\ P_{k+1|k} &= AP_{k|k-1}A^T + R_w - AK_k(CP_{k|k-1}C^T + R_e)K_k^TA^T \end{aligned}$$

where R_1 is the variance of the model noise w_k and R_2 is the variance of the measurement noise e_k (assumed independent here). If measurements y_k are unavailable due to packet drops or delays, only the model can be used for prediction, which is equivalent of letting $K_k = 0$ in the Kalman filter.

In state feedback LQG control these estimated states are used to derive the control signal. The separation principle is important to note here, which states that the estimation problem and the control problem can be solved as two independent problems. [44].

CHAPTER 7

Control over unreliable channels

To implement the control structures described previously, there needs to be controllers that calculate the control signal, actuators that apply the control signal on the process, and sensors that measure the outputs used by the controllers to calculate the control signals. As controllers, sensors and actuators are often positioned on different locations, it can be difficult or expensive to create reliable communication links between the components. Hence wireless communication channels, which are generally less reliable have seen increasing use. Therefore, the question of control over lossy networks is one of increasing importance. Some methods to optimize control algorithms over lossy channels are discussed in [46]–[48].

Depending on how data is sent and decoded, a communication channel could also be subject to delays. For instance, if tree codes are used to characterize the submission of the data from the controller to the actuators, and from the sensors to the controller, as discussed in [49] and [50], a lossy channel can be turned into a channel with a random delay. This delay is not bounded, but it follows a probability function that depends on the reliability of the channel.

7.1 Unreliable communication links

When there is a risk of packet loss or packet delays in a system there are no guarantees that the signal sent from the controller at time t is applied as this time. Hence, there is a need to distinguish between the signal the actuator applies and the signal the controller sends. The system studied here can be described as

$$x_{k+1} = Ax_k + Bu_k + w_k, \tag{7.1}$$

where u_k is the control signal applied by the actuator at time k. However, u_k is not necessarily the control signal our controller sends at time k, which is denoted as v_k. Consequently, if LQG control was to be implemented in cases with risk of packet loss or packet delays, the optimization problem would be posed as

$$J_N = \min_v E\left[\sum_{i=0}^{N} u_i^T Q u_i + \sum_{i=0}^{N} x_i^T R x_i + x_{N+1}^T S_{N+1} x_{N+1}\right]. \tag{7.2}$$

rather then the general case without packet delays or losses described in (6.2).

7.2 Hold input or zero input

When there is an unreliable channel between the controller and the actuator this means that the latest signal sent might not yet have arrived. When this occurs there are two basic strategies the actuator can adopt [51], [52]. One is to set the input to zero if the latest control signal sent is delayed or lost, which is known as zero-input. The other alternative is to continue to apply the latest input received until a more recent one arrives, which is known as hold-input.

7.3 TCP- or UDP- like case

There are also two basic types of unreliable communication links. In one the sender does not know if the sent packet has arrived, which is known as the UDP-like case. In the other one there is a system of acknowledgment that ensures that the sender knows if the sent packet has arrived. This is referred to

as the TCP-like case. While these acknowledgments can themselves be subject to packet losses and packet delays, we will be assuming that their arrival is instantaneous and reliable. There are two principal differences between the design of a controller for the UDP-like case and for the TCP-like case.

The first pertains to state estimation. If there is a random delay between the sensors and the controller, this means that at time n the controller may not have access to measurements from time n. An algorithm is then needed to estimate the states (as they may not have arrived). For example, in the case where the states can be measured directly without noise, an estimate of the state x_N at time t would be

$$\hat{x}_N(t) = \begin{cases} x_N & \text{if } x_N \text{ has arrived by time } t \\ A\hat{x}_{N-1}(t) + B\hat{u}_{N-1} & \text{if } x_N \text{ has not arrived by time } t \end{cases}$$

where $\hat{u}_N$ is an estimate of the control signal applied by the actuator at time N. In the TCP-like case the control signal applied by the actuator is known as there are acknowledgments that inform the controller when a control signals arrives at an actuator. This means that $\hat{u}_{N-1} = u_{N-1}$ and thus no uncertainty of the control signal will impact the quality of the estimation. However, for the UDP-like case $\hat{u}$ must be estimated, which leads to a more complex problem as the choice of control signal will impact the optimal estimation [46] and therefore one cannot necessarily treat the control problem as a problem separate from the estimation problem. However, the focus here is the TCP-like case, where the acknowledgments lead to the separation principle holding and thus the control problem can be solved separately from the estimation problem [46].

Another important difference is that the information of which control signal that is available in the TCP-like case is highly relevant when deriving subsequent control signals. Thus, controller design methods which utilize this knowledge can perform significantly better then those that do not, as is shown in Paper E.

7.4 Control structure for LQG control over unreliable channels

For LQG control without delays, the control signal is derived as discussed in Chapter 6, by $v_k = -Lx_k$, i.e. the control signal is proportional to the state. Since this is the optimal solution, the state can be seen to contain all relevant information required to derive the control signal. However in the case of delays with hold input, the optimal control solution will have a somewhat different structure. To understand this we examine an example where at time k,

$$u_k = v_{\tau_k},$$

In this case, as v_{τ_k} is applied by the actuator, we can understand that v_{τ_k+1},..., v_{k-1}, v_k have not yet arrived to the actuator. Now, when deriving the optimal v_{k+1}, the optimal solution will not only depend on x_{k+1}, but also on control signals that may be applied by the actuator before v_{k+1} has reached the actuator. In Paper E we show that for the finite horizon case the optimal solution for v_{k+1} becomes

$$v_{k+1} = L_k \begin{bmatrix} v_k \\ \vdots \\ v_{\tau_k} \\ x_{k+1} \end{bmatrix}, \tag{7.3}$$

where L_k can be derived using knowledge of the channel (probability of packet delays and packet losses). Now this assumes that the controller knows which signals have reached the actuator, which is the case in the TCP- like case and not in the UDP- like case. If the controller does not know which signals have reached the actuator, it will have to take into account all previous control signals, i.e. the solution would be in the form of

$$v_{k+1} = L_k \begin{bmatrix} v_k \\ \vdots \\ v_0 \\ x_{k+1} \end{bmatrix}. \tag{7.4}$$

In Paper D we derive an expression in the finite horizon case for the optimal

L_k in case of a specific type of random delay. This solution did not however fully utilize the knowledge of which control signals has reached the actuator, resulting in a control strategy as described in (7.4). In Paper E we correct this short-coming as well as expand the solution to cover a broader set of unreliable channels, encompassing both packet losses and packet delays.

