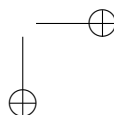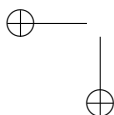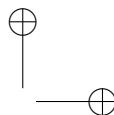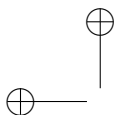# Model-based Process Monitoring and Control of Micro-milling using Active Magnetic Bearings

MODEL-BASED PROCESS MONITORING
AND CONTROL OF MICRO-MILLING USING
ACTIVE MAGNETIC BEARINGS

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof.ir. K.C.A.M. Luyben,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op

dinsdag 17 mei 2011 om 10:00 uur

door

Rogier Sebastiaan BLOM

elektrotechnisch ingenieur
geboren te Haarlem.

disc

Micro Ned

Printed in the Netherlands.

Acknowledgments

Those close to me will know me that the decision to pursue a Ph.D. has been a very conscious one. Over the past years, although having experienced both high and low points, the enthusiasm for doing Ph.D. research has never jaded. This had not been possible without the people around me, who contributed in different ways.

First of all, I am indebted to professors Paul Van den Hof and Rob Munnig Schmidt as my promotors. Their guidance and sharp insights have allowed me to develop as a person and to improve my work. I am also grateful to professor Bernhard Karpuschewski for accepting me as a Ph.D. candidate. On day-to-day basis, I have received the supervision of Hans Langen, André Hoogstrate and Marcel Achtsnick during different stages of the project, for which I offer them my appreciation. With the multiple disciplines involved in this research, it was particularly interesting to exchange ideas with them on how problems are described and approached. I would like to acknowledge professors Johan Schoukens, Gerd Vandersteen, Raymond de Callafon and Carsten Scherer for allowing me to use their time. Their comments, suggestions, and feedback have been very valuable.

One of the rewarding aspects of doing Ph.D. research is working in a diverse and stimulating environment. I have enjoyed being surrounded by many inspiring people in the DCSC and PME departments, it was a real pleasure working with you! Special thanks for all the many interesting discussions go to Jeroen Derkx, Xavier Bombois, Roland Tóth, and Vincent Henneken. Also thanks to the entire Microfactory team, and in particular Peiyuan, Maarten, Navin, Aleksandar and Guido. I also have very good memories of the collective lunches with the colleagues of the former PMA department, so thanks to all of you! Peiyuan and Iwan, I enjoyed sharing the office with you and our numerous stimulating discussions on cultural differences. A word of gratitude also to the supportive technical and administrative staff, in particular Harry, Jos, Marli, Olga, Inge, Birgit, Corinne, Ellen, Esther, and Kitty.

My words of deepest appreciation are to my dear wife Rachel. From the very

beginning you understood my desire to pursue a Ph.D. and gave your full support, knowing what it would take. Your selflessness in all these years amazes me to this day. Your love has been an indispensible source of enthusiasm, encouragement, inspiration and dedication. You gave what I needed and much more than that, for which I am immeasurably grateful.

A final word to our son Elian: it might take a few years until you are able to read this, but nevertheless I want everyone to know that I am very grateful for being your father. I am immensely proud of you.

Attenkirchen, Deutschland, April 2011
Rogier S. Blom

Contents

Contents

CONTENTS

Chapter 1

Introduction and research goal

1.1 Introduction

The advances in miniaturization have had an undeniable impact on our modern existence. It is hard to imagine our current daily lives without the effects of miniaturization on how we work and communicate, how we travel, how we treat the sick, and the way we live and relax. For a large part, this is owing to the enormous progress in the lithographic technologies. These enabled the fabrication of complex microelectronic systems, micro-electromechanical systems (MEMS), and micro-optoelectromechanical systems (MOEMS), resulting in a plethora of applications in many disciplines and industries. Still, the lithographic technologies have some principle shortcomings, limiting the types of miniaturized components that can be fabricated. The limitations are found in the inability of these techniques to produce components with arbitrary 3D features, the small range of materials that can be processed, and the inefficiency for smaller batch sizes. On the other hand, there is a still growing demand for even smaller and increasingly complex devices, requiring several mechanical, fluidic, electronic and/or chemical functions integrated together. Examples of such miniature products can be found in many areas:

Medical The body has limited place for 'extra' hardware, so implants like pain relief devices, drug delivery systems, hearing aid devices and various intelligent *in vivo* health monitoring or diagnostic devices are preferably as small as possible;

Aerospace In particular micro sensors, flow control devices, micro-turbines, and microscale fuel cells have been mentioned as applications [103];

Automotive In this area, small components are required for convenience, safety and entertainment. Popular examples are micro-parts for airbags, fuel injection nozzles and various actuators (ABS, headlight / mirror adjustment);

1. INTRODUCTION AND RESEARCH GOAL

Electronics & Communications In general, interest exists in miniaturized consumer products, such as digital cameras and mobile phones, combining several functions energy-efficiently. Other driving applications are found in fibre optic components.

This trend emphasizes the need of a wide range of reliable precision manufacturing techniques in the micro-domain. Recognition of this need has led to the development of several micro-fabrication techniques, including laser beam machining (LBM), micro-electric discharge machining (EDM), micro-electrochemical machining (ECM) and micro-milling, together with replication techniques for serial manufacture, such as micro-injection moulding, and hot-embossing.

In the past years, there has been a considerable amount of research progress in the development and improvement of these techniques. Multiple authors have reviewed and classified the broad range of micro-manufacturing techniques, indicating the distinct advantages and disadvantages of the various techniques over the others for given applications. A comprehensive comparative analysis of micro-manufacturing techniques is given by Masuzawa [111], while Alting et al. in [8] analyze the field of micro-engineering in general.

The focus in this thesis is on micro-milling. This micro-manufacturing process is particularly attractive because of its relatively large material removal rates compared to e.g. EDM and LBM, and its flexibility in producing different component sizes, shapes, features, and the ability to machine a variety of materials, including most metals and plastics. It is suitable for machining complex 3D micro-structures with high aspect ratios, and is mentioned specifically for its applicability for the fabrication of moulds for micro-forming processes [27].

This thesis reports on research into new directions for model-based process monitoring and control of the process of micro-milling, performed with setups having a spindle with active magnetic bearings (AMBs). The intent of this chapter is to introduce the reader to this research. In Section 1.2, we will briefly review the process of micro-milling. The key challenges encountered in micro-milling are discussed, as well as the main trends in the development of micro-milling machinery. Subsequently, in Section 1.3 a concise overview is given of the vast amount of results available for process monitoring and control in conventional manufacturing. From there, the implications of scaling for process monitoring and control are discussed. The state of the art of AMB spindle technology is covered in Section 1.4. The active magnetic bearing technology itself will be introduced, accompanied with an overview of available results in which the active nature is employed for monitoring or control purposes. Section 1.5 is the pivotal part of this chapter. In this section, the central research question is specified in more detail and the goal of this thesis is laid out. The concluding Section 1.6 is added as a guide to the reader and provides an overview of the content of the subsequent chapters.

(a)   (b)

Figure 1.1: Examples of small milling tools. Fig (a): A line up of various tools, from left to right: ⌀ 0.5 mm and ⌀ 0.3 mm tools with CrCN coating (for cutting of copper), and ⌀ 0.2 mm, ⌀ 0.3 mm, and ⌀ 1.0 mm tools with TiAlN coating (for cutting of steel). Fig (b): A zoomed-in comparison of the tooltip of a micro-mill with ⌀ 0.2 mm (left) and that of a mill with ⌀ 1.0 mm (right).

1.2 Micro-Milling: scaling of the cutting process to the micro-domain

Milling is the manufacturing process that involves the cutting and shaping of materials into products (*workpieces*) using rotating tools with defined cutting edges (*milling tools*). Material removal happens through the formation of chips at the cutting edge of the tool, while the tip of the rotating milling tool traces a contour along the workpiece (*tool path*). A machine tool capable to perform a milling job (*milling machine*) uses a *spindle* to provide rotation of the milling tool. It has several *axes* of motion to provide the relative movement of the milling tool and workpiece. The technology referred to as *micro-milling* is simply scaled down milling, which means that the milling process is performed using tools with rather small diameters (see Figure 1.1). There is no strict definition, but common understanding is that a micro-milling tool has a radius smaller than 1 mm [112].

The start of micro-milling research can be traced back to the mid 1990s. To date, it is still a very active area of research with many open questions. In this section, we will not give a full overview of the current state of research, for this the reader is referred to excellent surveys by Chae [31], Liu [104] and Li [98]. Instead, we will highlight the main challenges and issues with the micro-milling process. Indeed, reducing the tool diameter into the range below 1 mm results in a number of issues that make micro-milling fundamentally different from conventional machining. In Section 1.2.1 we discuss the most prevalent of these. In response to these issues, some trends can be observed in the development of machinery for micro-milling, which will be covered in Section 1.2.2.

1.2.1 Main issues in micro-milling

Compared to conventional — macro-scale — milling, the material removal mechanism of micro-milling is similar, i.e. material is removed through the formation of chips at the defined cutting edges of the rotating milling tool. However, there are some essential differences, which arise due to the effects of scaling. As a result, the chip formation process is different with scaled-down milling tools, while also several other factors start to play a more dominant role at these reduced tool diameters. In this section we will review some of the main issues which concern the chip formation, the required high rotational speeds, tool life, precision, and workpiece quality.

Chip formation

The first effect of scaling of the milling process to the micro-domain that we will discuss, is the impact of the limitations on the achievable sharpness of cutting tools. This limit depends on the properties of the tool material, where hard metal (tungsten carbide) tools are most commonly used to mill metallic materials. With careful grinding, edge radii of $2-3\mu$m can be reached with ultra-fine grain tungsten carbide tools. Lower values are unlikely, with its grains being nearly 1μm in diameter [142]. Although such tools can be considered very sharp at the macro-scale, at the micro-scale where chip thicknesses are in the orders of micrometers, the cutting process appears to take place with a rather blunt tool[1], see Figure 1.2. The implication of this, is that the chip formation process at the micro-scale is fundamentally different compared to milling at the macro-scale. In recent years, several studies have been carried out to gain better understanding of the chip formation process in micro-milling and the characteristics of the cutting forces during micro-milling [88, 18, 171, 185, 96, 25].

An important phenomenom that was noted in micro-chip formation is the Minimum Chip Thickness (MCT) effect [75, 179, 142, 87, 172]. In essence this effect can be described as follows. If the uncut chip thickness is smaller than some critical value (the MCT), no chip is formed, but instead the workpiece material is forced under the tool and deformed. When this happens, the cutting process can become intermittant, i.e. with each tool pass the uncut chip thickness accumulates, the machining forces increase, and high tool wear is incurred [50]. This happens for a number of passes, until the uncut chip thickness exceeds the MCT. Only then a chip is formed. The large forces and high tool wear experienced in this cutting mode clearly indicate that cutting with chip loads below the MCT should be avoided.

Studies show that the sharpness of the cutting tool predominantly determines the MCT, while the value also depends on the properties of the workpiece material [179, 172]. The reported ratio of the MCT to the cutting edge ratio is in the range of $10-40\%$ [87, 172].

---

[1]In literature this is referred to as the effective negative rake angle effect, see e.g. [104, 44].

4

Micro-Milling: scaling of the cutting process to the micro-domain
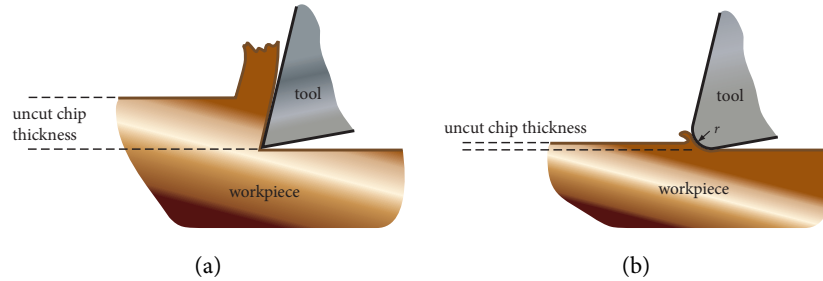
Figure 1.2: Tool-workpiece interaction in macro-scale cutting (a) and in micro-scale cutting (b). In micro-scale cutting the edge radius r is large compared to the uncut chip thickness.

Parameter	Unit	Symbol
Number of teeth	-	N
Tool diameter	mm	d
Rotational speed	rpm	n
Feed speed	mm/min	v_f
Depth of cut	mm	a_p
Width of cut	mm	a_e
Feed per tooth	μm	$f_z = \frac{v_f}{n \cdot N} \cdot 10^3$
Cutting speed	mm/min	$v_c = \frac{1}{2} n \cdot d$
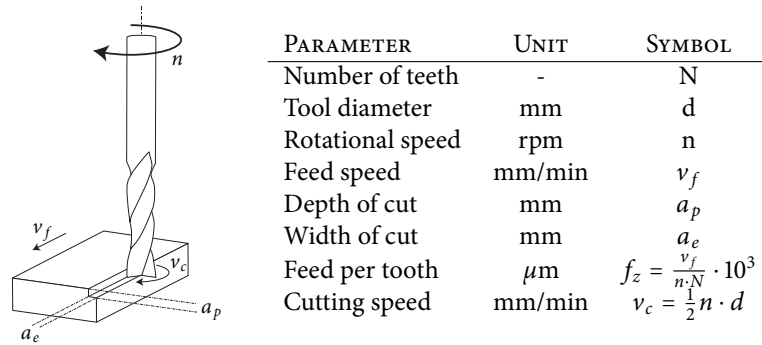
Figure 1.3: Definitions of common cutting parameters.

High rotational speeds

Micro-milling is mostly carried out at very high rotational speeds, often at several tenthousands revolutions per minute (rpm) [134, 96, 114, 103], or even at speeds of over 120, 000 rpm [50, 143, 21]. The first reason behind this is of an economical nature. When the tool diameter is reduced, the material removal rate drops correspondingly. Increasing the rotational speeds allows for compensating for this loss. The second reason is more fundamental. Many decades of research and experience in conventional machining have led to recommendations for cutting parameter selection for high speed milling, including recommendations for the cutting speed for machining different materials (the cutting speed is defined as the surface velocity of the cutting edge of the milling tool, see Figure 1.3). These guidelines are assumed to be applicable also for the micro-milling process. However, when reducing the diameter of cutters, obviously the rotational speed needs to be increased to

TOOL DIAMETER	SPINDLE SPEED (RPM)	
	Tool steel	Aluminum
20 mm	$3,000 - 7,500$	$7,500 - 30,000$
2 mm	$30,000 - 75,000$	$75,000 - 300,000$
0.5 mm	$120,000 - 200,000$	$200,000 - 800,000$
0.1 mm	$600,000 - 1,000,000$	$1,000,000 - 4,000,000$

Table 1.1: Optimal rotational speeds for decreased cutter diameters. The given numbers are based on recommended cutting speeds for high speed cutting of tool steel ($200 - 500$ m/min) and aluminium ($500 - 2,000$ m/min). These ranges are indicative and vary with tool geometry, cutting conditions and hardness of workpiece material.

maintain the same cutting speed. Table 1.1 shows a simple calculation, illustrating that to maintain optimal cutting speeds with tools with diameters of less than 0.5 mm when cutting materials like tool steel and to an even greater extent aluminum, rotational speeds far above $200,000$ rpm are needed. Such speeds exceed the limits offered by state-of-the-art spindle technology. It is for this reason, that numerous attempts have been done to find new solutions for extremely high speed rotation. High rotational speeds also imply that the feedrate should be high enough. This is to keep the uncut chip thickness per tooth above the MCT. Therefore, micro-milling machine tools also require high velocities and accelerations of the feed axes [179].

Tool life

The wear and breaking mechanism of micro-milling is very different compared to macro-scale tools [157]. During usage, macro-scale tools loose their form by wear[2], and cutting edges may chip off, generally one by one, and only partially. As a result, the performance of the milling process deteriorates gradually. The end of tool life is met when the performance no longer meets some defined quality standard.

In micro-milling however, premature catastrofic breakage of the shaft is one of the major factors limiting the tool life [157, 134]. The unpredictable tool wear and premature tool failure are serious issues in micro-milling. In an early case study by Li [98], a micro-mould was milled, which involved the removal of only ± 4 mm^3 of material. Nonetheless, it was reported that while executing this micro-milling task, 12 milling tools broke. Often when a tool breaks, the process needs to be restarted, since accurately realigning workpiece and tool is hardly feasible and the incurred loss of precision is generally unacceptable.

Hence, understanding of underlying causes of tool breakage in micro-milling is needed to maintain an efficient process. Rahman et al. [134] attribute premature tool failure to non-uniform tool wear, leading to a sudden increase of the cutting

---

[2] In conventional milling several wear mechanism are known. See [9] for a comprehensive overview.

forces. Tansel et al.[157] mention several causes for tool breakage, including fatigue related breakage, clogging of the cutting edge, and excessive stress related breakage.

Precision

Miniaturization puts a strong requirement on the precision of manufacturing processes. Indeed, in general when components are scaled-down, the demands on the relative tolerance level (i.e. the tolerance to feature size ratio) will remain the same. In this regard, micro-manufacturing is tightly connected to precision manufacturing. In precision manufacturing macro-scale components (e.g. mirrors and lenses for optical applications) are machined with extremely high relative tolerance levels of 10^{-7} or smaller. On the other hand, in micro-manufacturing components with dimensions of several millimeters and feature sizes in the micrometer range are machined with relative tolerance levels of 10^{-3} to 10^{-5}, requiring absolute precision levels in the same range as with precision manufacturing.

However, as a result of the reduced dimensions of the milling tool, the susceptibility of the micro-milling process for various error sources increases. Hence, while the process is scaled-down, it becomes increasingly more difficult to maintain the absolute precision of the process, let alone to satisfy the relative tolerance specification. In the remainder of this section we will illustrate this by discussing the three main sources of error: tool runout, tool deflection, and machine vibrations.

Tool Runout Tool runout occurs when a milling tool rotates around another axis than its main symmetry axis. There are two main causes for runout to occur. The first is related to the dynamics of the milling machine and is caused by the dynamic response of the machine to a mass unbalance in the rotor. The second is related to alignment errors of the tool, the toolholder and the rotor.

While the effect of runout in conventional milling is small, with the reduced uncut chip thicknesses in micro-milling, the effect of runout increases. When cutting with milling tools with two cutting edges, it is reported that due to run-out, the cutting process can become rather unbalanced with only one cutting edge performing most of the cutting, while the other edge hardly touches the workpiece [19, 45, 99]. Apart from a loss of dimensional precision, high runout is therefore detrimental to the micro-milling process. It leads to accelerated and non-uniform tool wear, as well as large cutting force variations, and — as a result — a serious reduction of the tool life.

Tool deflection Reduction of the tool diameter causes the tool to become increasingly more compliant. This scaling effect is quite strong. Applying cantilever beam theory, it can be noted that the stiffness of a tool is approximately proportional to the fourth power of its diameter (see e.g. [138]). With tool diameters of less than 1 mm, the loss of tool stiffness becomes very significant. Indeed, as shown by Uriarte et al. [163], in micro-milling up to 90% of the

total compliance at the tool tip can be attributed to the compliance of the tool itself. Reported stiffness values for 0.5 mm diameters tools are as low as $0.3 - 0.7$ N/μm [98]. Although cutting forces in micro-milling may only be a few Newtons, having such a limited stiffness, the micro-cutting tools will deflect by several microns. The loss of dimensional precision is therefore quite significant.

Machine vibrations Milling machine tools are complex dynamic systems and in general exhibit several vibration modes. When these vibration modes are excited — depending on the corresponding mode shape — they may cause the tool tip to deviate from the programmed contour, leading to a loss of precision. In micro-milling this is no different than in precision manufacturing. However, vibrations can have a similar effect as runout in micro-milling: even small vibration amplitudes at the tooltip can cause a deteriorated cutting process, as they may cause large variations of the uncut chip thickness.

The vibrations that arise during milling can be divided in three types: free vibration, forced vibration, and self-excited vibration. Free vibration is the vibration of the machine tool under any kind of external disturbance. Forced vibration is the response of the machine to the periodic forces that arise during the cutting process. In particular situations, these vibrations may be self-exciting. A vibrating tool edge leaves a wavy pattern in the workpiece after removing a chip. The spatial frequency of this wave pattern will be related to the frequency of the vibration and the rotational speed of the tool. The varying uncut chip thickness experienced at the passing of the next cutting edge, causes the cutting forces to vary with a similar frequency as the vibration experienced at the previous cut, and this force in turn excites the machine tool. Under specific conditions this mechanism of self-exciting vibrations can cause highly resonant behavior, which is generally referred to as *chatter*. Chatter is a very much unwanted process phenomenon, for causing severe surface quality deterioration, tool damage and possibly machine damage. For the micro-milling process chatter also constitutes a significant problem [30]. The excessive vibrations that arise due to chatter not only affect the workpiece surface quality, but can lead to catastrophic failure.