CHAPTER 8

Mechanical pulp refining

In order to implement the control strategies described previously on a existing system, an understanding of the system in question is required. In this thesis the pulp refining process is analyzed, being a potential candidate for decentralized or sparse control.

The purpose of pulp refining is to break down chips of wood into their individual fibers, the resultant pulp can then be used to produce paper, cardboard and other products. This is done using paper refiners, which utilize mechanical forces to grind down the wood chips. Before and during the refining process the chips are treated to soften up. This as otherwise the chips will be too brittle, causing fractures which result in shorter fibres. Pretreatment can be done by utilizing steam to heat the chips (this is referred to as thermomechanical pulping (TMP). If chemicals are used in conjunction with the steam to treat the chips, this is referred to as chemical thermomechanical pulping (CTMP). In this thesis we will examine data of both the TMP and CTMP refining process.

There are numerous different types of pulp refiners. The refiner type that is the focus for the modelling work presented in the thesis is the CD refiner. CD refiners consists of two serially linked refining zones called the flat zone

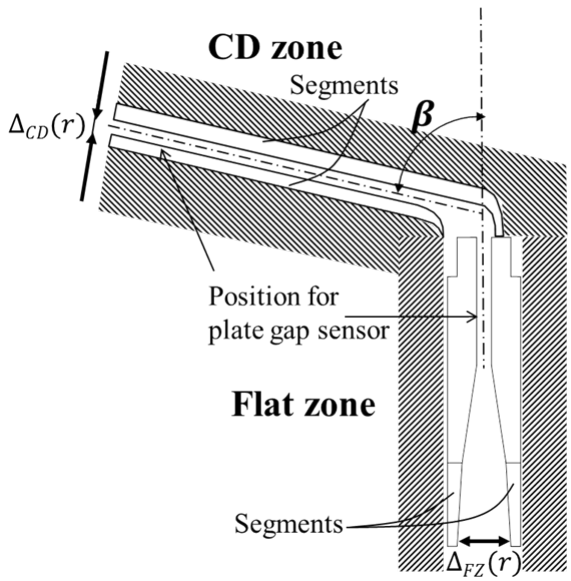

Figure 8.1: A schematic drawing of a CD refiner. The vertical flat zone (FZ) is directly linked to the conical zone (CD) via an expanding point. Plate gaps are represented by Δ_{CD} and Δ_{FZ}.

(FZ) and the conical zone (CD) as shown in Figure 8.1. The width of each zone, commonly known as the plate gap, is adjustable. Furthermore, dilution water can be added at the entrance to each zone.

After the refining process the resulting pulp is sent to a latency chest, in which the pulp is stirred and heated in order to straighten the fibers. Measurements of pulp properties are often taken after the latency chest. This results in a substantial and time varying measurement delay as the time spent in the latency chest is generally around 5-20 minutes depending on the size of the latency chest. Moreover the mixing in the latency chest means that the measurements should be interpreted as weighted averages over a time period.

8.1 Inputs

These are different factors which may affect the quality of the pulp and paper produced by the refining process and, thus, can potentially be used both for control of pulp properties and to model pulp properties. In the following sections, the main inputs for a refiner are described. The effect these inputs have on pulp and paper properties will later be examined, which can lay the foundation for design of control configurations.

Plate gaps

Plate gap, also referred to as disc clearance, is a measure of the width of the space where the chips are ground. Hence it can be seen as a measure to how closely the chips are ground.

Residence Time

Residence time is a measure of how much time the chips spend in each zone of the refiner. It can be seen as a measure of the time the refiner spends grinding the chips. Residence times are heavily influenced by both the plate gaps and the production rate. As the production influences the plate gaps in both zones, the plate gap of each zone is heavily interdependent.

Production

Production is a measure of how much pulp is produced. In practice, though, the production is estimated from the volume of input chips. As the density of input chips varies this can lead to production measurements being somewhat unreliable.

Specific energy

Specific energy is calculated as energy consumption divided by production. Hence, it can be seen as a metric of efficiency as it measures the amount of energy consumed per unit of product.

Dilution water feed rate

This is the amount of dilution water added to during the refining process. Dilution water can be added separately to each zone. Dilution water greatly affects the consistency of the flow.

Consistencies

The consistency is a measure of the percentage of chip content in the different zones. Note that the consistencies are not only affected by the dilution water feed rate, but also by the flow of chips, water and steam within the refiners.

Types of wood chips

Due to variation in production and availability of wood chips, pulp and paper refineries often complement their own produced wood chips with chips from external sources. These are often of different type and quality, which may effect the properties of the produced pulp or paper.

Internal and external variables

Some of the aforementioned inputs can be measured or concluded directly from operating conditions. These are what we refer to as external variables, and include the dilution water feed rate, saw mill chip type and plate gaps. We also consider the production to be an external variable, as its derivation is straight-forward from process inputs.

By adding sensor arrays within the refiner temperature profiles can be acquired over different parts of the refiner. These temperature measurements, along with the previously mentioned external variables, can be used in the extended entropy model [23]–[26], for estimation of the consistency profile and the fiber residence times in the FZ and CD zones [14]–[16]. These inputs, being estimated from internal temperature measurements, will be called internal variables, and they will be used along with external variables to derive models for pulp and paper properties in Paper G and Paper H as illustrated in Figure 8.2.

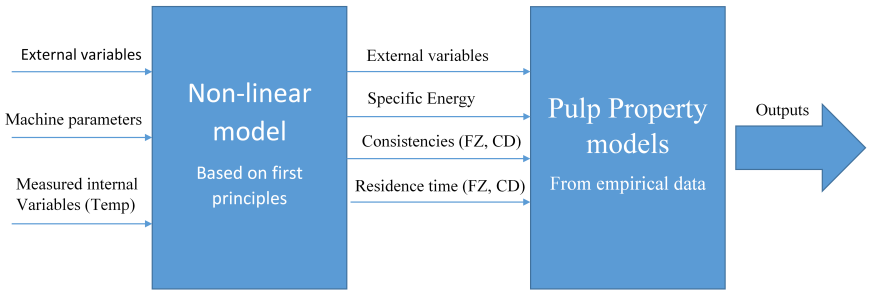

Figure 8.2: A combination of internal and external variables can be used to derive pulp refiner models for predicting pulp properties.

Interdependencies of internal variables

From the internal measurements we have access to estimates of the residence time and consistencies in both zones. There are many factors that affect these parameters. Some parameters, such as production and chip content, will affect these measurements in both zones. Plate gaps and dilution water feed rates are individually set for each zone, but due to the serial nature of the refining process, parameter values of the flat zone may affect the measurements in the conical zone, as illustrated in Figure 8.3. This leads to considerable interdependencies between the measurements in the different zone. This is especially noteworthy for the residence times as production heavily influences the residence times in both zones, while in normal operating conditions the affect of the separate plate gaps is considerably lower. These interdependencies make it difficult to isolate the impact of each of these variables in their separate zones.

8.2 Outputs

There are many potential measures of pulp and paper quality, which are interesting to model (estimate) and to control. The focus here will be on the paper quality measure *tensile index*, along with the pulp quality measures *fiber length*, *freeness* and *shives width*.

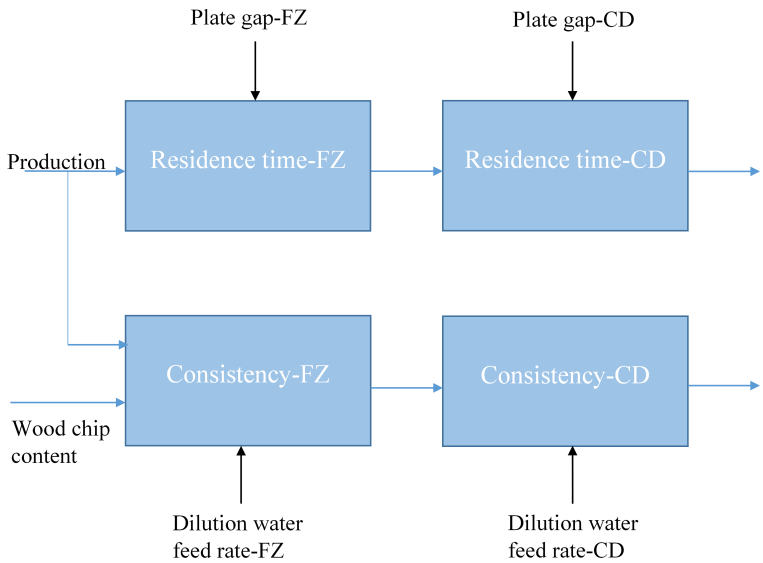

Figure 8.3: The serial nature of the process causes significant inderdependencies of the internal variables.

Fiber length

This is a measure of the average length of fibers in the pulp. Pulp with longer fibers generally yields paper with higher tensile strength, so this is an important quality measure. It can be measured automatically, for instance by a Pulp Quality Monitor (PQM).

Shives Width

Shives are bundles of unrefined fibers left in the pulp. These cause points of structural weakness in the resultant paper. Shives width is often measured by counting the number of shives in a sample over a certain width (PQM for example take widths greater than 150 micrometers).

Freeness

This is a measure of how quickly water drains from the pulp, which is commonly used as a measure of pulp quality. It can be measured automatically

using a PQM.

Tensile Index

Tensile index is a measure of the tensile strength of a handsheet. Tensile strength is tested by physically pulling a handsheet apart, to see the forces required for it to break.

To measure tensile index pulp is collected from the process, which is then used to make handsheets. These handsheets are then used to calculate tensile strength and from this derive tensile index. To understand how this is done we examine a case where the collected pulp is used to make 7 handsheets. Each handsheet is divided into 3 strip's (see Figure 8.4). Each strips tensile strength is then tested. From this there are 3 ways to calculate the tensile index of each strip:

$$\tau_{ij} = \begin{cases} \sigma_{ij}/\bar{\mu} & \text{Case A} \\ \frac{\sigma_{ij}}{\frac{1}{l}\sum_{k=1}^{l}\mu_{kj}} & \text{Case B} \\ \sigma_{ij}/\mu_{ij} & \text{Case C} \end{cases}$$

where σ_{ij} is the tensile strength of strip i from handsheet j, μ_{ij} is the corresponding basis weight and l is the number of strips. The denominator in Case A is the average basis weight for all handsheets, i.e, one measure for the complete batch of handsheets. The denominator in case B can be seen as the most logical, being the average basis weight of each handsheet. Case C requires information about both tensile strength and basis weight for each strip. In general, Case C is too time consuming and not so often used as a standard procedure. Instead, Case A is normally used, which of course can affect the accuracy when it comes to handsheet variations in the forming procedure.

There is generally a considerable variation of the tensile index of the strips from the same pulp sample [53]. The main reasons are that some strips will contain shives and there is a variation between the basis weight of the handsheets. Therefore, to get a single tensile index measurement from the pulp sample, the mean of all measured tensile indices from the sample is used.