Recently some studies on vibrations in micro-milling have been published. An experimental study of chatter in micro-milling has been performed by Baschin et al. [21]. Jun et al. have developed a dynamic model to predict cutting forces and micro-mill vibrations, and performed an experimental study to verify this model [81, 80].

Workpiece quality

The quality of a machined workpiece depends on various factors, which apart from the dimensional precision, include the surface finish (roughness), surface integrity,

and burr size. Several experimental studies have been carried out to find a relation between cutting parameters, tool characteristics and workpiece material properties on one hand, and the workpiece quality on the other hand (see e.g. [50, 114]).

It is noted that in particular burr formation is a dominant factor limiting workpiece quality in micro-milling. In manufacturing, burrs are the rough edges or small pieces of material remaining attached to the workpiece after a cutting operation. The problem of burr formation in micro-milling is twofold. First, in comparison to the uncut chip thickness, burrs arising in micro-milling are usually larger than in conventional cutting [97, 50]. This appears to be due to the edge radius effect [97]. Secondly, burrs resulting from micro-milling are hard to remove, as removal of burrs by secondary operations may introduce dimensional errors and residual stresses in the component [141, 97].

A few studies have been carried out on micro-burr formation. Lee et al. [97] studied the relationship between cutting conditions, tool life and burr size in cutting of stainless steel. In a machinability study of pure copper, Filiz et al. [50] report that burr formation increases considerably as the wear progresses, while lowest burr formation occurs at low cutting speeds and high feedrates. Aramcharoen et al. [15] have observed a relationship between burr formation and the ratio of uncut chip thickness to cutting edge radius in micro-milling of tool steel. They note that micro-milling with an uncut chip thickness equal to the edge radius gives the best result in terms of surface finish and burr size. This has been confirmed by the study performed by Mian et al. [114].

1.2.2 Trends in micro-milling machine tool technology

To perform an accurate, reliable, and efficient micro-milling process, special equipment is necessary. In this section, we will discuss the main trends in the developments of micro-milling machine tool technology.

Following from the discussion in the previous section, it can be stated in general that micro-machine tools need to have a submicron positioning accuracy, low spindle runout ($< 1\mu$m), and good vibration isolation. In that sense, the similarities between micro-milling machine tools and ultra-precision machine tools are high. Indeed, it is noted that several experimental micro-milling studies are carried out using precision or ultra-precision machine tools [168, 134, 21, 98], although it is also seen that research groups build their own machine to meet requirements on precision, stiffness, damping and flexibility needed for micro-milling research [179, 134, 142, 172, 17, 96]

A nice overview of the latest developments in micro-milling machine tool technology is given by Uriarte et al. in Chapter 11 of [106]. Here, a discussion is given of the main components of the micro-milling machine, like the drives, guides, measuring systems, and spindle, accompanied with the most recent advances in these components. It is noted that several manufacturers of ultra-precision machine tools now also offer special models for micro-milling (e.g. Kern, Sodick, Moore, Fanuc,

Kugler, and Makino). These models are built to manufacture smaller components and thus have smaller working areas, and are equipped with faster spindles to achieve the required cutting speeds with small diameter tools.

An important trend in machine tool development for micro-milling concerns the attempts to increase rotational speed capabilities of micro-milling spindles. Indeed, as indicated in the previous section, in order to maintain acceptable cutting speeds, the rotational speed requirements grow rapidly with reduced tool diameter. State of the art spindle technology is able to reach rotational speeds of up to 200, 000 rpm. High precision ball bearings are most common, but air bearings are also applied [106]. Research initiatives are seen to attain even higher rotational speeds. In [57], a rotational speed of up to 450,000 rpm is achieved by using the tool shank itself as the spindle shaft. A friction drive with drive ratio of 9:1 is used to transmit the torque from a commercially available high speed (90,000 rpm) spindle to achieve these speeds. Other concepts that have been explored for ultra-fast rotation are air turbine spindles [47], and AMB spindles [89].

Another noteworthy trend is the Microfactory concept [119], also referred to as Micro/Meso Mechanical Manufacturing (M4) [48]. The main philosophy here is that the manufacturing of miniature components is technically better realized by machines many orders of magnitude smaller than those used for conventional manufacturing. The claimed advantages of miniaturizing the equipment for the performance and accuracy of micro-manufacturing processes are that miniature machines are

- faster: they can accelerate faster due to lower inertia, and can rotate faster due to lower centrifugal forces; and

- more accurate: due to higher natural frequencies, smaller vibration amplitudes, and smaller thermal deformations.

Besides these technical advantages of using miniaturized machinery for micro-manufacturing processes, also several economical advantages are mentioned. Smaller machines are expected to decrease floor space requirements, reduce energy consumption, improve material resource utilization, and provide higher manufacturing flexibility (e.g. reconfigurability of the production lay-out). Moreover, the Microfactory concept is believed to enable on-site production of customized miniature products, which is expected to give a large impetus to miniaturization [119, 162].

Following this concept, numerous research efforts are directed towards miniaturization of micro-manufacturing equipment, typically combining several machining, inspection and assembly processes. Okazaki et al. review the concept, history and developments of the Microfactory in [119]. The WTEC report [48] contains a comprehensive overview of the worldwide research activities in this area.

One of the key challenges encountered with the reduced dimensions of miniature machines, is their relatively low structural rigidity, which forms an obstacle in achieving the high demanded precision and surface finish levels. In the Mi-

crofactory cluster of the MicroNed program [2], this issue is addressed through mechatronic design of miniaturized machinery and by applying modern control approaches for disturbance suppression [95]. The MASMICRO project [1] seeks to find a trade-off between the large, but accurate conventional ultra-precision machines and the small, but fast miniaturized machines. In this project a compact, energy-efficient bench-top five axis ultra-precision micro-milling machine was developed and realized — the UltraMill [73, 74].

1.2.3 Concluding remarks

It is recognized that the process of micro-milling is a promising technology for the fabrication of micro-parts in a wide range of materials and with arbitrary 3D features. However, some technological hurdles need to be taken to make this technology economically viable. More research is required to better understand the chip formation process at the micro-scale, in order to define cutting strategies and optimal combination of cutting parameters. Furthermore, improved micro-tool design is needed, targeted at extending the tool life and increasing the stiffness. In parallel to such studies, further improvement of micro-milling equipment will result in an efficient and reliable cutting process, yielding micro-parts with demanded precision and surface quality.

1.3 Monitoring and control of micro-milling

It is widely accepted that sensor based manufacturing is vital to achieve low downtime levels of manufacturing systems in conjuction with high quality levels of the manufactured components [28]. With advanced sensors and adequate signal processing, information about the process condition is obtained, allowing for process optimization and control.

When scaling down the process of milling, the importance of monitoring and control increases. Signs of problems are almost unnoticeable without the use of special equipment. This stresses the need for reliable systems that are able to detect —and possibly even predict — anomalies in the process. Systems are needed that online monitor the condition of the cutting process and the condition of the micro-milling tool. Besides, in order to reach the desired accuracies and to obtain an efficient milling process, control of cutting forces, and tool deflection is required. The purpose of this section is to discuss the implications of scaling down the milling process to the micro-scale for process monitoring and control.

Monitoring and control has been a topic of research for decades in conventional machining. Hence, in Section 1.3.1 we will start with a classification of the various types of monitoring and control in this area. Subsequently in Section 1.3.2, we will set out the reasons why additional research on monitoring and control is still needed for the micro-milling process. This will be accompanied with a discussion of the progress that has been made in this area.
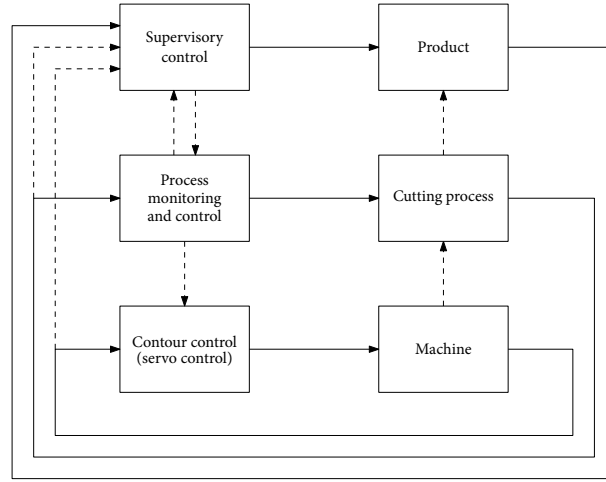
Figure 1.4: A simplified diagram illustrating the control hierarchy in manufacturing processes. At the lowest level, contour control is used to let the machine trace a predefined trajectory. At the level of process monitoring and control, process variables are monitored and controlled to ensure an optimal material removal process. Supervisory control maintains the effectivity and efficiency of the machining task.

1.3.1 Monitoring and control in conventional manufacturing

In many manufacturing processes a hierarchical control structure can be identified. This is illustrated in Figure 1.4 (a refinement of the structure proposed by Ulsoy in [173]). We will clarify this structure in particular for the milling process. At the lowest control level, the primary control objective is to let the tool trace a predefined contour along the workpiece. This is realized by coordinate motion of the multiple motion axes of the machine tool. A trajectory is planned along the contour, and translated to reference trajectories for the individual motion axes of the machine. Using measurements of its actual position, each of these axes is controlled through servo control. The precision of a machine tool is heavily dependent on the performance of the servo control loops, as well as on the mechanical properties of the construction responsible for translating the motion of individual actuators to the displacement of the tooltip. At the level of contour control, precision is improved by performing compensation for mechanical errors, thermal deformations, deformation due to tool bending, etc.

As soon as the rotating cutting tool hits the workpiece material, a cutting process starts. Control of this process forms the second level in Figure 1.4, where process variables such as cutting forces and cutting temperature are controlled to maintain an efficient material removal process, yielding good quality parts. Process moni-

toring is used to obtain relevant information about the process variables. Process control is achieved through adaptation of the cutting parameters (like cutting depth, cutting speed, feed per tooth). This may be realized directly through spindle speed variation, or indirectly through the servo control by means of adaptation of the contour or trajectory planning along the contour (this is indicated in Figure 1.4 by the dashed arrow from 'process monitoring and control' to 'contour control').

The highest level of control is supervisory control. At this level, the effectivity and efficiency of the machining task is controlled. Product related variables are measured, such as precision, and surface roughness. Through adaptation of the cutting parameters (indicated by the dashed arrow via the 'process monitoring and control' block), the objective of supervisory control is to maintain product quality and productivity. Supervisory control also ensures correct operation and monitors for faults and failures in either the process or machine. This may include monitoring of the tool condition, and detection of force overloads, collisions and machine chatter [153]. In order to do this, supervisory control receives input information on variables of the machine, the cutting process and/or information from the process monitoring and control task (see Figure 1.4).

We will direct our attention primarily to the process and supervisory control levels. At these levels, the degree of *automated control* may vary. Traditionally, human machine operators perform non-automated control: they use their sensory systems to obtain information about the cutting process and machine condition, and perform process control and supervisory control through direct intervention. In optimizing the manufacturing process through *automated monitoring and control*, in general three strategies can be defined, with increased level of automation (following Tönshoff in [153] and [159]):

Open-loop monitoring The system measures one or more physical quantities. By means of appropriate signal processing, a process variable or condition of the machine tool is derived from these measurements. It may prompt for human intervention in case of anomalies or faults;

Open-loop diagnostic systems The system attempts to determine a functional or causal relationship between a machine failure and its cause;

Closed-loop control The system uses sensory information according to some control law to control the machine's actuators and/or cutting parameters, in order to adapt the machine to changes in the process environment.

The remainder of this section is divided into two parts, which elaborate on the aspects of measurement and closed-loop control. Indeed, in either of the three above mentioned strategies, obtaining appropriate measurements is of crucial importance. We will further elaborate on this aspect in the first part. It would lead too far to give a review here of the plethora of signal processing, fault detection and fault diagnosis techniques that have been applied for monitoring and control

of manufacturing processes. We will comment on results on closed-loop process control though in the second part.

Measurement

Monitoring and control starts with the measurement of physical quantities in the machine setup. Most commonly used in literature are direct measurement of cutting forces [36] and acoustic emission[160, 46, 101], but also servo control signals, spindle current or power signals ([161, 100, 35]), accelerations and displacements of different machine components [67, 182], and temperature signals. For more detailed overviews see Karpuschewski [85], Byrne et al. [28] and Prickett and Johns [132]. Of the mentioned sensor systems, most applied in commercial machine tools are spindle current sensors, servo control current sensors and acoustic emission sensors, which is related to the low cost of these sensor systems and the ease of integration with the machine setup.

It is well accepted that information on the forces that arise during cutting processes gives the most direct knowledge on the cutting process itself. Machine tools in laboratory settings are therefore often equipped with force measurement platforms, that provide direct measurements of these forces. Good and reliable measurement results can be achieved with such equipment, and these are useful for research on process characteristics and for studying optimal choices of cutting conditions. However, when force signals are to be used for monitoring the process in a production environment, there are a number of drawbacks to use such force measurement platforms:

- force measurement setups are expensive (some tens of thousands of euros, depending on the specifications);

- the force measurement platforms are mounted between the workpiece and machine table, taking up space and changing the machine dynamics;

- the systems are fragile and very sensitive to overload.

These considerations have led researchers to find ways to obtain cutting force data from other sources, such as displacement or acceleration measurements from elsewhere in the machine, measurements of the spindle power, and workpiece stage controller signals. Results on indirect force signal estimation with both model-free and model-based estimation approaches are available. Model-free approaches are given by Kang et al. [84], who add a piezo force sensor in the drive system of the table, and Li et al. [102] who estimate the force signal from measurements from a current sensor installed on the servo motor of a turning center. In the model-based approaches, commonly the cutting force in each direction is modeled as the state of a random walk process, which is then estimated using a state observer. A sensorless approach has been proposed by Shinno et al. [151], who estimate the cutting forces from the compensation signals of the XY stage controller. A Kalman

filter approach to design a state observer for the cutting forces is followed by Park and Altintas ([10, 122]). In their research, a spindle of a vertical machining center is fitted with three pairs of piezo-electric sensors. As an alternative to using force sensors integrated in the spindle, Albrecht et al. [7] pursued a solution using capacitive displacement sensors, which increased the measurement bandwidth from 350 to 1000 Hz. Chae and Park [30] used the Kalman filter approach to compensate for the dynamics of a dynamometer setup, and increased the bandwidth of this approach by adding acceleration measurements. It is observed that in all of these Kalman filter approaches, the state and noise covariance matrices are simply used to tune the solution. No modeling of the force signal or noise is performed.

Control

The objective of process control is to vary process variables (such as feed, cutting speed, depth of cut), in order to obtain and maintain an optimal cutting process. This is daily practice in any machining environment and happens both offline and online. Offline process control takes place at the planning of a machining task. Process variables are chosen to ensure a favorable cutting process, yielding parts with good surface finish, low burr formation, etc., at low cost. This choice is often based on expert knowledge and machining handbooks, listing recommended cutting parameters for different machining tasks. Online process control is performed by an operator during the execution of the task. Process variables such as spindle speed and feed are changed, e.g. to avoid chatter and excessive cutting forces.

Besides this kind of non-automated process control, in the manufacturing literature also automated process control techniques have been proposed. In this community these solutions are commonly referred to as 'adaptive control' techniques, although they are often not adaptive in the sense as used in the systems and control community. In the manufacturing literature, process control techniques are broadly classified into two types: adaptive control with optimization (ACO) and adaptive control of constraints (ACC) [135, 153, 173]. ACC systems are essentially feedback systems in which the control objective is to regulate the cutting force, spindle power, or some other process quantity at a preset value, by varying cutting parameters like the feed rate and velocity. The control objective in ACO systems is to optimize a predefined performance index, e.g. the material removal rate or the tool wear rate, and adjust in the cutting parameters to achieve that goal.

1.3.2 Monitoring and control of milling in the micro-domain

As stated in the introduction of this section, process monitoring and control becomes of increasing importance when reducing the tool diameter. We will further support this statement by analyzing in more detail how particular process monitoring and control tasks address the effects of scaling:

Vibration monitoring In Section 1.2.1 it was discussed how several factors (e.g. the MCT effect, excessive run-out, and chatter) can cause an intermittant or even unstable cutting process. When such a cutting condition arises, the consequence can be severe reduction of the workpiece quality, and possibly even catastrophic breakage of the tool. In conventional machining, such conditions are often well observable by a human operator. At reduced diameters this is no longer the case. By online monitoring such vibrations, it becomes possible to interfere when excessive vibrations are detected, and the mentioned effects can be mitigated.

Tool wear monitoring As mentioned in Section 1.2.1, the life of a micro-milling tool frequently ends abrubtly due to catastrophic breakage. Tool breakage may cause damage to the workpiece and should be avoided. As one of the main reasons for breakage is progressed wear, monitoring of tool wear allows for replacing tools before they break. Additionally, workpiece quality can be maintained, as worn tools cause lower surface quality and larger burr formation.

Force control When the cutting forces are regulated during a micro-milling task, the workpiece quality can be enhanced and the tool-life can be prolonged. Workpiece precision is improved, since regulating of the cutting forces ensures limited tool bending. Since excessive stress is avoided and fatigue is reduced, tool life is extended.

Vibration monitoring, tool wear monitoring, and force control have been studied extensively in literature for improving conventional machining processes. However, due to the effects of the scaling, these results cannot be applied directly to the micro-milling process. New techniques are needed that are tailored to the challenges and consequences of the reduced tool dimensions. These are summarized as follows:

A. Mission-critical Monitoring techniques intend to replace human ears and eyes, hence the techniques need to be very reliable and robust for changing cutting conditions.

B. High bandwidth When the rotational frequency of the milling spindle increases, the bandwidth of any signal related to the cutting process increases correspondingly. Consequently, the required measurement bandwidth of sensors used to measure these signals increases likewise. For cutting force measurements in micro-milling, this constitutes a serious challenge. For example, with a rotational frequency of 120,000 rpm and a tool with two teeth, the tooth pass frequency will be 4 kHz. The cutting force signal will contain several harmonics of this frequency. However, commercially available state-of-the-art force measurement systems have measurement bandwidths limited

to $\pm 2kHz$[3]. Development of new sensors, fusion of various sensor signals and indirect force estimation methods are therefore needed to acquire cutting force information with sufficiently high bandwidth for micro-milling.

C. Reduced signal quality Due to the reduced scale, signal magnitudes are smaller, resulting in deteriorated signal-to-noise ratio (SNR). Moreover, as a result of the increased complexity of the milling process at smaller scales, the interpretation of measured signals is more difficult [154].

D. Fast processing Whenever an anomaly is about to happen, quick action is needed to avoid catastrophic failure. This implies that the monitoring techniques need to be efficient, and that prompt action can be taken at a small time scale.

A few results for monitoring and control of micro-milling are available, which are predominantly tool condition monitoring solutions. Each of them are tailored to the micro-milling process by addressing one or more of the above issues. In the remaining part of this section we provide an overview of these results.

Tansel and coworkers are among the first authors who have studied process monitoring of micro-milling. In [157] they have analyzed the relation between the static part of the force in the feed direction and the tool condition in micro-milling, and used this to develop tool breakage detection methods using segmental averages and wavelet transformations. This approach is able to predict tool breakage, and in a subsequent paper [156], the authors presented a Smart Workpiece Holder able to perform a quick move in the opposite direction of the feed force once breakage was likely to occur in the next few tool rotations. This increased the tool life by more than 30%. To circumvent the low SNR of the cutting force measurements, tool wear estimation and tool breakage detection methods based on acoustic emission signals are presented in [158]. In [154] tool wear estimation has been performed using force signal features and a neural network. It is noted that this method works well, provided the cutting conditions are identical to the training set. This is quite a limitation. The authors have presented an interesting solution for this in [155]. Here the wear status is inspected during a machining task by regularly pausing the task and cutting a slot in a reference material (aluminum in this case), while measuring the forces. These measurements are used with a trained neural network to estimate the wear status of the tool. Since in this approach the cutting conditions during the wear inspection epochs are identical and independent of the part being machined, the reliability of the wear monitoring could be increased.

A model-based approach to tool wear estimation has been given by Bao and Tansel in [20]. In this method, experimental cutting force data has been used to determine the parameters of a wear model in conjuction with a micro-cutting force

---

[3]Typically such platforms have a natural frequency in the range of 500 Hz - 5 kHz, and the recommended measurement bandwidth is $\frac{1}{3}$ of the natural frequency.

model. Genetic algorithms were used to find the parameter values that minimize the average absolute error between measured and modeled cutting forces.

Malekian et al. [109] propose an approach for tool wear monitoring for micro-milling using a fuzzy-neuro approach. In this approach, the necessary high bandwidth and higher reliability of the monitoring system is achieved by fusion of several sensor signals, i.e. force, acceleration and acoustic emission. Input of the fuzzy-neural network are the RMS values of these signals. Using a training epoch in which the wear status of the tool is measured using a microscope, the network is trained to predict the wear status using the sensor signals.