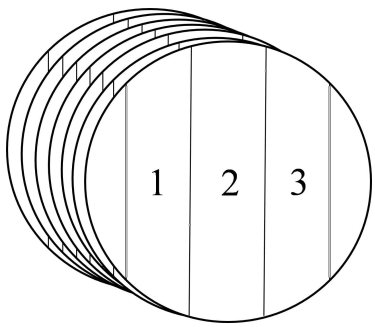

Figure 8.4: A schematic drawing of three strips obtained from each of the seven handsheets.

CHAPTER 9

Modelling

In mathematical modelling the goal is generally to derive a mathematical relationship from which one or several variables (referred to as the outputs or dependant variables), can be predicted using the remaining variables (referred to as independent variables or inputs). The structure of this mathematical relationship can vary from very simple linear models, to highly complex non-linear models. Choosing the correct model structure is a challenge and much depends on the expected relationship between the inputs and outputs, as well as the availability of data. With limited or noisy data, complex mathematical models may become highly unreliable, necessitating simpler model structures.

These models are important and in many cases necessary due to their use in predicting the output. Model predictions can complement or replace manual measurements, which are often expensive and time consuming to make, as well as being potentially unreliable. Another use for models, highly relevant in the control field is that of inference. This is that as models reflect the relationship from inputs to outputs they give understanding to how the outputs are affected by the inputs. This is highly useful when designing automatic controllers as it aids in determining which inputs are appropriate to use for control and which inputs do not sufficiently affect the outputs to be usable which can lay

the foundation when solving the input-output pairing problems. Note that the models need to be relatively simple in nature or inference becomes very difficult.

9.1 Static linear models

A commonly used simple model is the static linear model. For such models the equation used to predict the output is in the form of

$$y = \theta x,$$

where $y \in R^{1\times 1}$ is the output, $x \in R^{p\times 1}$ is a vector of the predictors (for static linear models this is generally the inputs, but more complex models may have other terms as predictors) , $\theta \in R^{1\times p}$ is a vector of what is called the parameters, and p is the number of predictors. The goal in linear modeling is to derive values for θ, so that from measurements of x, one gets as accurate as possible estimates of y.

The process of deriving linear models generally first consists of acquiring a data set containing measurements of both the output, and the predictors. These measurements can be grouped in matrices so a data set of size N can be expressed as

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \qquad X = \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_N^T \end{bmatrix},$$

where y_1 corresponds to the first measurement of the output, and x_1 are the measurements of the predictors at this time. The goal is then to derive parameters θ which best describe the relationship between Y and X. This relationship can be assessed by examining the mean square error of the difference between our predicted and measured outputs, i.e,

$$\frac{1}{N}\sum_{i=1}^{N}(y_i - \theta x_i)^2 = \frac{1}{N}(Y^T - \theta X^T)(Y - X\theta^T) \tag{9.1}$$

The θ which minimizes the above expression is:

$$\theta = (X^T X)^{-1} X^T Y$$

provided $X^T X$ is invertible.

Dynamic models

An alternative to the described linear static models is dynamic models, which are continuously updated as more data is acquired. This presents several advantages. As new data measurements are often continuously provided, by continuously updating our model this ensures that we are at all times utilizing all available data. Another benefit is that these dynamic algorithms can be adapted so that old data is given less importance or discarded entirely. This allows the models to adapt to changing process conditions. There are a few different variants of adaptive models, the one mainly used here is the sliding window adaptive model [54] in which data is discarded after a certain time, so the models are based only on the most recent data. This can be implemented using the following equations:

$$\begin{aligned}
\theta_k &= \theta_{k-1} + K_k \begin{bmatrix} \epsilon(k, \theta_{k-1}) \\ -\epsilon(k-m, \theta_{k-1}) \end{bmatrix} \\
K_k &= P_{k-1} \begin{bmatrix} x_k^T & x_{k-m}^T \end{bmatrix} \left(I + \begin{bmatrix} x_k^T \\ x_{k-m}^T \end{bmatrix} P_{k-1} \begin{bmatrix} x_k^T & x_{k-m}^T \end{bmatrix} \right)^{-1} \\
P_k &= P_{k-1} - K_k \begin{bmatrix} x_k^T \\ x_{k-m}^T \end{bmatrix} P_{k-1}
\end{aligned}$$

where $\epsilon(k, \theta) = y_k - \theta x_k$ and m is the length of the sliding window (measurements older than m time units will be disregarded with this algorithm).

ARX models

A group of models somewhat more complex than the static linear model is the ARMAX group of models. These models also model dynamic behaviour by including past values of inputs and outputs. One of the most common model structure from this group is the ARX model,

$$A(q)y(k) = B_1(q)u_1(k) + B_2(q)u_2(k) + ... + B_p(q)u_p(k) + e(k),$$

where $y(k)$ is the output at time k, $e(k)$ is white noise,

$$A(q) = 1 + a_1q^{-1} + a_2q^{-2}... + a_nq^{-n}$$
$$B_j = b_1 + b_2q^{-1} + b_{m+1}q^{-m},$$

where q is the shift operator (i.e. $q^{-1}y(k) = y(k-1)$) and n and m are parameters used to determine the size of the model. By setting $q = 1$ in the above equation the DC-gain can be calculated, which is the ARX models gain at low frequencies, and is comparable with the parameter vector from the static linear model.