Zhu et al. [186, 187] provide a solution to deal with the low SNR of sensor signals by training continuous Hidden Markov Models. The problem solved in this approach is to determine the most likely (hidden) wear state of the tool from the observed features in the cutting force (including statistics like mean, standard deviation, and temporal features like Daubechies wavelet coefficients).

Jemielniak et al. [76, 77] address the poor SNR of sensory information in micro-milling by proposing a multi-sensory approach. In this research a tool condition monitoring was developed that combines a large number of signal features from cutting force and acoustic emission measurements using a two-stage algorithm.

To the best knowledge of the author, no results have been published as of yet in which closed-loop process control has been implemented for the micro-milling process.

1.3.3 Conclusion

Sensor based machining has given an impetus to conventional machining by improving the reliability and efficiency of machining centers. However, due to the effects of scaling, new approaches to monitoring and control are needed for the micro-milling process. Sensor information is needed with sufficiently high bandwidth and SNR. Monitoring techniques need to be robust for changing conditions and should be fast. Some progress has been made in this area, although process control for micro-milling is still uncovered by the current state of research.

1.4 Active Magnetic Bearing Spindles

A key component of a milling machine is the spindle, which provides the rotation of the milling tool. The characteristic elements of a spindle are the rotor shaft to which end the milling tool is connected via a tool interface, a driving system (i.e. the motor), a bearing system, and a cooling system. Currently, a wide variety of spindle designs is available on the market, having different properties regarding stiffness, torque, power, speed, and thermal behavior. In this thesis we focus on the application of the particular class of spindles that have Active Magnetic Bearings (AMBs). As will become clear, the properties of such spindles make them particular attractive for application to micro-milling.

Displacement sensor

Current through
right coil

$$i_r = i_{bias} - i$$

$$i_l = i_{bias} + i$$

Current through
left coil
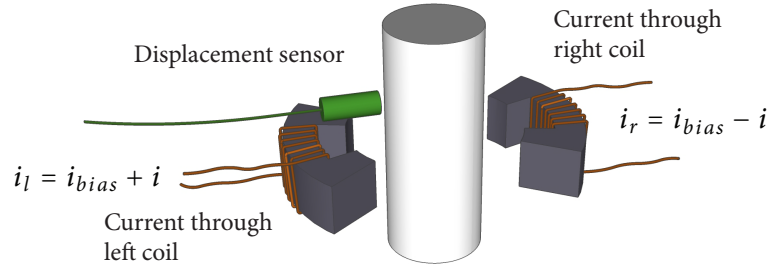
Figure 1.5: Schematic illustrating the basic working principle of an active magnetic bearing with two electromagnetic coils operating in differential driving mode.

The main function of the bearing system is to keep the rotating shaft in position, while minimizing the friction with the support. Active magnetic bearings achieve this by carrying the rotor in a controlled magnetic field. The basic principle of an AMB can be described as follows (see Figure 1.5). Electromagnets are positioned at opposite sides of a body. Most commonly these are operated in what is known as the *differential drive mode*. In this configuration, the same bias current i_{bias} flows through the coils of both electromagnets, causing an equal bias flux in the air gaps at each side. Additionally, a control current i is added to both coils, but with opposite sign. This implies that a nonzero control current will increase the flux in the air gaps of one coil by the same amount as it is reduced in the other, resulting in a net force on the body. This effect is used to control the position of the body. To do this, the actual position of the rotor is measured by a displacement sensor, and taken as input by a controller that controls the current levels in the coils, with the objective to keep the body at the reference position.

The most common configuration of a spindle using this principle of magnetic bearing to control the position of the rotor shaft, is depicted in Figure 1.6. At the top side of the rotor two magnetic bearings control the position in the X and Y plane; similarly two magnetic bearings control the X and Y position of the rotor at the bottom side. An axial bearing (not shown in Figure 1.6) is added to control the position of the rotor in the Z direction.

A main limitation of the applicability of AMB spindles for conventional milling is that the maximum sustainable bearing forces are relatively small compared to rolling element bearings [90]. However, this limitation will be less of a problem when applying AMB spindles for micro-milling, as the cutting forces in micro-milling are small. Moreover, active magnetic bearing spindles combine a number of favorable features that make them in fact particularly attractive for application to the micro-milling process:
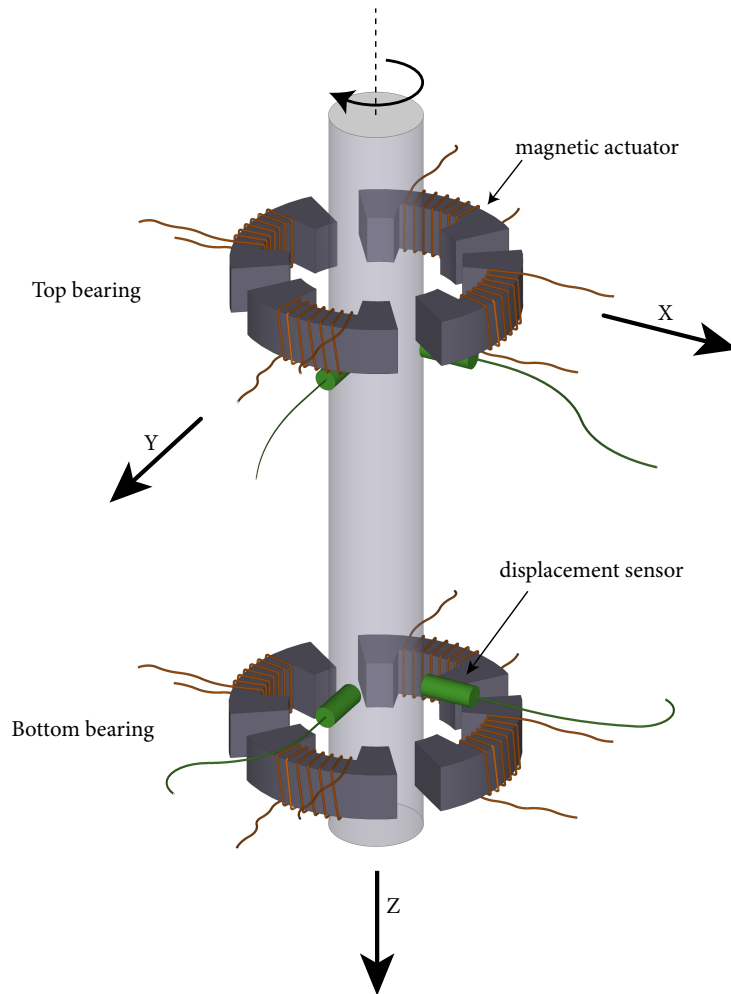
Figure 1.6: Standard configuration of a spindle with Active Magnetic Bearings: radial magnetic bearings are provided in x and y direction at the top and bottom. Additionally an axial magnetic bearing is added to constrain the motion in the z-direction (not shown).

1. The maximum achievable surface speed of magnetic bearings is much greater than of rolling element bearings. Rotational speeds up to the limit of material strength are possible. There is virtually no wear and no need for lubrication.

2. Because they are non-contacting, magnetic bearings are less prone to thermal problems as conventional bearings at high speeds.

3. The active nature opens a range of monitoring and control possibilities, a feature that is unique for magnetic bearings.

It is the last feature that forms the point of departure for this thesis. Hence, in Section 1.4.1 we will elaborate on this concept and discuss the potential applications. Subsequently in Section 1.4.2, we will discuss the available results in the literature in which the active character of AMB setups is used for monitoring and control.

1.4.1 Process monitoring and control of micro-milling with AMB spindles

The active nature of the AMB spindles can be exploited to improve the milling process. The advantage of this is that no additional sensors need to be added to the machine, but instead usage can be made of signals that are already available. Improvements are possible both on the level of contour control as well as the level of process monitoring and control. The underlying mechanisms at both levels can be described as follows:

Contour control by positioning of rotor shaft A unique feature of magnetic bearings is that there is some freedom in the positioning of the rotor shaft inside the air gaps. Typically AMB spindle controllers are designed by setting a fixed reference position for the rotor shaft at the location of each bearing. This reference position is mostly in the center of the air gap. Deviations from this are possible, which can be exploited to improve the contouring control:

Runout compensation Referring to the description of runout in Section 1.2.1, runout can be considered a periodic displacement of the tool tip. Compensation of such positioning errors is in principle possible by generating periodic forces, leading to a displacement of the rotor in opposite direction to the runout errors.

Tool bending compensation The rotor shaft can be displaced inside the airgaps to compensate for errors due to bending of the tool under cutting forces. This can be translated to the control objective to minimize the displacement of the tool tip under disturbance forces acting on the tool tip (as opposed to the abovementioned standard objective to minimize the displacement of the rotor shaft at the bearings).

Dual actuator As explained in Section 1.2.1, a consequence of the high rotational speeds is that milling machines also need to be fast. Higher control bandwidths of the relative positioning of the tool and workpiece can be achieved by combining the control of the main machine axes and the magnetic bearings in a dual actuator concept. The main idea here is that the positioning error of the slower machine axis (with a large stroke) is compensated by the positioning of the faster magnetic bearing (with a small stroke).

Process monitoring/control through interaction with the cutting process From the perspective of the AMB spindle, the interaction between tool and workpiece causes disturbance forces acting on the AMB spindle system. Ideally, the AMB controller is designed such that the response of the AMB spindle system to these disturbances is small. From this, it follows that the current and position signals contain filtered versions of the disturbance forces, where the filters are determined by the dynamics of the AMB spindle system and the AMB controller. This concept can be exploited for estimation of the cutting forces. Furthermore, fault monitoring (e.g. breakage detection and collision detection) is possible by monitoring the disturbance for abnormal changes. Morever, when it is known how the disturbance varies with changing process conditions such as progressing tool wear, condition monitoring becomes possible as well.

The main emphasis in this thesis is on the second mechanism, as our focus is on process monitoring and control of micro-milling. Nevertheless, the aforementioned opportunities for improvement of the micro-milling process at the level of contour control are worthwhile to explore further (see recommendations for further research in Chapter 8).

1.4.2 Review of results of process monitoring and control with active magnetic bearings

The idea that the active nature of a magnetic bearing setup provides a suitable starting point for developing the monitoring and control techniques, was recognized more than 20 years ago [13, 26]. Some results are available, which will be discussed in this section. To give a comprehensive overview, the scope is widened to any kind of rotating machinery having magnetic bearings, which apart from spindles includes pumps and flywheel systems.

In his thesis, Müller [116] explored the application of a high-speed AMB milling spindle for various monitoring techniques. A static approach to cutting force estimation approach was formulated that essentially ignores all transient effects as well as the effect of measurement noise. Cutting forces are estimated by first calculating the bearing forces from the current signals using a simplified analytical model of the

magnetic actuator, and subsequently using the lever law to translate these bearing forces to the forces acting on the tool tip (essentially a static model of the rotor-tool dynamics). Performance of this approach is poor, which Müller attributes to non-linear effects, thermal effects and the dependency of the currents on the angular position of the rotor. Furthermore, tool breakage detection and collision detection was proposed. The proposed methods are not model-based, but are based on detecting abnormal changes in the bearing signals, for which linear filters are designed. A model-free approach to tool wear monitoring was explored by looking for features in the estimated force signals that are independent of the cutting conditions, but this approach was reported unsuccessful.

A recent example of cutting force estimation using AMB spindles has been presented by Auchet and coworkers [16]. In this approach, a non-parametric frequency response of the transfer from the cutting force on the tooltip to the command voltage levels of the current amplifier[4] is estimated through impact testing. The cutting force estimation problem is solved in the frequency domain by computing the pseudo-inverse of the estimated frequency response and multiplying this with the Fourier transform of the measured command voltage levels. By windowing and inverse transformation, estimates of the cutting force in the time-domain are obtained. We observe that in this approach no noise modeling is performed, nor is the information in the displacement signal used. Still, the results are in good agreement with force measurements from a dynamometer platform for low frequencies. Moreover, the presented method allowed cutting force measurements with a bandwidth that exceeded that of the dynamometer by more than 20 times.

A failure detection system for a AMB spindle has been reported by Chevrier et al. [34]. Position signals and bearing forces (presumably derived from the current signals) in the bearings are constantly monitored. If any value is beyond a design threshold, an alarm signal is given to machine controller, causing the the machine and spindle to be stopped immediately.

Aenis [4] performs condition monitoring of a pump system using AMBs. Model-based diagnosis is proposed where the closed-loop frequency reponse function of the AMBs is modeled with and without faults. Fault diagnosis is done by comparing these modeled frequency responses with experimentally identified frequency responses of the system.

We conclude this overview with some results in which the active character of AMB spindles is employed to suppress chatter. As discussed in Section 1.2.1, chatter is the result of self-exciting vibrations between the tool and the workpiece. Kyung and Lee [93] formulate guidelines for selecting parameters of a PID controller to obtain chatter free cutting. Chen and Knospe [32] have studied the possibility to actively suppress machining chatter via AMB control design for the case of turning. A robust control strategy is followed to design controllers that guarantee that

---

[4]Note that in many AMB setups, the type of amplifier that is used is a *transconductance amplifier*, i.e. an voltage-controlled current source with a fixed gain.

regenerative vibrations cannot arise for three different cases, i.e. (i) at any rotational speed, (ii) at a given rotational speed, (iii) at any speed in a given rotational speed interval. As part of the control design, the maximum cutting depth is established such, that the control objective in each of the three cases is satisfied. In his thesis, van Dijk [166] gives two chatter control strategies using AMB control design. The first ensures robust chatter-free milling by automatic adaptation of spindle speed and feed. The objective of the second strategy is similar to the approach by Chen and Knospe, but then for the more complicated milling process.

1.4.3 Conclusion

Spindles with Active Magnetic Bearings are particularly interesting for the micro-milling process, not only for the achievable spindle speeds, but also because of the opportunities they offer to develop online process monitoring and control techniques. These include force monitoring, tool condition and breakage monitoring, and chatter control. However, even in conventional manufacturing literature, the available results implementing these techniques are quite limited. Model-based approaches are even scarser and to the best of the author's knowledge, no results are available that are tailored to the challenges of micro-milling.

1.5 Research goal and approach

At this point we are ready to formulate the research goal of this thesis. Indeed, in the previous sections we have highlighted the main challenges of micro-milling and have pointed out that process monitoring and control becomes of increasing importance when the tool diameter is reduced. It has been discussed that AMB spindles are an attractive technology for the micro-milling process, given the high attainable rotational speeds. We have reviewed prior research, demonstrating that the active nature of AMB setups can be utilized for process monitoring and control. However, it has been concluded that literature currently lacks results on monitoring and control using AMB spindles that are particularly focussed on the process of micro-milling. Given the above, the main goal of the thesis is:

Research goal

Investigate the opportunities for model-based process monitoring and control to improve the micro-milling process using the intrinsic properties of AMB spindles.

We will elaborate further on two aspects of this research goal. First and foremost, the model-based approach plays a central theme in this thesis. In this ap-

proach, prior knowledge of the dynamics of the AMB spindle setup, the cutting process, as well as of the disturbances in the system, are formulated as explicit dynamic models, which are used for synthesis of observers and controllers that accomplish the process monitoring and control objectives. Including such prior knowledge will allow for obtaining high reliability and performance of the resulting solutions. This is of high importance to the micro-milling process, as was reviewed in Section 1.3.2, and provides the justification for such approach.

Second, the objective to investigate the opportunities lies in the fact that it is not a priori evident that AMB spindles can be used successfully for development of process monitoring techniques for the micro-milling process. Given the low amplitude and high signal bandwidth of cutting forces in micro-milling, one can ask if monitoring based on the AMB signals is feasibile in the first place. Similarly, the question can be raised if any of the potential control applications as reviewed in Section 1.4.1 will have sufficient performance, given the limitations imposed by the physics of the AMB setup.

The approach towards this goal is summarized as follows:

A. Provide an approach to model-based cutting force estimation in micro-milling using the signals of the AMBs;

B. Provide a method for system identification of a high speed AMB micro-milling spindle;

C. Provide a compensation approach for the almost periodic disturbances in the bearing signals resulting from unbalance forces and roundness errors.

These steps are clarified as follows:

A. By modeling the dynamics of the closed-loop AMB system in a micro-milling configuration, a framework can be drawn that allows to formulate AMB-based process monitoring and control objectives as model-based observer or controller synthesis problems. In this thesis, we limit our attention to one of these problems in particular, i.e. the problem to estimate the cutting forces from the bearing signals. Indeed, as discussed in the preceding sections, the cutting forces are among the key physical quantities for development of process monitoring and control techniques. This is true in macro-scale, as well as in micro-scale cutting. We will take the position that the feasibility of process monitoring and control using AMB spindles depends on the ability to estimate cutting forces from the bearing signals with sufficient bandwidth and accuracy. AMB spindle based process monitoring, like tool condition monitoring, as well as process control, such as force control, can be considered an extention of cutting force estimation using AMB spindles. For this reason, the main emphasis in this thesis is on estimation of the cutting forces, and not on further application of the cutting force signal for more advanced

monitoring and control tasks. Reviewing the available results on (cutting) force estimation with AMB spindles, we conclude that some attempts have been made, but that no comprehensive model-based procedure is available, which solves the cutting force estimation problem while dealing with

- the dynamics of the AMB spindle, including its multivariable character and the increased coupling due to gyroscopy at higher rotational speeds;
- the closed-loop nature of the setup; and
- information on the spectral content of the cutting force signal and the noise disturbance on the bearing signals.

To solve this problem, an approach is developed for model-based optimal estimation of unknown inputs to multivariable closed-loop systems.

B. Obviously, the performance of any model-based cutting force estimator depends on the accuracy of the model of the dynamics of the AMB spindle system. Such model of the AMB spindle could be derived from first principles. However, given the complexity of the system and required accuracy of the model, the modeling efforts will rapidly increase. Considering this, while in addition several physical phenomena can at best be modeled in approximation (such as material dampling and eddy current effects), an attractive alternative approach is to perform modeling of the AMB spindle system from measured data sequences. Although a number of results are available on AMB spindle identification, literature is currently lacking an approach in which:

- careful experiment design is done, considering the intrinsic nonlinear nature of the AMB spindle system;
- black-box models are estimated of the dynamics of the full high-order multivariable AMB system operating at high rotational speeds, including both its rigid body, as well as its flexible body behavior.
- the transfer function from the force on the tip of a micro-mill to the displacement of the rotor at the bearings is also modeled.

In this thesis, a system identification approach is formulated that comprises these aspects.

C. The third and last step pertains to a practical obstacle to implementing any monitoring approach: the effects of mass unbalance and roundness errors of the rotor. These result in disturbances in the current and position measurements which are (almost) periodic. Proper compensation for these disturbance in the measured signals, prior to using them for monitoring tasks, is necessary to avoid making monitoring errors. In this thesis, we give an efficient approach to this.

1.6   Outline of the thesis

In this thesis, the results are presented of the research to the main problem given in the previous section. Apart from the introduction and the conclusion, this thesis comprises six main chapters:

Chapter 2 This chapter provides the foundation for the rest of the thesis. It will be discussed how modeling of the AMB spindle setup, as well as the cutting process, results in configuration with two closed-loops, i.e. the AMB control loop and the cutting process loop. From analysis of both closed-loops, as well the various disturbance sources, conclusions will be drawn on the feasibility of performing monitoring and control tasks within this setup. The cutting force estimation problem is studied in more detail and it is discussed how this problem can be treated as an input estimation problem.

Chapter 3 In this chapter, an approach will be given to solve the cutting force estimation problem by optimal estimation of unknown inputs to discrete-time multivariable systems from closed-loop data. A particular challenge in this problem is that controller knowledge might not be available. An approach is formulated by imposing an extra constraint to the optimal estimator, ensuring equal performance of it for any controller. Using Wiener filter theory and spectral factorization, solutions are formulated, where particular attention is given to the instablity of the plant dynamics. Smoothed estimators are derived allowing to obtain smaller estimation errors when a delay is tolerable.

Chapters 4, 5 and 6 deal with the problem of identifying the AMB dynamics from measured data sequences. In Chapter 4 we first give an algorithm for estimation of multi-input multi-outout (MIMO) Output Error (OE) models in matrix fraction description from frequency domain data. Subsequently, we treat the identification of the AMB spindle in two steps. In Chapter 5, identification of the bearing dynamics is covered, in which the results of Chapter 4 are applied. Identification of the tooltip dynamics is covered in Chapter 6. In more detail, the content of these three chapters is as follows:

Chapter 4 This chapter presents an iterative linear regression algorithm for identifcation of MIMO output error models from frequency domain data. A well-known method to approach the output error minimum by iterative linear regression steps has been formulated by Sanathanan and Koerner. A disadvantage of this approach is that in general convergence of the iterations only implies optimality under restrictive conditions. In literature, an alternative iterative linear regression procedure is available, which ensures optimality upon convergence, also in case of undermodeling. This algorithm is known for time-domain identification as the Simplified Refined Instrumental Variable method (SRIV), and was recently formulated for frequency domain

identification of SISO output error models. Here we generalize this formulation to MIMO identification of models in matrix fraction description.

Chapter 5 In this chapter we consider the problem of obtaining accurate models of the bearing dynamics of the AMB spindle in the frequency range relevant for process monitoring and control. The complexity of this problem is found in the instability of the dynamics, its high order and multivariable character, and the parametrically dependency of the dynamics on the rotational speed, as well as in the presence of nonlinearities in the system. A frequency domain approach is taken, consisting of two stages. In the first stage, accurate estimates of the multivariable frequency response function (FRF) of the bearing dynamics are made by excitation of the closed-loop with orthogonal random phase multisine signals. A method to detection of nonlinear distortions in the multivariabele FRF estimate is discussed, which enhances the variance analysis for detection of nonlinearities of Pintelon et al. in [131]. In the second stage, a parametric model of the AMB spindle is estimated using the estimated FRF and its covariance. A Schur-weighted Output Error criterion is formulated, and minimization of the criterion function is achieved by application of the IV-based iterative linear regression method of Chapter 4.

Chapter 6 In this chapter, identification of the tooltip dynamics is addressed. The main challenge in this identification problem is to apply a known excitation force to the tooltip. The route followed in this chapter is to identify the tooltip dynamics using data obtained during a milling experiment in which the cutting forces are measured. The amount of data that can be generated in this way is limited, as is the control over the spectral properties of the input. Hence, in order to reduce the complexity of the identification, usage is made of the observability and controllability properties of the system as derived in Chapter 2. This results in a particular closed-loop parameterization of the model-set and a known, but non-minimum phase noise model. For this particular identification problem, solutions are discussed.

The last chapter covers the final subgoal:

Chapter 7 This chapter considers model-based correction of runout disturbances in measurements of the positions and currents of AMB spindle. Such disturbances are synchronous with the rotation of the spindle and hence almost periodic. A parametrized truncated Fourier series expansion model for the runout disturbance as a function of the angular position is used, allowing to formulate runout identification as a parameter estimation problem. In correcting for the runout disturbances, the main issue is how to deal with the uncertainty in the angular position measurements, or the total lack of such measurements. Solutions are given that compensate for the errors introduced by this uncertainty, or estimate the the angular position from the available data using an Extended Kalman filter approach.

Chapter 2

A modeling framework for monitoring and control with AMB spindles

2.1 Introduction

In this thesis a model-based approach to process monitoring and control of micro-milling using spindles with Active Magnetic Bearings is pursued. The objective of this chapter is to lay down the foundation by providing a modeling framework, and by deriving and analyzing the necessary models. To this end, we will first discuss the default configuration of an AMB spindle in a micro-milling setup as considered in this thesis (Section 2.2). From this, a first principles model of the system is given, where we will restrict attention to a description of the dynamics of the system in radial direction, as these dynamics are the most relevant for our problem[1]. This model is obtained by first modeling the AMB spindle dynamics in Section 2.3, and subsequently modeling the cutting dynamics in Section 2.4. These submodels are integrated in Section 2.5. The resulting modeling framework enables us to analyze the feasibility of monitoring and control of micro-milling in greater detail and formulate monitoring and control objectives as model-based observer and control synthesis problems. As laid out in Chapter 1, the cutting force estimation problem is given emphasis in this thesis, and this problem is further analyzed in Section 2.6. A difficulty in solving this problem is that it requires knowledge of both the AMB spindle submodel and the cutting dynamics submodel. As will become clear in Section 2.4, obtaining accurate models of the latter is particularly difficult. Hence, the approach taken in Section 2.6 is to make a reasonable simplification that allows to

---

[1]Particularly when milling with square end-mills, the radial cutting forces dominate over the axial force.

reduce the force estimation problem to an input estimation problem. In the input estimation problem, the cutting forces are an unknown input to the closed-loop AMB spindle system, and in solving this problem, no model knowledge of the cutting dynamics is required. The analysis in Sections 2.2 to 2.6 is done in continuous time. However, in practice we will deal with discrete-time data. Hence, in the final section of this chapter, Section 2.7, equivalent discrete-time models are given. From the physics of the AMB system, properties concerning the observability and controllability of the resulting discrete-time models are derived, which will prove to be useful in later chapters when addressing the input estimation problem and the problem of identifying the dynamics of the setup from measured data sequences.

2.2 Configuration of the AMB spindle setup

The default configuration of a micro-milling setup with an AMB spindle considered in this thesis, is depicted in Figure 2.1. In this section we will discuss the main components of this setup. We start in Section 2.2.1 with a discussion of the AMB spindle hardware, consisting of the rotor with toolholder and tool, and the magnetic bearings. Subsequently in Section 2.2.2, some key properties of the AMB controller and the power amplifier are reviewed. Finally, in Section 2.2.3 we will comment on the feed drives needed to create the relative motion between the tooltip and the workpiece.

2.2.1 AMB spindle hardware

The results in this thesis apply to AMB spindles having a standard configuration as is described in this section. Such AMB spindles have a design with four radial bearings and one axial bearing. By convention, the Z-axis of the AMB spindle coordinate frame is aligned with the rotor axis in the direction of the milling tool, and the X- and Y-axis coincide with the directions of the radial bearings. Each of the bearings consists of two magnetic actuators positioned at opposite side of the rotor, which are configured in differential driving mode. The key principle of this operation mode is that the net force exerted by the set of the two magnetic actuators on the rotor can be controlled by varying a single control current, causing the magnetic flux to be simultanously increased in one actuator and decreased in the other. The exact properties of the magnetic actuators depend on their structural design, where eight-pole and homopolar configurations are the most common types [148]. The displacement sensor of each radial bearing measures the position of the rotor at, or close to, the force actuation point of the magnetic actuator.

The micro-milling tool is attached to the rotor of the AMB spindle, by means of an appropriate interface and a high-speed motor provides the rotation of the rotor/tool combination.
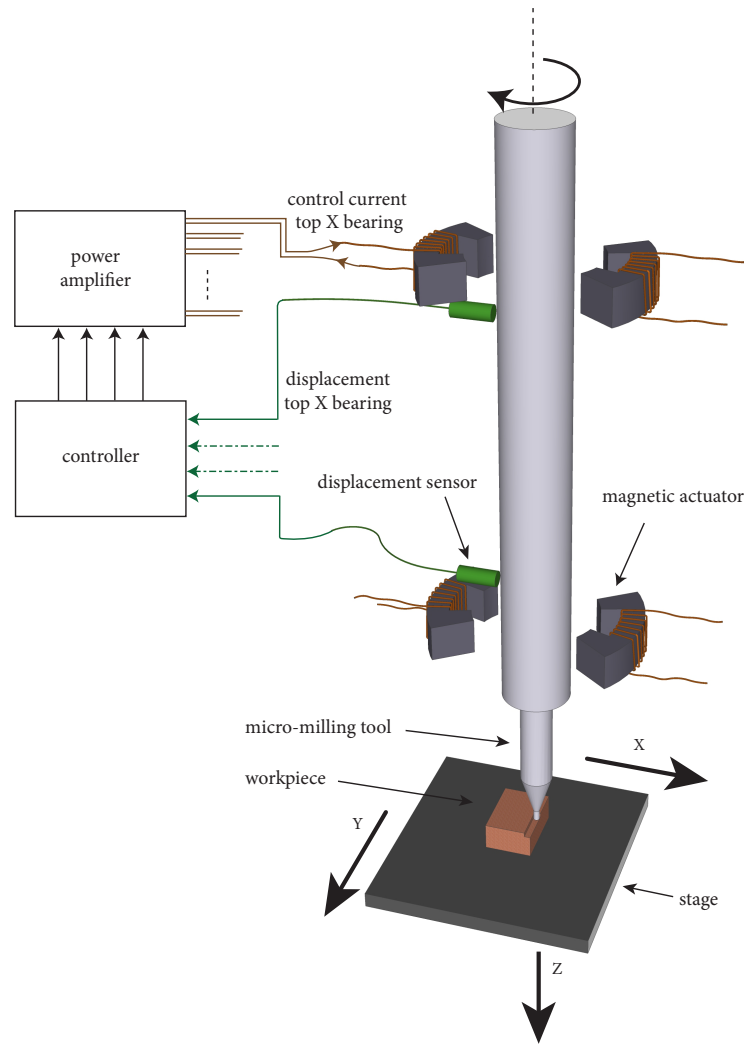
Figure 2.1: Schematic of the default configuration of an AMB spindle used for micro-milling (only the bearings in X direction are shown). The considered configuration shows the closed-loop AMB setup consisting of magnetic actuators, displacement sensors, controller and power amplifier. The micro-milling tool is attached to the rotor. Micro-milling is performed by relative displacement of the tool and a workpiece in X, Y and Z direction.

2.2.2 AMB controller and power amplifier

A magnetic bearing setup is unstable in general. Stable positioning of the rotor is obtained by closed-loop control. In the setting of this research we will restrict attention to setups with linear controllers operating in current control, meaning that the output of the controller defines the level of the control current[2]. Configurations with analog and digital controllers are considered.

The current through the actuators is supplied by power amplifiers. For this purpose, the coil current is measured and compared to the value commanded by the AMB controller. A controller adjusts the output voltage to obtain the desired coil current. Limitations on the output voltage, as well as other internal dynamics, imply that power amplifiers cannot be considered a static gain, but have dynamical behavior and a certain bandwidth, which is also determined by the reluctance and resistance of the windings of the coils.

2.2.3 Feed drives

In order to perform a micro-milling task, the workpiece and the tooltip need to perform a relative motion. For this, the micro-milling machine is equiped with feed drives, which operate under servo control to achieve the required coordinated motion of the axes. Although the dynamics of the cutting process and the dynamics of the controlled servos are directly coupled, in the context of this research we will not include this coupling, and consequently neither consider the design, nor the dynamics of the feed drive systems. We will justify this in more detail in Section 2.4 when discussing the milling dynamics.

2.3 Modeling of AMB spindle systems

In the literature dealing with the problem of designing controllers for AMB spindles, the topic of first principles modeling of AMB spindles has received much attention. In this context, models are often derived under some simplifying assumptions. The following assumptions are frequently made in the literature [148, 149]:

- the rotor of the AMB system is a fully rigid body, meaning that bending mode vibrations of the system are considered to be outside the operating range, and are hence ignored (often robustness against these resonances can simply be provided by adding high frequency roll-off to the controller);

- there is no coupling between the lateral motions of the rotor, meaning that the gyroscopic effects are ignored and that the dynamics in X and Y directions are treated as decoupled systems;

---

[2] We remark that the methods in thesis are easily adapted for AMB setups that are operated with voltage control.
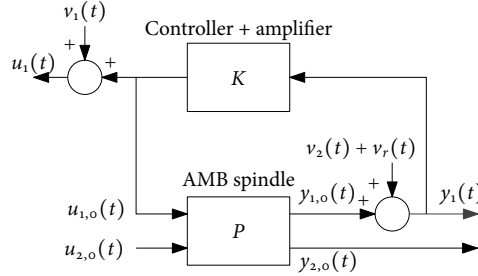
Figure 2.2: Block diagram of the AMB spindle system, with P representing the dynamics of the AMB spindle, and K the controller and amplifier. System P has as inputs the control currents through the bearings $u_{1,0}$ and the cutting forces $u_{2,0}$ (with $\dim(u_{1,0}) = 4$ and $\dim(u_{2,0}) = 2$). The outputs are the displacement of the rotor shaft at the bearings $y_{1,0}$ and the displacement of the tooltip $y_{2,0}$ (with $\dim(y_{1,0}) = 4$ and $\dim(y_{2,0}) = 2$).

- the actuator force depends linearly on the rotor position and the control current, while this relationship is intrinsically nonlinear. Most AMB applications use a linearization of this relationship in the operating point.

In this thesis we will not adopt the first two assumptions, i.e. in modeling the AMB spindle system, the effects of high speed rotation are taken into account, as is the flexibility of the rotor shaft. In micro-milling in general high rotational speeds are used, implying that the effects of gyroscopy are not negligible. Furthermore, particularly when considering monitoring applications, the frequency range of interest is wider than the frequency range that is relevant for control design alone, implying that the bending behavior of the system cannot be ignored. We will however make the third assumption, although the effect of the nonlinearities in the system is analyzed in Chapter 5 when dealing with the problem of identifying the AMB spindle system.

This section contains a method for first principles modeling of an AMB spindle system of the default configuration described in the previous section, resulting in a linear model incorporating the abovementioned phenomena. A key difference in this approach compared to the AMB literature, is that this approach does not only cover the AMB bearing dynamics, but also the tooltip dynamics, see Figure 2.2. This means that the model of the AMB spindle system P has as inputs the control currents through the bearings $u_{1,0}$, as well as the cutting force acting on the tool tip $u_{2,0}$. The outputs are the displacement of the shaft at the radial bearings $y_{1,0}$, and the displacement of the tooltip $y_{2,0}$. The AMB spindle model P can be obtained from two submodels, i.e. a model of the rotordynamics of the rotor with tool, and a model of the electromagnetic actuators. In the following two sections we will
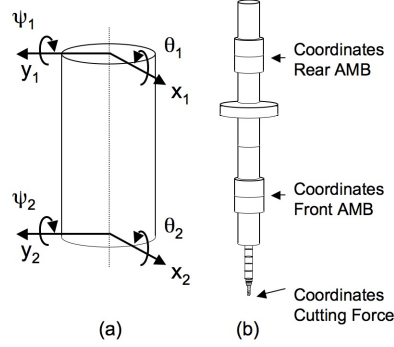
Figure 2.3: Finite Element model of the rotor system using cylinder elements. (a) The dynamics of the cylinders can be expressed using a set of eight coordinates describing the position and orientation of the top and bottom plane; (b) The rotor system is model as a series of cylinder elements with varying diameter.

cover how these submodels can be obtained, and subsequently in Section 2.3.3 we treat how these models can be combined to obtain a model of the AMB spindle. In Section 2.3.4 we discuss the disturbance sources in AMB setups. Finally, in Section 2.3.5 the modeling procedure is illustrated by showing the resulting model for the AMB spindle setup that was built for this research.

2.3.1 Dynamical model of the rotor and tool

The rotor, toolholder and tool can jointly be considered as a flexible beam system that rotates along its symmetry axis and exhibits bending vibrations. Flexible systems are in principal infinite dimensional, so that in practice finite element approximations are made to model the dynamics. For the rotor, a common approach to do this is by modeling the entire beam as a series of rigidly or flexibly connected Timoshenko cylinder elements (see Figure 2.3)[91, 94]. The material properties of such cylinder elements are defined by the Elasticity modulus, Mass density and the Poisson ratio. With this, the dynamics of the elements are described by a set of coordinates, representing the position and orientation of the top and bottom plane of the cylinder. Let x represent the vector that is obtained by stacking the coordinates of all cylinders, and Ω the rotational speed of the rotor, which is assumed to be known and constant. As is shown in Appendix A.1, the equation of motion of the rotor is then given by

$$M\ddot{x} + (D + \Omega G)\dot{x} + Kx = F. \tag{2.1}$$

The dimension of x depends on the number of cylinder elements that are used in the finite element approximation. Let this number be the denoted by N_{beam}. Then $n_x := \dim(x) = 4N_{\text{beam}} + 4$ (see Appendix A.1). Matrices M, D, G, K are all square with dimension n_x. In particular $M = M^T$ represents the mass matrix, $D = D^T$ the damping matrix, $K = K^T$ the stiffness matrix, and $G = -G^T$ the skew-symmetrical gyroscopic matrix. Matrix D represents the material damping, which is difficult to model from first principles. It is common to apply Rayleigh damping, or modal damping [91]. Rayleigh damping involves setting $D = \alpha M + \beta K$, where α and β are estimated constants. To apply modal damping first a modal decomposition of the undamped system (for $\Omega = 0$) is determined. Then for each mode a damping value is selected. With this, matrix D can be computed. The vector $F \in \mathbb{R}^{n_x}$ is the vector of generalized external forces associated with the coordinates of the rotor.

2.3.2 Model of the electromagnetic actuators

Each of the electromagnetic actuators exerts a force on the rotor. The second step in modeling the AMB spindle dynamics is to obtain the relation between this force, the control current through the coil, and the position of the rotor at the location of the bearing. In general this relation is nonlinear, where the exact form depends on the structural design of the actuator. However, often this relationship reduces to a linear form when the actuator is configured in differential driving mode and the rotor is exactly at its center position. Hence, it is quite common to approximate the behavior of the actuator by a first order linearization. As is shown in Appendix A.2, this yields

$$F_{\text{amb},j} = k_{x,j} x_{\text{amb},j} + k_{i,j} i_j \tag{2.2}$$

where j indicates the bearing (i.e. $j = 1$ correspondes to the bottom X bearing, $j = 2$ to bottom Y, $j = 3$ to top X, and $j = 4$ to top Y), $x_{\text{amb},j}$ the j-th bearing coordinate (i.e. the difference of the air gap length of bearing j compared to its nominal value), and i_j the control current of bearing j. The constant $k_{x,j}$ is referred to as the negative stiffness of the bearing, and the constant $k_{i,j}$ as the actuator gain. Both constants depend on the structural design of the bearings, and relevant properties are the number of windings and pole surface area of the coils, the bias current, and the nominal gap length (see Appendix A.2). Stacking the forces of all bearings in the vector $F_{\text{amb}} = \text{col}(F_{\text{amb},1}, F_{\text{amb},2}, F_{\text{amb},3}, F_{\text{amb},4})$ we obtain

$$F_{\text{amb}} = K_x x_{\text{amb}} + K_i u_{1,0} \tag{2.3}$$

with

$$u_{1,0} = \text{col}(i_1, i_2, i_3, i_4) \qquad\qquad K_x = \text{diag}(k_{x,1}, k_{x,2}, k_{x,3}, k_{x,4})$$
$$x_{\text{amb}} = \text{col}(x_{\text{amb},1}, x_{\text{amb},2}, x_{\text{amb},3}, x_{\text{amb},4}) \qquad K_i = \text{diag}(k_{i,1}, k_{i,2}, k_{i,3}, k_{i,4}).$$

Observe that the bearing coordinates are linearly related to the position of the rotor. Hence, let $B \in \mathbb{R}^{n_x \times 4}$ be a linear coordinate mapping such that $x_{\text{amb}} = B^T x$. Then we obtain

$$F_{\text{amb}} = K_x B^T x + K_i u_{1,0}, \tag{2.4}$$

which is the sought relationship between the actuator force, rotor position and control currents. Note that the obtained model is a static relation.

2.3.3 Combining the submodels

In order to obtain a model of the spindle, let us define $C_1 \in \mathbb{R}^{4 \times n_x}$ such that $C_1 x$ describes the displacement of the rotor at the location of the displacement sensors, $C_2 \in \mathbb{R}^{2 \times n_x}$ such that $C_2 x$ describes the displacement of the tool tip, and $B_2 \in \mathbb{R}^{n_x \times 2}$ a matrix that maps tool tip coordinates to the rotor coordinates (note that $B_2 = C_2^T$). Then combining (2.1) and (2.4) yields the set of equations

$$M\ddot{x} + (D + \Omega G)\dot{x} + (K - BK_x B^T)x = BK_i u_{1,0} + B_2 u_{2,0} \tag{2.5a}$$

$$y_{1,0} = C_1 x \tag{2.5b}$$

$$y_{2,0} = C_2 x \tag{2.5c}$$

where $u_{2,0}$ is the 2×1 vector containing the cutting forces in X- and Y- direction. This model can be reformulated in a state space representation as follows:

$$\dot{\xi} = \mathcal{A}\xi + \mathcal{B}_1 u_{1,0} + \mathcal{B}_2 u_{2,0} \tag{2.6a}$$

$$y_{1,0} = \mathcal{C}_1 \xi \tag{2.6b}$$

$$y_{2,0} = \mathcal{C}_2 \xi \tag{2.6c}$$

with $\xi = \text{col}(x^T, \dot{x}^T)^T$ and

$$A = \begin{bmatrix} 0 & I \\ -M^{-1}(K - BK_x B^T x) & -M^{-1}(D + \Omega G) \end{bmatrix}$$

$$\mathcal{B}_1 = \begin{bmatrix} 0 \\ M^{-1}BK_i \end{bmatrix}, \qquad \mathcal{B}_2 = \begin{bmatrix} 0 \\ M^{-1}B_2 \end{bmatrix}$$

$$\mathcal{C}_1 = \begin{bmatrix} C_1 & 0 \end{bmatrix}, \qquad \mathcal{C}_2 = \begin{bmatrix} C_2 & 0 \end{bmatrix}$$

Depending on the number of elements that are used to model the rotor system, this typically leads to high order systems. Appropriate model reduction can yield lower order models that include only the relevant modes of the rotor. We have now obtained a linear model of the setup, of which we can express the transfer function P as:

$$\begin{pmatrix} y_{1,0} \\ y_{2,0} \end{pmatrix} = \begin{bmatrix} P_{11}(s) & P_{12}(s) \\ P_{21}(s) & P_{22}(s) \end{bmatrix} \begin{pmatrix} u_{1,0} \\ u_{2,0} \end{pmatrix} \tag{2.7}$$

We discuss some key characteristics of the dynamics of AMB spindles:

Pole locations and rotational speed dependency Note that if the rotor is at stand-still, i.e. $\Omega = 0$, the dynamics of the plant in X- and Y-direction are decoupled and two separate systems describing the dynamics in each of these planes can be derived. Typically the dynamics in each plane then combine a set of real poles (both stable and unstable) related to the rigid body modes, and a set of stable poles close to the imaginary axis pertaining to the flexible modes of the rotor. When the rotational speed increases, two phenomena can be observed due to the effects of gyroscopy [94, 91]:

- A coupling between the dynamics in X- and Y-direction increasing with the rotational speed. The real modes become complex.