For most cases ARX models can be derived in the same way as static linear regression models, by simply expanding the predictor matrix X to also contain past inputs and outputs. However, for many of the systems we examine here this is not possible due to the fact that the inputs are sampled much more frequently than the output. This means that relevant past outputs are not always available. This can be solved as described in [55] by designing the model recursively and estimating missing outputs using the model, i.e.

$$\begin{aligned}
\hat{\psi}_k &= \begin{bmatrix} -\hat{y}_{k-1} & ... & -\hat{y}_{k-n} & u_{k-1} & ... & u_{k-m} \end{bmatrix}^T, \\
\gamma_k &= \frac{P_{k-1}\hat{\psi}_k}{\lambda + \hat{\psi}_k^T P_{k-1}\hat{\psi}_k}, \\
\theta_k &= \theta_{k-1} + \gamma_k(y_k - \hat{\psi}_k^T\theta_{k-1})r_k, \\
P_k &= \frac{1}{\lambda}(I - \gamma_k\hat{\psi}_k^T)P_{k-1}r_k + P_{k-1}(1 - r_k)
\end{aligned}$$

where $r_k = 1$ if y_k is available and $r_k = 0$ otherwise. $\hat{y}_k$ is our output estimate, which when measurements are unavailable can be estimated from our model,

$$\hat{y}_k = \theta_k^T\hat{\psi}_k.$$

Evaluating models

A common challenge when conducting linear modeling is deciding which parameters to use. As some parameters effect on y might be small or highly nonlinear, including these parameters may not improve our linear model quality. To be able to determine which parameters are beneficial a method to evaluate model quality is required. One way is to use the cost (9.1) to evaluate the model. This is a metric useful for comparing different models. However, we also need a measure of how well the model corresponds to the data. To resolve this there is what is known as the coefficient of determination, which compares the linear model to the best placed horizontal line, and it is expressed as

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - f(x_i))^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2},$$

where $\bar{y}$ denotes the mean of the outputs and $f(x_i)$ denotes the output predicted by the model for the inputs x_i. The R^2 has the benefit of giving a clear scale of model quality. An R^2 of 1 indicates a perfect fit, an R^2 of zero indicates that the model is no better than a horizontal line, while negative R^2 indicates an even worse model. A problem with the R^2 metric is that increasing the number (p) of parameters will always improve the R^2. To solve this the adjusted R^2 can be used [56],

$$adj.R^2 = 1 - \frac{N-1}{N-p-1}\frac{\sum_{i=1}^{N}(y_i - f(x_i))^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2},$$

which punishes models with many parameters and hence ensures that adding additional parameters will only improve the $adj.R^2$ if they have a greater impact than what is expected by chance.

It is important to note that no metric of model quality is perfect (R^2 for example has the limitation that it emphasises predictions of the major and enduring shifts in output values). Hence, it is often important to combine the metrics of model quality with visual inspection of the models predictions to analyse each model strengths and limitations.

When evaluating models, especially if the purpose of the models is for prediction, it is important to ensure that the model works well on new data and not only on the data used to make the model. To ensure this before designing the model, the data set is divided into two sets, one training and one valida-

tion set. The training set is used to derive the model while the validation set is then used to evaluate the models, which can be done using either the R^2 or by using the quadratic cost described in (9.1).

A common issue in linear modeling is that the inputs may be dependent on each other. This may mean that it is difficult to individually asses the impact of each input, resulting in less reliable models. A measure of the linear interdependency of the predictors is the variance inflation factor (VIF) [57], defined as

$$VIF_i = \frac{1}{1 - R_i^2},$$

where R_i^2 is the R^2 value obtained by regressing the ith predictor on the remaining predictors. A $VIF_i = 1$ means that there is no linear correlation between the kth predictor and the other remaining predictor variables. If $VIF > 4$, a general rule is that further analysis should be performed, while $VIF > 10$ indicates serious multicollinearities and a need to find a modified set of predictors.

Outlier detection

A common issue when modeling is the possibility of outliers. These are data points whose value differs significantly from that predicted by the model. This could be due to a number of factors such as measurement errors, sampling issues or anomalous process conditions. Outliers can have a considerably detrimental impact on model quality, especially in cases where the number of data points is low. There are a number of methods designed to detect outliers. One is Cook's distance [58] which gauges the impact of removing a data point i by:

$$D_i = \frac{\sum_{k=1}^{N} (\hat{y}_k - \hat{y}_{k(i)})^2}{ps^2},$$

where $\hat{y}_k$ is the prediction of the data at time k with a model derived using the full data set, $\hat{y}_{k(i)}$ is the prediction of the data at time k with a model derived using the same data set but with the data point i removed. s^2 is the mean squared error, defined as:

$$s^2 = \frac{\sum_{i=1}^{N}(y_i - \hat{y}_k)^2}{(N-p)}$$

A high Cook's distance indicates that removing the data point considerably impacted the model, which in turn indicates that the point may be an outlier.

An issue with outlier detection methods is the possibility of masking, in which outliers can hide each other. When modeling using small data sets consisting of quite unreliable data we propose a method for analysis and validation of models that is specifically tailored to cases with small data sets: use roughly half the samples to design a model (e.g. 10 samples for a set of 19). Evaluate the resulting model by calculating the adjusted R^2. Then repeat this for all possible combinations of samples. This allows us not only to get a general impression of model quality but also to evaluate the quality of each measurement sample. For example, if a sample only appears in poor models, it is indicative that the sample itself is an outlier.

CHAPTER 10

Summary of included papers

This chapter provides a summary of the included papers.

10.1 Paper A

Fredrik Bengtsson, Torsten Wik, Elin Svensson
Resolving issues of scaling for gramian based input-output pairing methods
International Journal of Control, Taylor & Francis, 2020 .

In this article we examine ways to resolve the issues of input and output scaling that the gramian based input output pairing methods have. We propose a new method based on the Sinkhorn-Knopp algorithm which removes the scaling dependence of the gramian based measures. We then test this along with other scaling methods on a large number of systems using the MIMO generator described in Paper C. It is found that using the Sinkhorn-Knopp algorithm to scale the systems yields significant improvements compared to the other scaling methods.

10.2 Paper B

Fredrik Bengtsson, Torsten Wik
Finding feedforward configurations using gramian based interaction measures
Submitted to Modeling, Identification and Control .