- Flexible mode resonances splitting into two frequencies, centered around the mode frequency at stand-still. The frequency separation increases with increasing rotational speed.

Non-linear effects In the previous section it was mentioned already that the relationship between actuator force, control currents and displacements of the rotor is nonlinear by nature. In reality, this is not the only nonlinearity. Other nonlinear effects include Eddy-current losses and the nonlinear permeability of ferromagnetic materials used for the cores. Moreover, the permeability of these materials also depends on the history of the material's magnetization, which leads to the well known-magnetic hysteresis effect [149]. The implication of this is that the linear model that we use is valid only in the chosen operating point and for small signals. We will return to the consequences of this when addressing the problem of identifying a linear model of the AMB spindle from measured data (Chapter 5).

2.3.4   Disturbance sources

There are several sources of disturbances in the AMB spindle system. These include noise-like disturbances in the mechanical and electrical parts of the system, as well as measurement disturbances. It is assumed in this context that all these disturbances can be modeled as additive stochastic noise processes, where we denote v_1 the noise on the measurements of the control currents and v_2 the noise on the displacement measurements (see Figure 2.2).

In an AMB spindle system in general a specific non-stochastic disturbance is also experienced, which is caused by runout. The phenomenon of runout can be described as follows. The displacement of the rotating shaft is measured at fixed locations in the spindle housing. Consequently, any roundness errors or eccentricity of the shaft will appear as varying displacements when the shaft is rotating. Obviously these variations will be periodic in nature with a fundamental frequency

2. A MODELING FRAMEWORK FOR MONITORING AND CONTROL WITH AMB SPINDLES

equal to the rotational frequency of the spindle[3]. Zooming in on the causes of these roundness errors and eccentricity we note the following:

Roundness errors Due to limitations in the fabrication process, rotor shafts will generally have a cross section that deviates from a perfect circular profile. Such deviations are referred to as roundness errors. Evidently when the shaft is rotated, these deviations will appear in the displacement signal and the resulting pattern will be periodic with the rotation of the shaft.

Eccentricity It is easily verified that a varying displacement signal will be obtained when the shaft is not rotating around its geometrical center, but some other axis. This can simply be caused by misalignment, but another important cause of eccentricity in rotor systems is mass unbalance, which like roundness errors, is caused by limitations in the fabrication process.

Mass unbalance occurs when the center of mass deviates from the geometrical center, or, in terms of the 3D shaft, when the geometrical center line does not coincide with one of the principal axes of rotation. As a result of such mass unbalance, the rotor will experience a centrifugal force, of which the amplitude is proportional to the square of the rotational frequency [91]. As a result of this rotational frequency dependency, also the response of the AMB spindle system to this mass unbalance forces depends on the rotational frequency. At low rotational speeds these forces are small, and they will not lead to a significant eccentricity, meaning that the rotor will rotate around its geometrical center. When increasing the rotational speed, it may pass through the eigenfrequencies of the rotor-bearing system (known as the the critical frequencies). The mass unbalance forces will excite these modes and depending on the damping of the modes, high eccentricities may be observed. At high rotational speeds, the centrifugal forces will cause the rotor to rotate around its center of mass, again resulting in an eccentric motion.

Another possible source of eccentriticy arises when the motor has a negative stiffness that varies with the angular position of the rotor, which is also referred to as magnetic unbalance. As a result the rotor experiences a varying force while rotating, resulting in a periodic displacement synchronous with the rotation of the rotor.

For our purposes, runout is modeled as an additive disturbance source on the position measurements[4], here denoted as v_r (see also Figure 2.2). The reasoning behind

---

[3] As these displacement variations are periodic with the rotation of the shaft, they are referred to as *repeatable runout* or *synchronous runout*.

[4] Note that we have focused on how runout results in periodic variations in the displacement signals of the magnetic bearings. We add that another effect of runout is that the bearing forces of the actuators will also be affected as a result of the resulting variations of the gap length of the bearings. However, as in general these variations are small in comparison to the gap length, the effect on the bearing forces can be neglected.
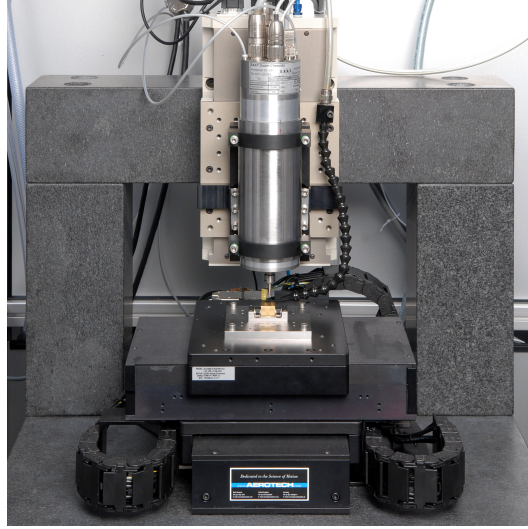
Figure 2.4: Realized experimental micro-milling setup. The setup is built around a high-speed micro-milling AMB spindle from EAAT with maximum rotational speed of 120 krpm. A three axis milling machine is constructed, with linear motors for horizontal displacement of the workpiece and a vertical stage for displacement of the spindle.

this is that we consider y_1 the position of the axis of rotation of the rotor, of which the measurement is disturbed by the periodic distortion v_r.

2.3.5 Example

For the research described in this thesis, a micro-milling setup has been realized with a high speed AMB spindle, see Figure 2.4. A three axis milling machine was created using a horizontal XY stage to position the workpiece and a vertically mounted Z-stage for translation of the spindle. Together these stages are mounted on a granite table with a bridge for the Z-axis. All stages are manufactured by Aerotech. The horizontal drives have linear motors, with a positioning accuracy of $< 0.3 \mu$m and a repeatability of $< 0.12 \mu$m. The vertical drive has a lead screw and has a positioning accuracy of $< 0.75 \mu$m and a repeatability of $< 0.1 \mu$m. The selected AMB spindle is from EAAT GmbH Chemnitz (Elektrische Automatiserungs- und Antriebstechniek). This is a medium-sized spindle with a relatively high maximum rotational speed of 120,000 rpm. The rotor length is 250 mm and the rotor mass is 1.1 kg. The displacement sensors used in the bearings have a resolution of 0.1 μm. The airgap length of the radial bearings is 400 μm. The AMB controller hardware also provides measurements of the currents through the coils. The controller itself

2. A MODELING FRAMEWORK FOR MONITORING AND CONTROL WITH AMB
SPINDLES

	rotor	tool
Elasticity modulus [MPa]	225	690
Mass density [kg/m^3]	7850	15700
Poisson ratio [·]	0.30	0.24

Table 2.1: Material properties used for modeling the rotor (tool steel) and the tool (tungsten carbide) in the EAAT AMB spindle model.

	Bottom X and Y	Top X and Y
N (Number of windings) [·]	42	42
A Pole surface [m^2]	$3.9 \cdot 10^{-4}$	$2.6 \cdot 10^{-4}$
i_{bias} (bias current) [A]	4.4	4.4
g (nominal gap length) [m]	$0.4 \cdot 10^{-3}$	$0.4 \cdot 10^{-3}$

Table 2.2: Properties of the magnetic bearings in the EAAT spindle.

	Bottom X and Y	Top X and Y
k_i [N/A]	23.8	15.8
k_x [N/m]	$2.62 \cdot 10^5$	$1.74 \cdot 10^5$

Table 2.3: Parameter values of the actuator model derived for the AMBs of the EAAT spindle.

is an analog PID with additional digitally implemented notch filtering to suppress the response at the bending mode frequencies.

Following the procedure set out in the previous sections, a model for this spindle was derived using Matlab/Simulink. For this, the rotor and tool geometry were together modeled with a set of 91 cylinder elements. For the rotor, made of tool steel, and the tool, made of tungsten carbide, different material properties were selected, as summarized in table 2.1. The properties of the bearings are listed in table 2.2. With these, and using the modeling procedure in Appendix A, the constants of the linear AMB actuator model as listed in table 2.3 were derived. The amplitude frequency response of this model after balanced truncation [118], resulting in a state space model with 24 states, is depicted in Figure 2.5. In this figure, a plot is made of the frequency response for low and high rotational speed. Here the typical dynamics of an AMB spindle can be observed:

- On the diagonal, second order dynamics with cut-off frequency in the lower frequency range, pertaining to the rigid body modes of the rotor;

- Lightly damped modes in the high frequency range, related to the bending modes of the rotor. These modes are present in all elements of the transfer function matrix;

- The dynamics in the same plane are coupled (see e.g. element $(1, 3)$: the transfer function from current top X to displacement bottom X). This coupling is not subject to gyroscopic effects and hence does not vary significantly with the rotational speed;

- The dynamics between the X and Y directions are increasingly coupled with higher rotational speeds (see e.g. element $(1, 2)$: the transfer function from current bottom Y to displacement bottom X). This coupling is indeed caused by the effects of gyroscopy.

To illustrate the mode-splitting effect at high rotational speeds as described in Section 2.3.3, a zoomed-in plot of element $(1, 1)$ of P_{11} in Figure 2.5 is made, see Figure 2.6. This plot clearly shows that at high rotational speeds, flexible mode resonances split into two frequencies.

2.4 Model of the cutting dynamics

The second building block of the modeling framework in this chapter, is the model of the micro-milling dynamics. Modeling of manufacturing processes has been a topic of research in manufacturing for several decades. It is widely acknowledged that modeling the process of formation of chips using a rotating cutter is difficult, and even more so for milling with miniaturized tools. For what follows in this thesis, we do not need a detailed model of the milling dynamics (in fact, we will give an approach to cutting force estimation for which modeling of the milling dynamics is not needed at all). We will therefore not present a full model here, but limit attention to the following more high-level model, which will prove sufficient to understand the complexity of the modeling efforts.

When modeling the cutting dynamics, the objective is to describe the relation between the position of the tooltip $y_{2,0}$ and the forces $u_{2,0}$ that are generated as a result of the cutting process and that act on the tip of the tool, see Figure 2.7. A common approach in literature to model this relationship consists of three steps (see e.g. [11, 9, 49]):

1. Consider a tool with N teeth. For each tooth of the tool $i = 1 \ldots N$, the uncut chip thickness $h_i(t)$ is computed in a coordinate frame attached to the rotating tool. Mechanical modeling of the milling process involves computation of the trochoidal paths traced by the milling teeth, given the rotational motion of the tool and the linear motion of the workpiece towards the tool (the feed). Using knowledge of the milling operation (side milling, slot milling,

2. A MODELING FRAMEWORK FOR MONITORING AND CONTROL WITH AMB SPINDLES
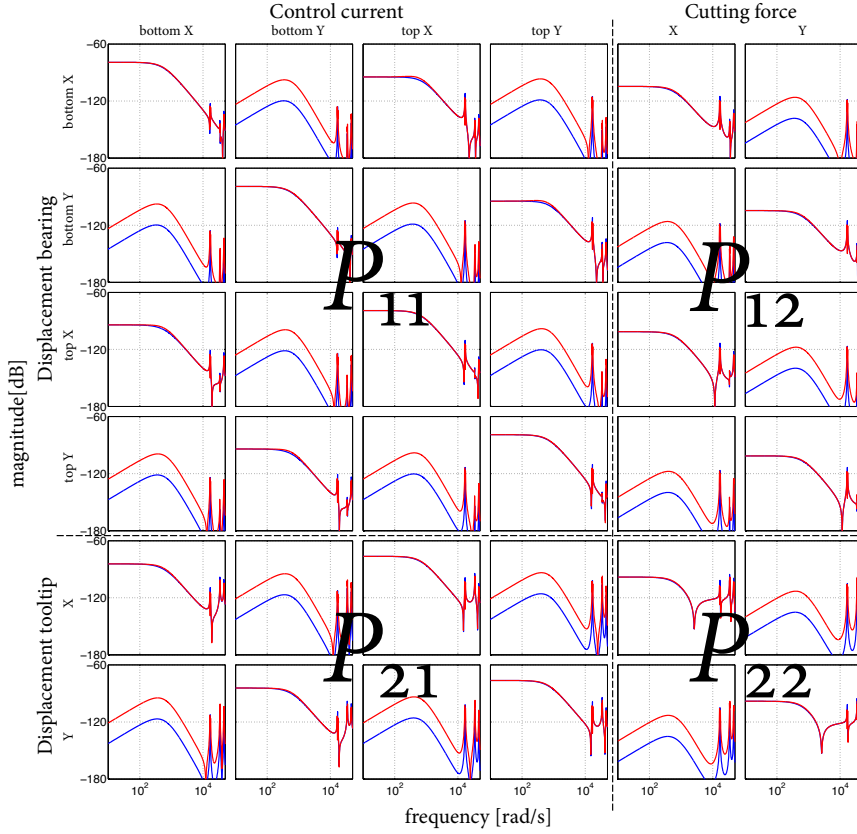
Figure 2.5: Amplitude frequency response of a model of the high-speed micro-milling AMB spindle from EAAT at low and high rotational speed (blue: Ω = 10 krpm, red Ω = 120 krpm). The plot shows the rigid body behavior in the low frequency range, and the resonant behavior of the bending modes in the high frequency range. Coupling of the dynamics between the various directions is present, where the cross terms related to dynamic coupling between the X- and Y-axis increase with rotational speed.

etc), the periodically varying uncut chip thickness can be computed. Extension to dynamic modeling of the milling process is obtained by observing that the feed is not a constant, but a varying quantity that depends on the positioning error of the stages and the position of the tool $y_{2,0}$, as well as their values at the previous tooth pass[5] (i.e. $y_{2,0}(t - \tau)$, with $\tau = \frac{1}{n\Omega}$, where n is

---

[5] To see this, note that the path traced by a particular cutting tooth determines the material that is

Control current bottom X to displacement bottom X
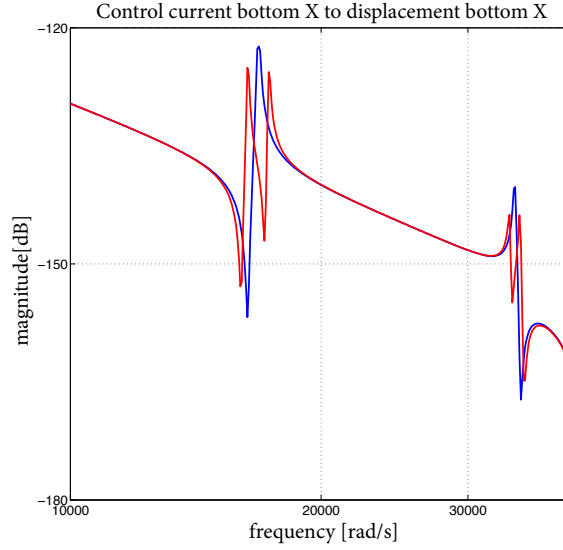
Figure 2.6: Zoomed-in amplitude frequency response of element $(1,1)$ of P_{11} in Figure 2.5 (Blue: Ω = 10krpm, red Ω = 120 krpm): an effect of high speed rotation is that bending modes split.

Cutting dynamics

t, feed, cutting conditions, material properties, tool condition
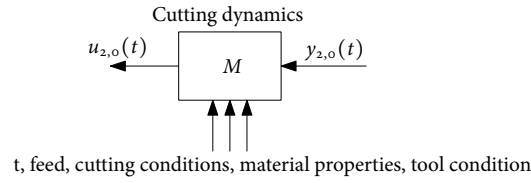
Figure 2.7: Block diagram representing the milling dynamics M, having as input the displacement of the tooltip $y_{2,0}$ (dim($y_{2,0}$) = 2), and as output $u_{2,0}$, the cutting forces (dim($u_{2,0}$) = 2). Exogenous inputs of the milling dynamics include t (to indicate the time-dependency of the dynamics), the feed, the material properties and tool condition.

the number of teeth).

2. The cutting forces on each tooth in tangential and radial direction (see Figure 2.8) are modeled as a static nonlinear function of the uncut chip thickness, where a common form is $F(t) = a_p K_c h_i(t)^\gamma + a_p K_e$, with a_p the cutting

removed, and consequently defines the uncut chip thickness experienced by the next tooth.

43
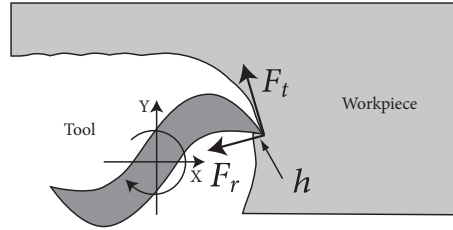
Figure 2.8: Schematic illustrating the modeling of the milling process: at each time instant
the uncut chip thickness h is calculated, from which the force in tangential direc-
tion F_t and radial direction F_r are derived. By coordinate transformation these
forces are mapped in the XY plane. Doing this for all engaged teeth yields the
cutting forces in the X and Y direction.

depth (see Chapter 1), and K_c, K_e and γ are parameters that are material de-
pendent and vary with cutting parameters and the condition of the tool [9].

3. A coordinate transformation is performed to obtain the cutting forces in the
machine coordinate system. This is done for all teeth, the results of which are
summed, resulting in the cutting forces $u_{2,0}$.

From this description, it follows that even a rather simplified modeling of the cut-
ting dynamics between $y_{2,0}$ and $u_{2,0}$ results in a complicated model: it is nonlinear,
time dependent and includes a delay term.

In this context, we remark that the cutting dynamics and the dynamics of the
positioning servos are coupled (in Figure 2.7 this means that there exists a dynamic
feedback of the cutting force $u_{2,0}$ to the feed input of M). Indeed, the forces that
arise during milling, are disturbances to the servos. Depending on the performance
of the servos, these will lead to positioning errors, and consequently to variations
in the uncut chip thickness. In the context of this research we will not further study
this coupling, but will simplify the situation by considering the feed as an exogenous
input of the system. Observe that this does not mean that the dynamics of the
positioning stages are entirely ignored, which would imply that a constant value for
the feed is a assumed. Instead, variations of the feed are allowed, but the causes of
these variations are assumed to lie outside our system boundary.

2.5 Framework for monitoring and control of micro-milling

A model of the AMB spindle in a milling configuration can be obtained by inter-
connection of the AMB spindle model of Section 2.3 and the model of the milling
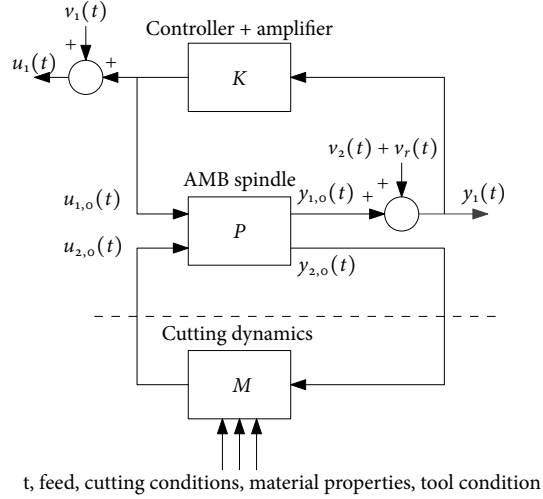
Figure 2.9: Generic model of the AMB spindle in a micro-milling configuration

dynamics of Section 2.4, resulting in the model that is depicted in Figure 2.9. The model of Figure 2.9 exhibits a structure with two coupled closed loops:

i. The control loop of the AMB system, consisting of P and K,

ii. The nonlinear feedback loop caused by the cutting process, consisting of P and M.

For the sake of clarity, we remark that the measured signals in this setup are u_1 and y_1, representing the bearing currents and displacements respectively. The cutting forces $u_{2,0}$ and tooltip displacement vector $y_{2,0}$ cannot be observed without additional hardware.

Within the general framework of this figure, process monitoring and control of micro-milling with AMB spindles can be formulated as model-based observer and control design problems, as follows:

Force estimation: Estimation of the cutting force can be formulated as the objective to design an observer that with measurements of the signals in the first closed-loop — i.e. control currents u_1 and displacements y_1 — estimates the signal $u_{2,0}$ in the second closed-loop.

Tool monitoring: Implementation of tool monitoring in this framework can be formulated as the objective to design a filter F that with u_1 and y_1 as inputs, classifies if the tool is broken or not, or estimates particular parameters of the

cutting dynamics, e.g. a parameter indicating the tool wear status. In both cases, the filter F is used to obtain information about one or more parameters of the dynamics M, using signals of the first closed-loop. In case of tool breakage detection, F is a fault detection filter, in the case of wear monitoring, F is an online parameter estimator. Hence, crucial in the development of such techniques is to derive simplified models of the cutting dynamics M, which explicitly depend on the parameters that need to be estimated.

Force control: The problem of force control can be formulated as the objective to control the cutting parameters (one or more of the exogenous inputs of M), based on some measure of the cutting forces u_2. Generally in force control, the objective is to regulate only the low frequent behavior of the cutting forces. This means that the cutting forces are low-pass filtered, and subsequently the difference with a reference value is used as input for a control algorithm to adapt the cutting parameters.

Chatter control: As reviewed in Chapter 1, chatter is the phenomenon of regenerative vibrations in machining operations. In the context of Figure 2.9, the occurence of chatter can be described as instability of the second closed-loop. Chatter stability research has very much concentrated on predicting for which combinations of cutting conditions this loop is stable, where the dynamics of the machine play an important role (see e.g. [11] and the references therein). With an AMB spindle, this problem can be turned around: chatter control can be formulated as the problem to design a controller K, so that the double closed loop system is stable for a particular choice of cutting parameters.

As argued in Chapter 1, in contributing to the development of model-based monitoring and control for micro-milling using AMB spindles, we will put the main emphasis on the cutting force estimation problem. In the next section we will further specify this problem.

2.6 Force estimation as unknown input estimation

The cutting force estimation problem as formulated in the previous section, involves estimation of the unknown force signal $u_{2,0}$ from the measured bearing signals u_1 and y_1. The complexity and nonlinearity of the milling dynamics is a highly complicating factor in solving this problem. In order to avoid this, the approach that we take is to make a simplification to the framework of Figure 2.9. The simplification involves that we ignore the existence of the closed loop from $y_{2,0}$ to $u_{2,0}$, and treat the cutting forces $u_{2,0}$ as an external disturbance to the AMB spindle system. The obvious advantage of this simplification is that we avoid the complications of modeling of the cutting dynamics for the cutting force estimation problem. Essentially
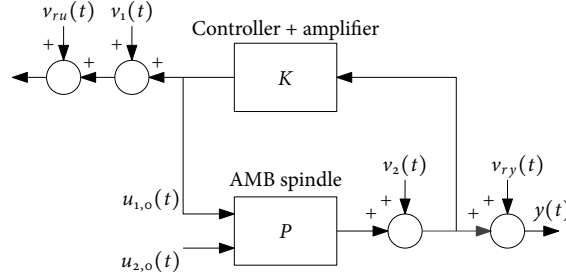
Figure 2.10: Simplified of model of the AMB spindle in a micro-milling configuration for the purpose of cutting force estimation

this simplification means that we ignore the correlation between the noise v_2 and the forces $u_{2,0}$, and instead treat these signals as independent.

For brevity, we will continue with the following change of notation. As the second output of P is no longer relevant, it is omitted and the subscript 1 is dropped of the first output.

The resulting model is the block scheme of Figure 2.10. In this figure another simplification is made regarding the runout disturbance v_r. Based on the superposition principle, disturbance v_r has been replaced by two equivalent noise sourses v_{ry} and v_{ru} outside of the closed-loop, where v_{ry} is the periodic disturbance present in the displacement signal (and is equal to the v_r filtered by the sensitivity of the AMB closed loop), and v_{ru} is the periodic disturbance present in the current signals (and is equal to v_r filtered by the input sensitivity of the AMB closed loop). Observe that the consequence of this operation is that the signals in the closed loop are redefined and no longer include the effect of the runout disturbance. This we will we use to our advantage, as will become clear.

Figure 2.10 is the point of departure for the cutting force estimation approach in this thesis. We treat the cutting forces as an unknown input to the closed-loop AMB spindle setup, and estimate this input from measurements of the bearing signals. In Chapter 3 we will define this unknown input estimation problem more precisely and provide solutions to synthesize optimal input estimation filters. By moving the runout disturbance 'out of the closed-loop', we can treat the effect of runout as a measurement disturbance in the measurements of the control current and displacement signals. Correction of this disturbance can be achieved by using the property that it is periodic with the rotational frequency of the rotor. Hence, the objective of the runout correction procedure is to estimate a correction $\hat{v}_{ry}(t)$ and $\hat{v}_{ru}(t)$ such that $v_{ry}(t) - \hat{v}_{ry}(t)$ and $v_{ru}(t) - \hat{v}_{ru}(t)$ are small. It is the main subject of Chapter 7 to develop such run-out correction. Consequently, we can separately consider the force estimation problem and the runout correction problem, which

2. A MODELING FRAMEWORK FOR MONITORING AND CONTROL WITH AMB
SPINDLES
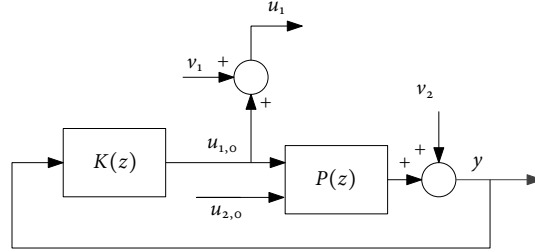
Figure 2.11: Block diagram of the discrete-time equivalent model of the AMB spindle setup.

means that we develop the force estimator under the assumption that no runout distortion is present (which is equivalent to the assumption that perfect correction can be achieved). Hence, for simplicity, we will therefore omit v_{ru} and v_{ry} henceforth, and postpone the correction for these measurement disturbances until Chapter 7.

2.7 Discrete-time equivalent models

Since in practice we will work with sampled data of the control current inputs of the AMB spindle, as well as of the displacement sensor outputs, we will concentrate on discrete-time methods in this thesis. In order to develop these, discrete-time equivalent models of the AMB spindle are needed. Naturally, such equivalent models include a discrete-time representation of the continuous-time cutting force signal. This requires us to make an additional assumption on the intersample behavior, i.e. on how the continuous-time cutting force signal varies between the samples of the discrete-time representation. By making the assumption that the cutting force signal is *Band-Limited* (BL), equivalent discrete-time models of the AMB spindle system can be obtained. This will be done for two configurations:

I. **Direct measurement**: In this configuration the bearing signals are measured directly by means of sampling of the analog control current and displacement sensor signals (after appropriate anti-aliasing filtering). This configuration would typically be used when the AMB spindle setup is designed either with an analog controller, or with a digital controller, of which the internal data is inaccessible. We remark that this is the configuration of the EAAT spindle described in Section 2.3.5.

II. **Integration with digital control**: In this configuration, the discrete-time bearing signals are obtained from a digital control environment.

In Appendix A.3 equivalent discrete-time models for both configurations are derived, based on some reasonable assumptions concerning the sampled signals and

properties of the continuous-time cutting force signal. As follows from the analysis in the appendix, the discrete-time equivalent model of both configurations can be represented by the block diagram in Figure 2.11, yet there are some key differences. In the integrated configuration, the equivalent discrete-time model $P(z)$ includes the dynamics of the D/A conversion and current amplifier, while in the direct measurement configuration these dynamics are included in $K(z)$ (meaning that the frequency response of $P(z)$ and $P(s)$ are the same up to the Nyquist frequency[6]). Furthermore, in the integrated configuration, measurement noise v_1 is nil.

This model will be used as the basis for the development of the force estimator (Chapter 3) and the identification approach of P (Chapter 5 and 6). In order to do so, we will make the following assumptions on P, which are justified from the physics of the systems.

The first assumption is that system P is a strictly proper real rational transfer function matrix with input dimension $m = \dim(u) = 6$ ($m = m_1 + m_2$ with $m_1 = \dim(u_1) = 4$ and $m_2 = \dim(u_1) = 2$) and output dimension $p = \dim(y) = 4$, and which is unstable. Strictly properness of P follows from the fact that the control currents have no direct feedthrough to the output of the system, see equation (2.6). In line with this, it is assumed that a minimal state-space representation for P exists, given by[7]

$$x(t+1) = Ax(t) + B_1 u_{1,0}(t) + B_2 u_{2,0}(t) \tag{2.8a}$$
$$y_0(t) = Cx(t) \tag{2.8b}$$

with $A \in \mathbb{R}^{n \times n}$, $B_1 \in \mathbb{R}^{n \times 4}$, $B_2 \in \mathbb{R}^{n \times 2}$, and $C \in \mathbb{R}^{4 \times n}$.

The second assumption is the property that in the discrete-time model, all modes of the system are controllable through the control current input of the system $u_{1,0}$. This assumption is equivalent to (A, B_1, C) is being minimal in the above state space representation. We will utilize this property in Chapter 6 when identifying the tooltip dynamics. For an AMB spindle with given properties, this assumption can be verified by following the modeling procedure in this chapter, and computing the controllability and observability matrices of the resulting the model. However, this property can also be justified from physical insights directly. In practice, the magnetic bearing spindle system has many modes, of which several are in fact uncontrollable from the magnetic actuators. We will argue that these modes will not be observable in the output of the discrete-time model. Indeed, these modes are typically the vibration modes where the rotor shaft does not move, but only the far end of the cutting tool performs an oscillation motion. However, these modes are in general well above the frequency range of interest, viz. the Nyquist frequency (see also Figure 2.12). This implies that these modes are unobservable in the sampled

---

[6] The relevance of this property becomes clear in Chapter 5, where we will use this for validation of the identification result.

[7] With a slight abuse of notation, but in order to follow common convention, we reintroduce symbol t here. While it was previously used to denote the continuous time, here it is used for the discrete time.

2. A MODELING FRAMEWORK FOR MONITORING AND CONTROL WITH AMB SPINDLES
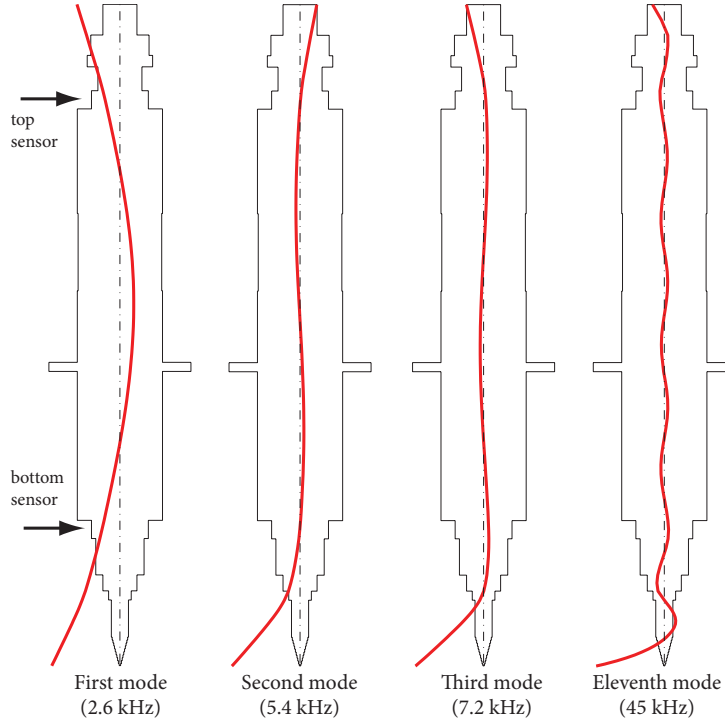
Figure 2.12: Mode shapes of the EAAT AMB spindle system. This figure illustrates the typical mode shapes of the AMB rotor with tool. The slender tool attached to the larger rotor results in mode shapes with large deviations at the tip. Nevertheless, even for high bending modes, the displacement of the rotor shaft is still measurable at the sensor locations. Bending modes for which this is no longer the case have frequencies that well exceeds the frequency range of interest.

output (taking into account that before sampling adequate anti-aliasing filtering is applied)[8]. Observe that this property does not rule out the possibility that some of the modes of P are uncontrollable from $u_{2,0}$. We will make no further statements on this.

.

---

[8]For the sake of completeness: we remark that we tacitly assume that all modes *within* the frequency region of interest are both observable and controllable. This means that the AMB spindle is designed such that the knots of the corresponding bending modes shapes do not exactly coincide with the location of the displacement sensors or the actuation point of the bearings, which is a reasonable assumption.

2.8 Conclusions

In this chapter, a generic model of an AMB spindle in a micro-milling configuration was given, comprising a model of the AMB spindle dynamics and a model of the cutting dynamics. The interconnection of these models results in a configuration with two closed-loops, which enables to formulate monitoring and control of micro-milling as model-based observer and control design problems. With a reasonable simplification of this configuration, the force estimation problem is reduced to an input estimation problem, where the cutting forces are an unknown input to the closed-loop AMB spindle system. Using measured data of the control current and displacement signal and knowledge of the dynamics of the closed-loop, the objective of the cutting force estimator is to estimate this unknown input. It is shown that runout disturbances can be treated as measurement disturbances in the bearing signals, for which compensation is attainable by using their periodic nature. Finally it was discussed that by making a band-limited assumption on the cutting force signal, equivalent discrete-time models can be obtained, facilitating the use of sampled data of the bearing signals. Based on the physics on the setup, assumptions concerning the controllability and observability of the resulting discrete-time models are given and justified.
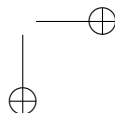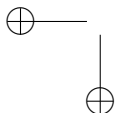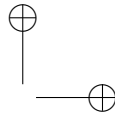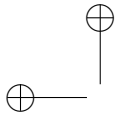
Chapter 3

Unknown input estimation from closed-loop data

3.1 Introduction

In the previous chapter we have described how the problem of estimating the cutting forces from the signals of an AMB spindle can be translated to the problem of estimating unknown inputs to a closed-loop system. This chapter deals with this problem and formulates solutions for optimal linear input estimation filters.

The problem of reconstructing inputs of dynamic systems from measurements of the output has been extensively studied in literature. As our main focus is on model-based approaches, we concentrate on liturature performing input reconstruction for systems for which models of the dynamics are available. It is worthwhile to notice that input reconstruction is closely related to the problem of state estimation of systems with unknown inputs. In literature, estimation of the unknown input is frequently performed hand-in-hand with estimation of the state. The methods can be broadly categorized according to the assumptions that are made concerning the unknown input, as well as the assumption on the noise disturbances on the output measurements, see table 3.1. The first category consists of input estimators that assume no prior knowledge on the unknown input, and which are designed in a deterministic setting (category A in table 3.1). This kind of input estimator is known as the Unknown Input Observer (UIO), which is predominantly considered for continuous-time systems [92, 70, 174, 24, 61, 69, 181, 40]. In structure the UIO is either a reduced-order observer (i.e. of the Luenberger-type) or a full-order observer, having as special property that the unknown input is decoupled from the state estimate. With this, the unknown input can be estimated using the state estimate, and the known inputs and outputs. The UIO is also reported to be of use when dealing with fault detection and isolation [61]. Particularly noteworthy in

	Deterministic	**Stochastic**
No prior information on unknown input assumed	Category A: Unknown Input Observers (UIO)	Category B: Two-stage Kalman Filter (TSKF), Minimum Variance Unbiased (MVU) approaches
Prior information on unknown input used	Category C: Waveform-structured input observers	Category D: Least mean squared error (MSE) observers, Wiener/Kalman filtering

Table 3.1: Approaches to the estimation of unknown inputs to linear systems can be categorized according to the usage of prior information on the unknown inputs, as well as the use of statistical information on the measurement noise

this context are the contributions of Corless and Tu [37] and Park and Stein [124], both which apply UIO designs to cutting force estimation problems in machining. The second category (category B in table 3.1) concerns approaches that also use no prior knowledge on the unknown inputs, but are designed on the basis of noisy output measurements [71, 38, 39, 72, 58, 59]. This category includes approaches that achieve optimal estimates building upon UIO design, as well approaches extending the two-stage Kalman filter approach due to Friedland [52]. The third and fourth category of input observers are characterized by the use of prior knowledge on the input. In category C this is done in a deterministic setting [78, 79, 68, 113]. A key step in these methods is the development of an autonomous state space model that generates the input according to the input model, which is then augmented to the plant model. Input estimation is performed by state estimation of the augmented plant, by design of a (deterministic) state observer.

Category D involves approaches in which the input is modeled as a noise-like signal, i.e. it is considered a realization of a stationary stochastic process of which spectral information is available, and in which the output measurement is disturbed with noise, of which the statistical properties are also known. These approaches have a long history. Indeed, the filtering problem in which the objective is to achieve minimum mean squared error (MSE) is due to Wiener [180], and the literature on this kind of input estimation problem is vast (see [14, 83, 66] and the references therein). Applications are found in several contexts, such as channel equalization in communications, seismology, acoustics, and speech coding. We remark that in addition, various results in literature are available that minimize a H_∞ criterion

instead of minimizing the MSE (being equivalent to H_2 optimization), see [117, 60, 65]. The key idea is that the resulting estimators are more robust with respect to model uncertainties and lack of statistical knowledge of the unknown input.

In this thesis, we have chosen to perform cutting force estimation by input estimation from the closed-loop AMB spindle data within the framework of category D. The main argument is that some a priori information regarding the input signal is available, given the knowledge of the cutting dynamics. From this it is known that cutting forces consist of a slowly varying offset related to the cutting depth, with in addition a signal that is periodic in nature due to the rotation of the milling tool. In pursuit of an observer that includes this prior information, we note that the noise on the measurements, as well as stochastic character of variations in the cutting process — and consequently also the stochastic character of the cutting force signal — make a category D input observer a more logical choice than a category C observer.

In literature, MSE estimation is generally performed in an open-loop setting. In contrast, in this chapter estimation of unknown inputs is addressed from data obtained from a MIMO closed-loop configuration. Here the main issues are the following:

- If full knowledge on the closed-loop is available, classical Wiener filter theory can be applied to perform input estimation. However, this requires the spectrum of the observations to be positive definite. It will be studied what conditions the plant and controller need to satisfy to ensure this.

- In some cases, no knowledge on the controller is available. An approach to deal with this is formulated, which ensures equal performance of the input estimator for all possible controllers. This is achieved by deriving a constraint optimal filter design problem, to which solutions are given.

- Depending on the application of the cutting force estimates, it is of more or less importance to have the estimates available instantly. In various applications small processing delays can be tolerated. In the approach of this chapter, such a delay is therefore included explicitly as a design parameter, turning the estimation problem into a *fixed-lag smoothing* problem. For the closed-loop configuration, both with and without controller knowledge, this problem is addressed.

In addressing these issues, we follow a polynomial matrix approach, allowing us to derive the general structure of the input estimators in a compact form and to analyze their properties. Following causal Wiener filter theory, the solutions rely on spectral factorization and derivation of causal parts. For the open-loop estimation problem it is well established that these operations can be carried out explicitly when a state space realization of the plant is available [83]. With some modifications, this is also the case for the closed-loop configuration, as will become clear in this chapter.

This chapter is organized as follows. We start in Section 3.2 by introducing some notation, relevant mathemetical concepts and main results regarding MSE estimation. Subsequently in Section 3.3 the closed-loop configuration is discussed and the input estimation problem is defined precisely. In Section 3.4, we give general solutions to the input estimation problem in polynomial matrix form, both for the case full knowledge is available on the controller, as well as for the case that no such information can be used. In Section 3.5 explicit solutions for the causal input estimators are formulated via a state space realization. In Section 3.6 several measures are introduced that can be used to assess the performance of the resulting estimators. We conclude this chapter with simulation results obtained by applying the methods in this chapter for cutting force estimation from simulated AMB spindle measurements, Section 3.7.