In this article we expand on the work carried out in Paper A, examining how the scaling schemes can be modified to design feed-forward controllers. The different methods are evaluated using the MIMO generator described in Paper C, and it is found that a hybrid scaling scheme, where Sinkhorn-Knopp scaling is used to design a decentralized control scheme and then other scaling schemes are used to determine which elements to use for feedforward yields the best result.

10.3 Paper C

Fredrik Bengtsson, Torsten Wik
A multiple input, multiple output model generator.
A multiple input, multiple output model generator. Technical report, Department of Signals and Systems, Chalmers University of Technology, 2017 .

In this technical report a MIMO model generator is presented. This generator allows the user to generate a large number of linear MIMO systems with predefined properties, such as system size, stability, time constants, delays etc. These systems can then be used to evaluate various control tuning methods.

10.4 Paper D

Fredrik Bengtsson, Babak Hassibi, Torsten Wik
LQG control for systems with random unbounded communication delay
Proceeding of the 55th Conference on Decision and Control (CDC), pp 1048-1055. 2016 .

Here we examine LQG control when there is a random unbounded delay between the controller and the actuator. For a specific probability distribution

of the delay we derive an optimal controller for the TCP-like case and evaluate it in simulations of an example system. The acknowledgements of which signals has arrived to the actuator are here only used for the state estimation and not in the optimization of the controller.

10.5 Paper E

Fredrik Bengtsson, Torsten Wik
Stochastic optimal control over unreliable communication links
To be submitted .

This article expands on the work in Paper D, to present a solution for a more general type of unreliable communications with any type of random delay, as well as packet losses. Moreover, we fully utilize the knowledge of which control signals have arrived to derive a solution that yields a lower cost than the one in Paper D.

10.6 Paper F

Fredrik Bengtsson, Anders Karlström, Torsten Wik
Modeling of Tensile Index using Uncertain Data Sets
Nordic Pulp and Paper Research Journal,
vol. 35, no. 2, pp. 231-242, 2020 .

In this article we examine data from a paper refiner and examine the modeling of tensile index from a number of available inputs and outputs. We explore to what extent various inputs are useful for predicting tensile index as well as examine how to best conduct modeling of tensile index considering the limited availability of data.

10.7 Paper G

Fredrik Bengtsson, Anders Karlström, Torsten Wik
On the modeling of pulp properties in CTMP processes
Submitted to Nordic Pulp and Paper Research Journal .

In this article we examine data from the CTMP refining process and examine the modeling of fiber length, shives width and freeness from a number of available inputs and outputs. Both static and dynamic models are explored, and it is shown that reasonable models can be derived for fiber length and shives width.

CHAPTER 11

Concluding remarks and future work

In this thesis control of MIMO systems has been investigated. Firstly, we examined methods for decentralized and sparse control and proposed modifications to the gramian based interaction measures for the input-output pairing problem. This modification consist of rescaling the IMs from the gramian based interaction methods using the Sinkhorn-Knopp algorithm. Furthermore we propose a new method of using the gramian based measures to design sparse controllers, by better taking into account the interactions of the system.

To evaluate control configuration selection methods we proposed a MIMO system generator that allows for a statistical evaluation of different control configuration methods. By evaluating our proposed modifications, both with the MIMO generator and on models of real systems, we demonstrate that our modifications significantly improve the results, both for decentralised and sparse control structures. In addition, using the Sinkhorn-Knopp algorithm to rescale the IM has the advantage that it yields identical results regardless of the scaling of the original system.

Moreover we investigate another MIMO control method, LQG control, in the case where there are unreliable communication channels between the con-

troller and actuators and between the sensors and controller. For this case we derive an optimal finite horizon control LQG control law.

Finally, the pulp and paper refining process was examined. We derived models for predicting tensile index, fiber length and shives width using a combination of external and internal variables.

There are multiple areas in this thesis work which can be expanded on. When it comes to evaluation of methods, further work is needed in the area of automatic controller tuning. The automatic controller tuning should ensure that comparison of methods for control configuration selection is as little dependent on the tuning of individual control loops as possible. However, it is not always clear what the best method to use is, and if each controller should be found with the other outputs in open loop [59], or if one should successively close the loops with each implemented controller [60].

Using a MIMO generator to evaluate different control structure selection methods proved to be effective. As the MIMO generator allows for generation of systems with many different types of properties, it may be fruitful to expand the evaluations with other types of systems to analyse which methods should be used for which type of system.

In the area of LQG control for lossy channels there are still some expansions possible. Our solution, while optimal is computationally very demanding. Although we have proposed a few methods to reduce the online computational burden, this is an area which can be explored further.

Moreover, we have only presented work for the TCP-like case. The UDP-like case, where there are no acknowledgments between the sensor and actuator, is still an open problem.

Regarding the modeling of pulp and paper outputs there is considerably more to do. There are numerous other pulp and paper properties which it would be of interest to model. Furthermore, the modeling was done on a data set of quite limited size. Studies conducted on a more expansive data set, where an effort has been made to individually excite the inputs, would potentially allow for more detailed and accurate modeling of pulp and paper properties. A larger data set would also allow for more possibilities to explore more complex model structures.

There are multiple potential uses for our models of the pulp and paper properties. One important step would be to utilize the derived models to aid in the design of controllers for the refining processes. At the simplest,

our models can be seen to be a gauge of how much each input affects the outputs. This can be utilized to perform control configuration selection by finding which inputs are appropriate to be used for control. Furthermore with more frequent output measurements, dynamic models could be derived, which could be utilized to derive MIMO controllers, such as LQG control, or to improve a decentralized control structure by using more sophisticated control configuration selection methods such as the gramian based methods.

Finally, the modelling work can be expanded to create digital twins, which can be used to aid operators and to optimize the refining process.

References

[1] S. Skogestad and I. Postlethwaite, *Multivariable feedback control: analysis and design*. Wiley New York, 2007, vol. 2.

[2] A. Khaki-Sedigh and B. Moaveni, *Control configuration selection for multivariable plants*. Springer, 2009, vol. 391.