3.2 Preliminaries

We start by introducing some notation and relevant mathematics. The analysis in this chapter is performed for linear time-invariant MIMO discrete-time systems that can be described by the mapping of the input $u \in \mathbb{R}^m$ to $y \in \mathbb{R}^p$, where $\mathbb{R}$ denotes the field of real numbers, according to $y(t) = \sum_{j=-\infty}^{\infty} T_{t-j} u(j)$ where T_i is a sequence of matrices that belong to $\mathbb{R}^{p \times m}$. The transfer function matrix of this mapping is given by the bilateral z-transform $T(z) = \sum_{i=-\infty}^{\infty} T_i z^{-i}$ for those regions of the complex plane $\mathbb{C}$ for which the infinite summation converges. The regions $\mathbb{C}^-$, $\mathbb{C}^O$, and $\mathbb{C}^+$ represent the open unit disc, the unit circle, and the area outside the unit disc respectively. The function $T(z)$ as defined above will be a rational matrix in z. The set of all rational transfer functions is denoted as RH. The subset of RH containing all stable transfer functions, i.e. all rational transfer functions $T(z)$ that are analytical for all $z \notin \mathbb{C}^-$, is denoted as RH_∞. The hermitian transpose of a transfer function is denoted as $T^H(z) = T^T(1/z)$. Transfer function $T(z)$ will be causal if and only if $T_i = 0$ for $i < 0$ or equivalently, if and only if $T(z)$ is analytic in the exterior of some annulus, $|z| > \alpha > 0$. Similarly, $T(z)$ is anti-causal if $T_i = 0$ for $i > 0$, or equivalently, if and only if $T(z)$ is analytic in the interior of some annulus, $|z| < \alpha < \infty$. Moreover, $T(z)$ is strictly causal (strictly anti-causal) if and only if it is causal (anticausal) and $T(\infty) = 0$ ($T(0) = 0$). The causality and anti-causality operators are denoted by $\{\cdot\}_+$ and $\{\cdot\}_-$ respectively and are defined as follows:

$$\{T(z)\}_+ = \sum_{i=0}^{\infty} T_i z^{-i} \qquad\qquad \{T(z)\}_- = \sum_{i=-\infty}^{-1} T_i z^{-i}.$$

The class of signals considered in this chapter consists of realizations of stationary random processes. We will use the following notions. Let $x(t)$ and $y(t)$ denote two random processes. The autocorrelation function of $x(t)$ and the cross correlation function between $x(t)$ and $y(t)$ are respectively given by

$$R_x(\tau) = \mathbb{E}x(t)x(t-\tau)^T \qquad\qquad R_{yx}(\tau) = \mathbb{E}y(t)x(t-\tau)^T$$

The z-spectrum (or briefly, spectrum) of $x(t)$ and the cross z-spectrum (or briefly, cross spectrum) between $y(t)$ and $x(t)$ are defined by the bilateral z-transform of $R_x(\tau)$ and $R_{yx}(\tau)$,

$$\Phi_x(z) = \sum_{\tau=-\infty}^{\infty} R_x(\tau) z^{-\tau} \qquad \Phi_{yx}(z) = \sum_{\tau=-\infty}^{\infty} R_{yx}(\tau) z^{-\tau}.$$

The power spectral density of $x(t)$ is found by evaluating $\Phi_x(z)$ on the unit circle, i.e. $\Phi_x(e^{j\omega})$. Similarly, the cross spectral density of $x(t)$ and $y(t)$ is given by $\Phi_{yx}(e^{j\omega})$. The power of process $x(t)$ is given by

$$\sigma_x^2 = \mathrm{tr} R_x(0) = \frac{1}{2\pi} \mathrm{tr} \int_0^{2\pi} \Phi(e^{j\omega}) d\omega.$$

This expression shows that process $x(t)$ will have finite power if and only if $\Phi_x(z)$ is bounded on the unit circle. A sufficient (but stronger than necessary) condition for this is that $R_x(\tau)$ is exponentially bounded, i.e., that there exists a positive definite constant matrix $T > 0$ and a positive scalar $\alpha < 1$, such that $R_x(\tau) < T\alpha^{|\tau|}$ for $-\infty < \tau < \infty$, which ensures absolute convergence of $\Phi_x(z)$ on an annulus containing the unit circle $\alpha < |z| < \alpha^{-1}$ [83].

As indicated in the introduction, minimum MSE input estimation strongly relies on the notion of spectral factorization, which is defined in the following theorem.

Proposition 3.2.1 (Spectral Factorization [14, 83, 140, 66]) *Consider* $\Phi(z) \in RH$ *with the following properties:*

- $\Phi(z)$ *has full rank for almost every* $z \in \mathbb{C}$

- $\Phi(z) = \Phi^H(z)$

- $\Phi(z)$ *is analytic on* $|z| = 1$, *and* $\Phi(z) > 0$ *for all* $|z| = 1$.

There exist a factorization of $\Phi(z)$, *given by* $\Phi(z) = M(z)M^H(z)$ *where*

- $M(z) \in RH$ *is analytic for* $|z| > \alpha$, $\alpha > 0$ *(i.e.* $M(z)$ *is causal),*

- $M(z)$ *has constant rank in* $\mathbb{C}^+$, *or equivalently,* $M^{-1}(z)$ *is analytic for* $|z| \geq 1$ *(*$M^{-1}(z)$ *is causal and stable)*

$M(z)$ *is called a causally invertible spectral factor of* $\Phi(z)$.

We remark that often in the literature on spectral factorization, the formulation of the above theorem is somewhat more restrictive in the sense that $M(z)$ is not only causally invertible, but also stable. The relaxed formulation of proposition 3.2.1 will prove to be useful for our purposes, as will become clear later.

With the above, we can now formulate the following two classic results concerning MSE estimation of random processes.

3. UNKNOWN INPUT ESTIMATION FROM CLOSED-LOOP DATA

Proposition 3.2.2 (The Wiener Smoother) *Consider two zero mean jointly station-ary random processes $u(t)$ and $y(t)$ with known spectra $\Phi_u(z) \in RH$, $\Phi_y(z) \in RH$ and known cross spectrum $\Phi_{uy}(z) \in RH$. Assume that $\Phi_y(z) > 0$ for all $|z| = 1$. Let u be estimated by a linear filter on y, i.e. $\hat{u}(t) = F(z)y(t)$, with $F(z) \in \mathbb{F}$, where $\mathbb{F}$ is defined as the set of all (causal and non-causal) filters that yield a bounded estimate for all y:*

$$\mathbb{F} := \{F \in RH | F(z) \text{ is analytic on the unit circle}\}.$$

Define the estimation error $\epsilon(t) = u(t) - \hat{u}(t)$. Then the solution to the MSE estima-tion problem $F_0 = \arg\min_{\mathbb{F}} \mathbb{E}\epsilon(t)^T \epsilon(t)$ is given by

$$F_0(z) = \Phi_{uy}^H(z)\Phi_y^{-1}(z). \tag{3.1}$$

The corresponding minimum MSE is given by

$$\mathbb{E}\epsilon(t)^T \epsilon(t) = \frac{1}{2\pi}\text{tr} \int_0^{2\pi} \Phi_\epsilon(e^{j\omega})d\omega \tag{3.2}$$

with estimation error spectrum

$$\Phi_\epsilon(z) = \Phi_u(z) - \Phi_{uy}(z)\Phi_y^{-1}(z)\Phi_{uy}^H(z). \tag{3.3}$$

Proof See e.g. [83]

Proposition 3.2.3 (The N-causal Wiener filter) *Consider two zero mean jointly sta-tionary random processes $u(t)$ and $y(t)$ with known spectra $\Phi_u(z) \in RH$, $\Phi_y(z) \in RH$ and known cross spectrum $\Phi_{uy}(z) \in RH$. Assume that $\Phi_y(z) > 0$ for all $|z| = 1$. Let u be estimated by a linear filter on y, i.e. $\hat{u}(t) = F(z)y(t)$, with $F(z) \in \mathbb{F}_{\text{caus},N}$, where $\mathbb{F}_{\text{caus},N}$ is defined as the set of all filters that are causal and stable when delayed by N time steps:*

$$\mathbb{F}_{\text{caus},N} := \{F \in RH | z^{-N}F(z) \text{ is analytic on and outside the unit circle}\}$$

Define the estimation error $\epsilon(t) = u(t) - \hat{u}(t)$. Then the solution to the MSE estima-tion problem $F_{0,N} = \arg\min_{\mathbb{F}_{\text{caus},N}} \mathbb{E}\epsilon(t)^T \epsilon(t)$ is given by

$$z^{-N}F_{0,N}(z) = \{z^{-N}\Phi_{uy}^H(z)M^{-H}(z)\}_+ M^{-1}(z). \tag{3.4}$$

where $M(z)$ is found from the spectral factorization of $\Phi_y(z) = M(z)M^H(z)$ ac-cording to proposition 3.2.1. The corresponding minimum MSE is given by

$$\mathbb{E}\epsilon(t)^T \epsilon(t) = \frac{1}{2\pi}\text{tr} \int_0^{2\pi} \Phi_\epsilon(e^{j\omega})d\omega \tag{3.5}$$

with estimation error spectrum

$$\begin{aligned}\Phi_\epsilon(z) = \Phi_u(z) - &\Phi_{uy}(z)\Phi_y^{-1}(z)\Phi_{uy}^H(z)+ \\ &+ (\{z^{-N}\Phi_{uy}^H(z)M^{-H}(z)\}_-)(\{z^{-N}\Phi_{uy}^H(z)M^{-H}(z)\}_-)^H. \end{aligned} \tag{3.6}$$

Proof See e.g. [83].

The set $\mathbb{F}_{\mathrm{caus},N}$ is the subset of $\mathbb{F}$, of which the elements themselves are not necessarily causal, but which are causal when delayed by N timesteps. It is worthwhile to elaborate on the properties of the filters in $\mathbb{F}_{\mathrm{caus},N}$. For this, let us expand an arbitrary element $F_N(z) \in \mathbb{F}_{\mathrm{caus},N}$ as $F_N(z) = \sum_{i=-N}^{\infty} F_i z^{-i}$. With this, the expression for the estimate $\hat{u}$ can be written as

$$\hat{u}(t) = \sum_{i=-N}^{-1} F_i y(t-i) + \sum_{i=0}^{\infty} F_i y(t-i).$$

This expression shows that $\hat{u}(t)$ is calculated using the full past of $y(t)$ plus future observations up to a fixed horizon. The causal input estimator with $N = 0$ calculates $\hat{u}(t)$ using present and past observations of $y(t)$ only. Now consider the relation between u and y is given by the mapping $y = Pu + v$, with v a stationary zero mean random process, uncorrelated with u. If P is strictly proper, implying that $y(t)$ is a not function of the input $u(t)$ (but only its past), then the estimate $\hat{u}(t)$ obtained with this estimator will neither be a function of the true input $u(t)$. Instead, it is a *prediction* based on the known cross-correlation function of u and y. Depending on this cross-correlation, as well as on the properties of v, the obtained prediction may be poor.

In contrast, for $N \geq k+1$, with k the number of delays in P, one ore more observations correlated with $u(t)$ are used to make the estimate $\hat{u}(t)$. Indeed, the estimation error obtained with $F_{0,N}$ decreases monotonely with increasing delay N and it is bound from below by the error obtained with F_0. To see this, let ϵ_N denote the estimation error obtained with $F_{0,N}$ and ϵ denote the estimation error obtained with F_0. Observe that for any $N_1 > N_2$, $\mathbb{F}_{\mathrm{caus},N_2} \subset \mathbb{F}_{\mathrm{caus},N_1} \subset \mathbb{F}$. Consequently $\mathbb{E}\epsilon_{N_2}^T \epsilon_{N_2} > \mathbb{E}\epsilon_{N_1}^T \epsilon_{N_1} > \mathbb{E}\epsilon^T \epsilon$. When implementing these filters in practice, the price to be paid for the lower error with increasing N is that a larger delay will be incurred in getting the estimate available. Indeed, since $z^{-N}F_N$ is causal, it is easy to see that practical implementation of the N-causal input estimator causes a delay of N time steps: $\hat{u}(t-N) = z^{-N}F_N y(t)$. A filter with such structure is generally referred to as a *fixed lag smoother* [83].

In the following sections we will apply the Wiener smoother and filter to a closed-loop configuration, and perform the analysis using polynomial equations. Hence, we conclude this section with some properties of polynomial matrices (for a more detailed discussion of these, see [82]). A $p \times m$ polynomial matrix is a matrix of the form $A(\xi) = A_0 + A_1 \xi + \cdots + A_n \xi^n$ with ξ an indeterminate variable taking values in the complex plane, and $A_i \in \mathbb{R}^{p \times m}$, $i = 0 \ldots n$ constant matrices. A polynomial matrix $A(\xi)$ has *full column (row) normal rank* if it has full column (row) rank everywhere in the complex plane, except at a finite number of points. A polynomial matrix is *left (right) prime* if it has full row (column) rank everywhere in the complex plane. A square matrix $U(\xi)$ is *unimodular* if its determi-
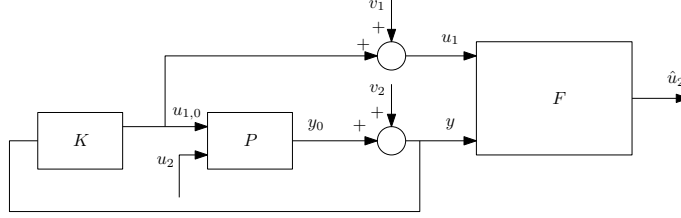
Figure 3.1: Configuration of the input estimation problem for closed-loop data. Filter F estimates unknown input u_2 from the closed-loop data u_1 and y.

nant $\det(U(\xi))$ is a nonzero constant. The inverse of $U(\xi)$ is then again a polynomial matrix. Polynomial matrices $A_1(\xi)$ and $A_2(\xi)$ with equal row dimension are *left coprime* if $[A_1(\xi)\ A_2(\xi)]$ is left prime. This is equivalent to saying that the greatest left common divisor of $A_1(\xi)$ and $A_2(\xi)$ is unimodular. Similarly, polynomial matrices $B_1(\xi)$ and $B_2(\xi)$ with equal column dimension are *right coprime* if $\mathrm{col}(B_1(\xi), B_2(\xi))$ is right prime. This is equivalent to saying that the greatest right common divisor of $B_1(\xi)$ and $B_2(\xi)$ is unimodular. The *zeros* of $A(\xi)$ are the points in the complex plane where $A(\xi)$ loses rank. A left matrix fraction description (MFD) $T(z) = D_L^{-1}(z^{-1})N_L(z^{-1})$ is called *irreducible* if $D_L(\xi)$ and $N_L(\xi)$ are left coprime. Similarly, a right MFD $T(z) = N_R(z^{-1})D_R^{-1}(z^{-1})$ is called irreducible if $N_R(\xi)$ and $D_R(\xi)$ are right coprime.

3.3 Configuration and problem statement

We consider the closed-loop system depicted in Figure 3.1, consisting of plant $P(z) = [P_1(z)\ P_2(z)]$ of dimensions $p \times (m_1 + m_2)$ and controller $K(z)$ of dimensions $m_1 \times p$. Plant $P(z)$ and controller $K(z)$ are both linear time invariant, real-rational causal transfer functions, where in addition $P(z)$ is strictly proper, and controller $K(z)$ is proper. Input $u_2 \in \mathbb{R}^{m_2}$ is an unknown input to plant P. Output $u_1 \in \mathbb{R}^{m_1}$ is the measurement of control input $u_{1,0}$, corrupted by additive noise v_1. Output $y \in \mathbb{R}^p$ is the output of plant P, corrupted by additive noise v_2. Let $T(z)$ be the transfer function of $\mathrm{col}(u_2, v_1, v_2)$ to $\mathrm{col}(u_1, y)$:

$$\begin{pmatrix} u_1 \\ y \end{pmatrix} = \underbrace{\begin{bmatrix} T_{u_1u_2}(z) & T_{u_1v_1}(z) & T_{u_1v_2}(z) \\ T_{yu_2}(z) & T_{yv_1}(z) & T_{yv_2}(z) \end{bmatrix}}_{T(z)} \begin{pmatrix} u_2 \\ v_1 \\ v_2 \end{pmatrix} \tag{3.7}$$

where

$$T_{u_1 u_2}(z) = K(z)S(z)P_2(z) \qquad T_{u_1 v_1}(z) = I \qquad T_{u_1 v_2}(z) = K(z)S(z)$$
$$T_{y u_2}(z) = S(z)P_2(z) \qquad T_{y v_1}(z) = 0 \qquad T_{y v_2}(z) = S(z)$$

and $S(z) = (I - P_1(z)K(z))^{-1}$. Is is assumed that the closed-loop is stable. Concerning the external input signals u_2, v_1 and v_2 the following assumption is made.

Assumption 3.3.1 *It is assumed that*

1. *Unknown input u_2, and noise processes v_1 and v_2 are exponentially bounded stationary random processes. The power spectral density functions of $u_2(t)$, $v_1(t)$, and $v_2(t)$ are known, rational and strictly positive:*
   $\Phi_u(e^{j\omega}) > 0$, $\Phi_{v_1}(e^{j\omega}) > 0$, *and* $\Phi_{v_2}(e^{j\omega}) > 0$, $\forall \omega \in [0, 2\pi]$.

2. *Inputs $u_2(t)$, $v_1(t)$ and $v_2(t)$ are mutually uncorrelated.*

Furthermore, it is assumed that the spectrum of $\mathrm{col}(u_1, y)$ is positive definite:

Assumption 3.3.2 *Let $\Sigma(z)$ be the spectrum of $\mathrm{col}(u_1, y)$. It is assumed that $\Sigma(z) > 0$ for all $|z| = 1$.*

Our goal is to design a linear transfer function $F(z) = [F_1(z) \ F_2(z)]$ that optimally estimates the unknown input sequence u_2 using the observations u_1 and y. The estimate is denoted as $\hat{u}_2$, which can be expressed as $\hat{u}_2 = F_1(z)u_1 + F_2(z)y$. Let the estimation error ϵ be denoted as $\epsilon(t) = u_2(t) - \hat{u}_2(t)$. We can now define the optimal non-causal and N-causal input estimation problems as follows.

Definition 3.3.3 (Optimal non-causal input estimator) *The optimal input estimator F_0 is defined as:*

$$F_0 = \arg \min_{\mathbb{F}} \mathbb{E}\epsilon(t)^T \epsilon(t) \tag{3.8}$$

where $\mathbb{F}$ is defined as the set of (causal and non-causal) filters that yield bounded estimation error:

$$\mathbb{F} := \{F \in RH | F(z) \text{ is analytic on the unit circle}\}$$

Definition 3.3.4 (Optimal N-causal input estimator) *The N-causal optimal input estimator $F_{0,N}$ is defined as:*

$$F_{0,N} = \arg \min_{\mathbb{F}_{\text{caus},N}} \mathbb{E}\epsilon(t)^T \epsilon(t) \tag{3.9}$$

where $\mathbb{F}_{\text{caus},N}$ is defined as the set of filters that are causal and stable when delayed by N time steps:

$$\mathbb{F}_{\text{caus},N} := \{F \in \mathbb{F} | z^{-N}F(z) \text{ is analytic on and outside the unit circle}\}$$

3.4 Polynomial solutions of the input estimation problem from closed-loop data

In the following sections we will give the solutions to the input estimation problems defined in the previous section. First in Section 3.4.1 we will do this for the case we have full knowledge on controller $K(z)$. For this case, we can directly apply the Wiener filter results from literature as formulated in Section 3.2. In addition, we will give conditions on the closed-loop that are necessary in order to apply the Wiener filter results. Section 3.4.2 extends the results for the case no information on $K(z)$ is available. An approach is formulated by imposing an extra constraint to the optimal estimator, ensuring equal performance of it for any controller. In the last Section 3.4.3 we investigate under which conditions the input can be observed from the output measurements.

3.4.1 Full controller knowledge

We will first consider the situation that full information on the controller $K(z)$ is available, in which case we can directly apply the Wiener filter results in Section 3.2. The result of that for the optimal non-causal input estimator is summarized in the following proposition:

Proposition 3.4.1 *Under assumptions 3.3.1 and 3.3.2, the optimal non-causal input estimator exists and is given by*

$$F_0(z) = \begin{bmatrix} 0 & \Phi_u(z)P_2^H(z)\Psi^{-1}(z)S(z)^{-1} \end{bmatrix} \tag{3.10}$$

with $\Psi(z) = P_2(z)\Phi_u(z)P_2^H(z) + \Phi_{v_2}(z)$. *The spectrum of the estimation error* $\Phi_{\epsilon,0}(z)$ *obtained with* $F_0(z)$ *is given by*

$$\Phi_{\epsilon,0}(z) = \Phi_u(z) - \Phi_u(z)P_2^H\Psi^{-1}(z)P_2(z)\Phi_u(z) \tag{3.11}$$

Proof Since by assumption $\Sigma(z) > 0$ for $|z| = 1$, it follows from proposition 3.2.3 that the Wiener smoother for u_2 from $\mathrm{col}(u_1, y)$ is given by

$$F_0(z) = \Phi_u(z) \begin{pmatrix} T_{u_1 u_2}(z) \\ T_{y u_2}(z) \end{pmatrix}^H \Sigma^{-1}(z). \tag{3.12}$$

Now note that we can express for $\Sigma(z)$

$$\Sigma(z) = \begin{pmatrix} I & KS \\ 0 & S \end{pmatrix} \begin{bmatrix} \Phi_{v_1} & 0 \\ 0 & \Psi \end{bmatrix} \begin{pmatrix} I & KS \\ 0 & S \end{pmatrix}^H \tag{3.13}$$

with $\Psi(z) = P_2\Phi_u P_2^H + \Phi_{v_2}$. From this, we find that

$$\Sigma^{-1}(z) = \begin{bmatrix} I & 0 \\ -K^H & S^{-H} \end{bmatrix} \begin{bmatrix} \Phi_{v_1} & 0 \\ 0 & \Psi \end{bmatrix}^{-1} \begin{bmatrix} I & -K \\ 0 & S^{-1} \end{bmatrix}. \tag{3.14}$$

Substituting in (3.12) the expressions for $T_{u_1 u_2}$ and $T_{y u_2}$ from equation (3.7), and expression (3.32), yields

$$F_0 = \begin{bmatrix} 0 & \Phi_u P_2^H \Psi^{-1} S^{-1} \end{bmatrix} \tag{3.15}$$

The expression for the estimation error spectrum is derived similarly from expression 3.3. ∎

Observe that the first element of $F_0(z)$ is zero. This is logical in the case we have full information on the controller. Indeed, output u_1 is a corrupted measurement of $u_{1,0}$, which is the result of the mapping $u_{1,0} = K(z)y$. Knowing y and having exact information on $K(z)$, we have exact information on the sequence $u_{1,0}$ [1]. Hence, no increased performance can be expected by using sequence u_1.