[3] M. van de Wal and B. de Jager, "A review of methods for input/output selection", *Automatica*, vol. 37, no. 4, pp. 487–510, 2001, ISSN: 0005-1098.

[4] E. Bristol, "On a new measure of interaction for multivariable process control", *IEEE Transactions on Automatic Control*, vol. 11, pp. 133–134, 1966.

[5] Z.-X. Zhu, "Variable pairing selection based on individual and overall interaction measures", *Industrial & Engineering Chemistry Research*, vol. 35, no. 11, pp. 4091–4099, 1996.

[6] W. Birk and A. Medvedev, "A note on gramian-based interaction measures", in *Proceedings of European Control Conference (ECC)*, 2003, pp. 2625–2630.

[7] A. Conley and M. E. Salgado, "Gramian based interaction measure", in *Proceedings of the 39th IEEE Conference on Decision and Control*, Citeseer, vol. 5, 2000, pp. 5020–5022.

[8] B. Wittenmark and M. E. Salgado, "Hankel-norm based interaction measure for input-output pairing", in *Proceedings of the 2002 IFAC World Congress*, vol. 139, 2002.

[9] M. E. Salgado and A. Conley, "MIMO interaction measure and controller structure selection", *International Journal of Control*, vol. 77, no. 4, pp. 367–383, 2004.

[10] R. Sinkhorn and P. Knopp, "Concerning nonnegative matrices and doubly stochastic matrices", *Pacific Journal of Mathematics*, vol. 21, no. 2, pp. 343–348, 1967.

[11] A. Johnson, "LQG applications in the process industries", *Chemical engineering science*, vol. 48, no. 16, pp. 2829–2838, 1993.

[12] Y. Tipsuwan and M. Y. Chow, "Control methodologies in networked control systems", *Control engineering practice*, vol. 11, no. 10, pp. 1099–1111, 2003.

[13] R. A. Gupta and M. Y. Chow, "Networked control system: Overview and research trends", *IEEE Transactions on Industrial Electronics*, vol. 57, no. 7, pp. 2527–2535, 2010.

[14] A. Karlström and J. Hill, "CTMP process optimization part I: Internal and external variables impact on refiner conditions", *Nord. Pulp Pap. Res. J.*, vol. 32, no. 1, pp. 35–44, 2017.

[15] ——, "CTMP process optimization part II: Reliability in pulp and handsheet measurements", *Nord. Pulp Pap. Res. J.*, vol. 32, no. 2, pp. 253–265, 2017.

[16] ——, "CTMP process optimization part III: On the predicition of scottbond, z-strength and tensile index", *Nord. Pulp Pap. Res. J.*, vol. 32, no. 2, pp. 266–279, 2017.

[17] B. Strand, "Model-based control of high-consistency refining", *Tappi journal (USA)*, vol. 10, pp. 140–146, 1996.

[18] E. Härkönen, E. Huusari, and P. Ravila, "Residence time of fibre in a single disc refiner", *Pulp & Paper Canada*, vol. 101, pp. 330–335, 2000.

[19] M. Sabourin, N. Wiseman, J. Vaughn, *et al.*, "Refining theory considerations for assessing pulp properties in the commercial manufacture of TMP", in *55th Appita Annual Conference*, Appita Inc., 2001, pp. 195–204.

[20] E. Härkönen, J. Kortelainen, J. Virtanen, and P. Vuorio, "Fiber development in TMP main line", in *Int. Mech. Pulping Conf., Quebec, Canada*, 2003, pp. 171–178.

[21] B. Strand and B. Grace, "Implementation of advanced supervisory control within a TMP refiner quality control system", in *Int. Mech. Pulping Conf., Helsinki, Finland*, 2014.

[22] E. Nelsson, "Improved energy efficiency in mill scale production of mechanical pulp by increased wood softening and refining intensity", PhD thesis, Mid Sweden University, 2016.

[23] A. Karlström and K. Eriksson, "Fiber energy efficiency part I: Extended entropy model", *Nord. Pulp Pap. Res. J.*, vol. 29, no. 2, pp. 322–331, 2014.

[24] ——, "Fiber energy efficiency part II: Forces acting on the refiner bars", *Nord. Pulp Pap. Res. J.*, vol. 29, no. 2, pp. 332–343, 2014.

[25] ——, "Fiber energy efficiency part III: Modeling of bar-to-fiber interaction", *Nord. Pulp Pap. Res. J.*, vol. 29, no. 3, pp. 401–408, 2014.

[26] ——, "Fiber energy efficiency part IV: Multi-scale modeling of refining processes", *Nord. Pulp Pap. Res. J.*, vol. 29, no. 3, pp. 409–417, 2014.

[27] A. Karlström, J. Hill, R. Ferritsius, and O. Ferritsius, "Pulp property development part I: Interlacing undersampled pulp properties and TMP process data using piece-wise linear functions", *Nord. Pulp Pap. Res. J.*, vol. 30, no. 4, pp. 599–608, 2015.

[28] A. Karlström, J. Hill, O. Ferritsius, and R. Ferritsius, "Pulp property development part III: Fiber residence time and consistency profile impact on specific energy and pulp properties", *Nord. Pulp Pap. Res. J.*, vol. 31, no. 2, pp. 300–307, 2016.

[29] A. Karlström, J. Hill, R. Ferritsius, and O. Ferritsius, "Pulp property development part II: Process nonlinearities and their influence on pulp property development", *Nord. Pulp Pap. Res. J.*, vol. 31, no. 2, pp. 287–299, 2016.

[30] M. C. Arranz, W. Birk, and G. Nikolakopoulos, "A survey on control configuration selection and new challenges in relation to wireless sensor and actuator networks", *IFAC World Congress*, vol. 50, no. 1, pp. 8810–8825, 2017.

[31] M. Witcher and M. T, "Interacting control systems: Steady state and dynamic measurement of interaction.", *ISA transactions*, vol. 16, no. 3, pp. 35–41, 1977.

[32] P. Grosdidier, M. Morari, and B. R. Holt, "Closed-loop properties from steady-state gain information", *Industrial & Engineering Chemistry Fundamentals*, vol. 24, no. 2, pp. 221–235, 1985.

[33] M. C. Arranz and W. Birk, "New methods for structural and functional analysis of complex processes", in *Control Applications,(CCA) & Intelligent Control,(ISIC)*, IEEE, 2009, pp. 487–494.

[34] P. A. Knight, "The Sinkhorn–Knopp algorithm: Convergence and applications", *SIAM Journal on Matrix Analysis and Applications*, vol. 30, no. 1, pp. 261–275, 2008.

[35] T. P. Chiang and W. L. Luyben, "Comparison of the dynamic performances of three heat-integrated distillation configurations", *Industrial & Engineering Chemistry Research*, vol. 27, no. 1, pp. 99–104, 1988.

[36] K. J. Åström and T. Hägglund, "Benchmark systems for PID control", *IFAC Proceedings Volumes*, vol. 33, no. 4, pp. 165–166, 2000.

[37] F. Bengtsson and T. Wik, "A multiple input, multiple output model generator", Department of Signals and Systems, Chalmers University of Technology, Tech. Rep., 2017.

[38] K. J. Åström, T. Hägglund, C. C. Hang, and W. K. Ho, "Automatic tuning and adaptation for PID controllers-a survey", in *Adaptive Systems in Control and Signal Processing 1992*, Elsevier, 1993, pp. 371–376.

[39] K. H. Ang, G. Chong, and Y. Li, "PID control system analysis, design, and technology", *IEEE Transactions on Control Systems Technology*, vol. 13, no. 4, pp. 559–576, Jul. 2005, ISSN: 1063-6536.

[40] J. Higham, "Single-term control of first-and second-order processes with dead time", *Control*, vol. 12, no. 116, p. 136, 1968.

[41] E. Dahlin *et al.*, "Designing and tuning digital controllers", *Instruments and Control systems*, vol. 41, no. 6, pp. 77–83, 1968.

[42] H. Panagopoulos, T. Hägglund, and K. J. Åström, "The lambda method for tuning PI controllers", Department of Automatic Control, Lund Institute of Technology, Tech. Rep. 7564, 1997, p. 5316.

[43] D. E. Rivera, M. Morari, and S. Skogestad, "Internal model control: PID controller design", *Industrial & engineering chemistry process design and development*, vol. 25, no. 1, pp. 252–265, 1986.

[44] T. Glad and L. Ljung, *Control theory, Multivariable and Nonlinear Methods*. Taylor & Francis, 2000.

[45] K. J. Åström and B. Wittenmark, *Computer-controlled systems: theory and design*. Courier Corporation, 2013.

[46] O. C. Imer, S. Yüksel, and T. Basar, "Optimal control of LTI systems over unreliable communication links", *Automatica*, vol. 42, no. 9, pp. 1429–1439, 2006, ISSN: 0005-1098.

[47] C. N. Hadjicostis and R. Touri, "Feedback control utilizing packet dropping network links", in *Proceedings of the 41st IEEE Conference on Decision and Control*, IEEE, vol. 2, 2002, pp. 1205–1210.

[48] L. Schenato, B. Sinopoli, M. Franceschetti, K. Poolla, and S. S. Sastry, "Foundations of control and estimation over lossy networks", *Proceedings of the IEEE*, vol. 95, no. 1, pp. 163–187, 2007.

[49] R. T. Sukhavasi and B. Hassibi, "Error correcting codes for distributed control", 2011, Available on-line at https://arxiv.org/abs/1112.4236v.

[50] R. Sukhavasi and B. Hassibi, "Tree codes improve convergence rate of consensus over erasure channels", in *IEEE 51st Annual Conference on Decision and Control (CDC)*, IEEE, 2012, pp. 32–37.

[51] J. Yu and L. Fu, "A new compensation framework for LQ control over lossy networks", in *2013 IEEE 52nd Annual Conference on Decision and Control (CDC),*, Dec. 2013, pp. 6610–6614.

[52] L. Schenato, "To zero or to hold control inputs with lossy links?", *IEEE Transactions on Automatic Control*, vol. 54, no. 5, pp. 1093–1099, 2009.

[53] O. Ferritsius, R. Ferritsius, J. Hill, A. Karlström, J. Ferritsius, and K. Eriksson, "Process considerations and its demands on TMP property measurements–a study on tensile index", *Nordic Pulp & Paper Research Journal*, vol. 32, no. 1, pp. 45–53, 2017.

[54] P. C. Young, *Recursive estimation and time-series analysis: An introduction for the student and practitioner*. Springer Science & Business Media, 2011.

[55] R. Sanchis and P. Albertos, "Recursive identification under scarce measurements-convergence analysis", *Automatica*, vol. 38, no. 3, pp. 535–544, 2002.

[56] R. V. Hogg, J. McKean, and A. T. Craig, *Introduction to mathematical statistics*. Pearson Education, 2005.

[57] D. A. Belsley, E. Kuh, and R. E. Welsch, *Regression Diagnostics. Identifying Influential Data and Sources of Collinearity*. Hoboken, USA: John Wiley & Sons, 1980.

[58] R. D. Cook, "Detection of influential observation in linear regression", *Technometrics*, vol. 19, no. 1, pp. 15–18, 1977.

[59] S. Skogestad and M. Morari, "Robust performance of decentralized control systems by independent designs", *Automatica*, vol. 25, no. 1, pp. 119–125, 1989.

[60] H. Ito, H. Ohmori, and A. Sano, "Robust performance of decentralized control systems by sequential designs", in *[1992] Proceedings of the 31st IEEE Conference on Decision and Control*, IEEE, 1992, pp. 1333–1339.

Part II

Papers