Furthermore, observe that under the given conditions the second element of $F_0(z)$ is the Wiener smoother for u_2 from observations y. Indeed, ignoring observations u_1, we can simplify the estimation problem to the problem of estimating u_2 from observations $y = T_{y u_2}(z)u_2 + S(z)v_2$. Assumption 3.3.2 implies $\Phi_y(z) > 0$ for $|z| = 1$. Hence, the Wiener smoother for input u_2 is given by

$$F^w = \Phi_{u_2 y} \Phi_y^{-1} \tag{3.16}$$

$$= \Phi_u T_{y u_2}^H (T_{y u_2} \Phi_u T_{y u_2}^H + S \Phi_{v_1} S^H)^{-1}, \tag{3.17}$$

which with the definitions of $T_{y u_2}$ and S is identical to the second element of $F_0(z)$.

In the above proposition, usage was made of the assumption that $\Sigma(z) > 0$ for $|z| = 1$ (assumption 3.3.2). We will analyze the implications of this assumption for the closed-loop structure. With equation (3.7), we can express for $\Sigma(z)$

$$\Sigma(z) = T(z) \begin{bmatrix} \Phi_u(z) & 0 & 0 \\ 0 & \Phi_{v_1}(z) & 0 \\ 0 & 0 & \Phi_{v_2}(z) \end{bmatrix} T^H(z). \tag{3.18}$$

By assumption 3.3.1, the center matrix in this product is positive definite for all $|z| = 1$. Hence $\Sigma(z) > 0$ for $|z| = 1$ iff $T(z)$ has full row rank for all $|z| = 1$. We will derive what conditions P and K must satisfy for $T(z)$ to have this property. For that, let us introduce MFDs for P and K. Let

$$P(z) = D_P^{-1}(z^{-1}) \begin{bmatrix} N_{P,1}(z^{-1}) & N_{P,2}(z^{-1}) \end{bmatrix} \tag{3.19}$$

be an irreducible left MFD of $P(z)$, i.e. $D_P(\xi)$ and $\begin{bmatrix} N_{P,1}(\xi) & N_{P,2}(\xi) \end{bmatrix}$ are left coprime. Furthermore, let $N_K(z^{-1})D_K^{-1}(z^{-1})$ be an irreducible right MFD for $K(z)$. By assumption the closed-loop is stable, which implies that any common factor of $D_P(\xi)$ and $N_{P,1}(\xi)$ is stable. We will however make the stronger assumption

---

[1]Of course this statement holds for the case the system has reached its steady state.

that $D_P(\xi)$ and $N_{P,1}(\xi)$ are left coprime, implying that all modes of the system are controllable from input u_1. Note that not all modes of P are necessarily controllable from input u_2. Stated otherwise, $D_P(\xi)$ and $N_{P,2}(\xi)$ are not necessarily left-coprime. We define $W(\xi)$ as the largest common left divisor of $D_P(\xi)$ and $N_{P,2}(\xi)$, i.e. $W(\xi)$ satisfies

$$D_P(\xi) = W(\xi)\bar{D}_P(\xi)$$
$$N_{P,2}(\xi) = W(\xi)\bar{N}_{P,2}(\xi)$$

such that $\bar{D}_P(\xi)$ and $\bar{N}_{P,2}(\xi)$ are left coprime. Using this notation, we can express $S(z)$ as

$$S(z) = (I - D_P^{-1}N_{P,1}N_K D_K^{-1})^{-1} \tag{3.21}$$

$$= D_K(D_P D_K - N_{P,1}N_K)^{-1} D_P \tag{3.22}$$

where we note that $\gamma := D_P D_K - N_{P,1}N_K$ is stable by the assumption of stability of the closed-loop. With this, we obtain

$$T(z) = \left[\begin{array}{cccc} N_K \gamma^{-1} W \tilde{N}_{P,2} & I & N_K \gamma^{-1} W \tilde{D}_P \\ D_K \gamma^{-1} W \tilde{N}_{P,2} & 0 & D_K \gamma^{-1} W \tilde{D}_P \end{array} \right]. \tag{3.23}$$

From this, it is immediate that $T(z)$ has full row rank for all $|z| = 1$ iff $W(z^{-1}) > 0$ and $D_K(z^{-1}) > 0$ for all $|z| = 1$. This condition is equivalent to saying that

1. None of the modes of $P(z)$ that are uncontrollable from input u_2 are on the unit circle, and

2. $K(z)$ has no poles on the unit circle.

We will now proceed to the solution for the N-causal input estimator, which is given in the following proposition:

Proposition 3.4.2 *Under assumptions 3.3.1 and 3.3.2, the N-causal optimal input estimator exists and is given by*

$$z^{-N}F_{0,N}(z) = \left[\begin{array}{cc} 0 & \{z^{-N}\Phi_u(z)P_2^H(z)S^H(z)M_S^{-H}(z)\}_+ M_S^{-1}(z) \end{array} \right] \tag{3.24}$$

where $M_S(z)$ is a causally invertible spectral factor of

$$\Psi_S(z) = S(z)(P_2(z)\Phi_u(z)P_2^H(z) + \Phi_{v_2}(z))S^H(z). \tag{3.25}$$

The spectrum of the estimation error obtained with $F_{0,N}$ is denoted as $\Phi_{\epsilon,0,N}$ and is given by

$$\Phi_{\epsilon,0,N}(z) = \Phi_u(z) - \Phi_u(z)P_2^H(z)\Psi_S^{-1}(z)P_2\Phi_u(z)+$$
$$+ \left(\{z^{-N}\Phi_u(z)P_2^H(z)S^H M_S^{-H}(z)\}_-\right)\left(\{z^{-N}\Phi_u(z)P_2^H(z)S^H M_S^{-H}(z)\}_-\right)^H. \tag{3.26}$$

Polynomial solutions of the input estimation problem from closed-loop data

If in addition $K(z)$ is stable, then the N-causal optimal input estimator is given by

$$z^{-N} F_{0,N}(z) = \begin{bmatrix} 0 & \{z^{-N} \Phi_u(z) P_2^H(z) M^{-H}(z)\}_+ M^{-1}(z) S(z)^{-1} \end{bmatrix} \qquad (3.27)$$

where $M(z)$ is a causally invertible spectral factor of

$$\Psi(z) = P_2(z) \Phi_u(z) P_2^H(z) + \Phi_{v_1}(z). \qquad (3.28)$$

The corresponding estimation error spectrum is

$$\Phi_{\epsilon,0,N} = \Phi_u - \Phi_u(z) P_2^H \Psi^{-1} P_2 \Phi_u +$$
$$+ \left(\{z^{-N} \Phi_u(z) P_2^H(z) M^{-H}(z)\}_- \right) \left(\{z^{-N} \Phi_u(z) P_2^H(z) M^{-H}(z)\}_- \right)^H \qquad (3.29)$$

Proof Since by assumption $\Sigma(z) > 0$ for $|z| = 1$, it follows from proposition 3.2.2 that the N-causal Wiener filter for u_2 from $\mathrm{col}(u_1, y)$ is given by

$$F_0(z) = \{z^{-N} \Phi_u(z) \begin{pmatrix} T_{u_1 u_2}(z) \\ T_{y u_2}(z) \end{pmatrix}^H M_\Sigma^{-H}(z)\}_+ M_\Sigma^{-1}(z) \qquad (3.30)$$

where $\Sigma(z) = M_\Sigma(z) M_\Sigma^H(z)$ is a spectral factorization according to proposition 3.2.1. Now note that we can express $\Sigma(z)$ as

$$\Sigma(z) = \begin{pmatrix} S^{-1} & K \\ 0 & I \end{pmatrix} \begin{bmatrix} S \Phi_{v_1} S^H & 0 \\ 0 & \Psi_S \end{bmatrix} \begin{pmatrix} S^{-1} & K \\ 0 & I \end{pmatrix}^H \qquad (3.31)$$

with $\Psi_S(z) = S(z)(P_2(z) \Phi_u(z) P_2^H(z) + \Phi_{v_2}(z)) S(z)$. From this, we find that

$$\Sigma^{-1}(z) = \begin{bmatrix} S^H & 0 \\ -(SK)^H & I \end{bmatrix} \begin{bmatrix} S \Phi_{v_1} S & 0 \\ 0 & \Psi_S \end{bmatrix}^{-1} \begin{bmatrix} S & -SK \\ 0 & I \end{bmatrix}. \qquad (3.32)$$

Note that the left and right matrix in this product are stable. As by definition $M_\Sigma(z)$ is a causally invertible spectral factor, $M_\Sigma^{-1}(z)$ is stable, from which it follows that we can write

$$M_\Sigma^{-1}(z) = \begin{bmatrix} S & 0 \\ -(SK)^H & I \end{bmatrix} \begin{bmatrix} * & 0 \\ 0 & M_S^{-1}(z) \end{bmatrix}. \qquad (3.33)$$

where it is used that $M_S(z)$ is a causally invertible spectral factor of $\Psi_S(z)$. Substituting in (3.30) the expressions for $T_{u_1 u_2}$ and $T_{y u_2}$ from equation (3.7), and expression (3.33), yields the optimal filter $F_{0,N}$. The expression for the estimation error spectrum is derived similarly from expression (3.6).

If in addition $K(z)$ is stable, we can write $M_S(z) = S(z) M(z)$. This can be observed as follows. First note that using the MFD for P introduced in equation (3.19), we can write

$$\Psi = D_P^{-1} \left[W(\tilde{N}_{P,2} \Phi_u \tilde{N}_{P,2}^H + \tilde{D}_P \Phi_{v_2} D_P^H) W^H \right] D_P^{-H} \qquad (3.34)$$

from which we see that D_P^{-1} is a left divisor of $M(z)$. Also, from equation (3.21) we know that D_P is a right divisor of $S(z)$. Hence, we can write

$$\Phi_S(z) = S\Psi S^H \qquad (3.35)$$

$$= D_K \gamma^{-1} \left[W(\tilde{N}_{P,2} \Phi_u \tilde{N}_{P,2}^H + \tilde{D}_P \Phi_{v_2} D_P^H) W^H \right] \gamma^{-H} D_K^H. \qquad (3.36)$$

If $K(z)$ is stable, we conclude that $D_K \gamma^{-1}$ is a left divisor of $M_S(z)$, and hence $S(z)$ is a left divisor of $M_S(z)$. Substituting $M_S(z) = S(z)M(z)$ in equation (3.24) and (3.26), yields equation (3.27) and (3.29) respectively. ∎

3.4.2 No controller knowledge

The optimal solution in the previous section was obtained by using full knowledge of the controller. This information is not always available. One way to deal with this is by formulating the input estimator design problem in a robust estimation framework. This would involve usage of nominal knowledge of K combined with an uncertainty description to include the variations of controller K. The goal then is to design F so, as to minimize the variations of Φ_e in some to be defined norm for all K in the uncertainty set. We will not further analyze this problem here.

Instead, we take the following somewhat more conservative approach. We define the optimal filter for the case K is unknown as the filter that gives minimum estimation error *for all K* stabilizing P. The optimal filter according to this definition will be independent of K and therefore also independent of the performance of the closed-loop. The practical advantage is that controller and estimator can be designed independently. In this section, the solution to this filter design problem is given. In addition we will analyze the conservatism in requiring equal performance of the input estimator for all stabilizing K, and draw conclusions on the applicability of this approach.

Equal performance of F can be obtained for all possible controllers, when F satisfies an extra constraint, as expressed by the following proposition.

Proposition 3.4.3 *Consider the stable closed-loop with controller K and the input estimator F, according to the configuration detailed in Section 3.3. If a filter F satisfies $F_1 = -F_2 P_1$, then Φ_e is independent of K.*

Proof Observe that we can express $u_{1,0}$ as

$$u_{1,0} = T_{u_1 u_2}(z) u_2 + T_{u_1 v_2}(z) v_2. \qquad (3.37)$$

Then

$$\begin{pmatrix} u_1 \\ y \end{pmatrix} = \begin{bmatrix} I & 0 & I & 0 \\ P_1(z) & P_2(z) & 0 & I \end{bmatrix} \begin{pmatrix} u_{1,0} \\ u_2 \\ v_1 \\ v_2 \end{pmatrix}. \qquad (3.38)$$

Polynomial solutions of the input estimation problem from closed-loop data

With this, the error spectrum can be expressed as

$$
\Phi_\epsilon(z) = \begin{bmatrix} -(F_1 + F_2 P_1) & (I - F_2 P_2) & -F_1 & -F_2 \end{bmatrix} \cdot
$$
$$
\cdot \begin{pmatrix} T_{u_1 u_2} \Phi_u T_{u_1 u_2}^H + T_{u_1 v_2} \Phi_{v_2} T_{u_1 v_2}^H & T_{u_1 u_2} \Phi_u & 0 & T_{u_1 v_2} \Phi_{v_2} \\ \Phi_u T_{u_1 u_2}^H & \Phi_u & 0 & 0 \\ 0 & 0 & \Phi_{v_1} & 0 \\ \Phi_{v_2} T_{u_1 v_2}^H & 0 & 0 & \Phi_{v_2} \end{pmatrix} \begin{bmatrix} -(F_1 + F_2 P_1)^H \\ (I - F_2 P_2)^H \\ -F_1^H \\ -F_2^H \end{bmatrix} .
$$
$$(3.39)$$

In the expression for Φ_ϵ the only terms that vary with K are $T_{u_1 u_2}$ and $T_{u_1 v_2}$. Without loss of generality, let us assume that $F_1 = -F_2 P_1 - A$, where A is a real rational matrix with appropriate dimensions. With this, we can write equation (3.39) as

$$
\Phi_\epsilon(z) = \begin{bmatrix} A & (I - F_2 P_2) & F_2 P_1 + A & -F_2 \end{bmatrix} \cdot
$$
$$
\cdot \begin{pmatrix} T_{u_1 u_2} \Phi_u T_{u_1 u_2}^H + T_{u_1 v_2} \Phi_{v_2} T_{u_1 v_2}^H & T_{u_1 u_2} \Phi_u & 0 & T_{u_1 v_2} \Phi_{v_2} \\ \Phi_u T_{u_1 u_2}^H & \Phi_u & 0 & 0 \\ 0 & 0 & \Phi_{v_1} & 0 \\ \Phi_{v_2} T_{u_1 v_2}^H & 0 & 0 & \Phi_{v_2} \end{pmatrix} \begin{bmatrix} A^H \\ (I - F_2 P_2)^H \\ (F_2 P_1 + A)^H \\ -F_2^H \end{bmatrix}
$$
$$
= \begin{bmatrix} A & (I - F_2 P_2) & -F_2 \end{bmatrix} \cdot
$$
$$
\cdot \begin{pmatrix} T_{u_1 u_2} \Phi_u T_{u_1 u_2}^H + T_{u_1 v_2} \Phi_{v_2} T_{u_1 v_2}^H & T_{u_1 u_2} \Phi_u & T_{u_1 v_2} \Phi_{v_2} \\ \Phi_u T_{u_1 u_2}^H & 0 & 0 \\ \Phi_{v_2} T_{u_1 v_2}^H & 0 & 0 \end{pmatrix} \begin{bmatrix} A^H \\ (I - F_2 P_2)^H \\ -F_2^H \end{bmatrix} +
$$
$$
+ \begin{bmatrix} (I - F_2 P_2) & F_2 P_1 + A & -F_2 \end{bmatrix} \begin{pmatrix} \Phi_u & 0 & 0 \\ 0 & \Phi_{v_1} & 0 \\ 0 & 0 & \Phi_{v_2} \end{pmatrix} \begin{bmatrix} (I - F_2 P_2)^H \\ (F_2 P_1 + A)^H \\ -F_2^H \end{bmatrix}
$$

Observe that in the last expression for the error spectrum $\Phi_\epsilon(z)$, we have obtained two terms, the first term depends on $T_{u_1 u_2}$ and $T_{u_1 v_2}$, the second term is independent of $T_{u_1 u_2}$ and $T_{u_1 v_2}$. When $A = 0$, the first term is zero, which proves the proposition. ∎

Note that in proposition 3.4.3 only sufficiency of the contraint $F_1 = -F_2 P_1$ is proven. Derivation of necessary conditions for the filter F to be independent of K will result in detailed geometrical conditions following from the structure of P, which is beyond the scope of this work.

Having obtained this result, we can proceed as follows. Finding the optimal estimator that yields equal performance for all K stabilizing P, implies minimizing estimation error under the constraint $F_1 = -F_2 P_1$. Here, the superscript K is used to denote K-independence.

Definition 3.4.4 (Optimal K-independent non-causal input estimator) *The optimal input estimator F_0^K, yielding equal performance for all K stabilizing P, is defined as:*

$$
F_0^K = \arg \min_{\mathbb{F}} \mathbb{E}\epsilon(t)^T \epsilon(t) \text{ subject to } F_1 = -F_2 P_1. \tag{3.40}
$$

Definition 3.4.5 (Optimal K-independent N-causal input estimator) *The N-causal optimal input estimator $F_{\text{caus},N}^K$, yielding equal performance for all K stabilizing P, is defined as:*

$$F_{\text{caus},N}^K = \arg \min_{\mathbb{F}_{\text{caus},N}} \mathbb{E}\epsilon(t)^T \epsilon(t) \text{ subject to } F_1 = -F_2 P_1. \tag{3.41}$$

The solution to these input estimation problems is given by the following two propositions.

Proposition 3.4.6 *Under assumption 3.3.1, the K-independent non-causal optimal input estimator exists and is given by*

$$F_0^K = \Phi_u(z) P_2^H(z) \bar{\Psi}^{-1}(z) \begin{bmatrix} -P_1(z) & I \end{bmatrix} \tag{3.42}$$

with $\bar{\Psi}(z) = P_2(z)\Phi_u(z)P_2^H(z) + P_1(z)\Phi_{v_1}(z)P_1^H(z) + \Phi_{v_2}(z)$. The estimation error spectrum obtained with F_0 is denoted as $\Phi_{\epsilon,0}$ and is given by

$$\Phi_{\epsilon,0}^K = \Phi_u - \Phi_u(z) P_2^H \bar{\Psi}^{-1} P_2 \Phi_u. \tag{3.43}$$

Proof From the proof of proposition 3.4.3 we know that if $F_1 = -F_2 P_1$, the error spectrum is given by

$$\Phi_\epsilon = \begin{bmatrix} (I-F_2 P_2) & F_2 P_1 & -F_2 \end{bmatrix} \begin{pmatrix} \Phi_u & 0 & 0 \\ 0 & \Phi_{v_1} & 0 \\ 0 & 0 & \Phi_{v_2} \end{pmatrix} \begin{bmatrix} (I-F_2 P_2)^H \\ (F_2 P_1)^H \\ -F_2^H \end{bmatrix} \tag{3.44}$$

$$= \begin{pmatrix} F_2 & I \end{pmatrix} \begin{bmatrix} \bar{\Psi} & -P_2\Phi_u \\ -\Phi_u P_2^H & \Phi_u \end{bmatrix} \begin{pmatrix} F_2 & I \end{pmatrix}^H \tag{3.45}$$

with $\bar{\Psi}(z) = P_2(z)\Phi_u(z)P_2^H(z) + P_1(z)\Phi_{v_1}(z)P_1^H(z) + \Phi_{v_2}(z)$. Observe that we can express

$$\bar{\Psi} = \begin{bmatrix} P_2 & P_1 & I \end{bmatrix} \begin{bmatrix} \Phi_u & 0 & 0 \\ 0 & \Phi_{v_1} & 0 \\ 0 & 0 & \Phi_{v_2} \end{bmatrix} \begin{bmatrix} P_2 & P_1 & I \end{bmatrix}^H. \tag{3.46}$$

Noting that the center matrix in this product is positive definite on the unit circle by assumption 3.3.1, and $\begin{bmatrix} P_2(z) & P_1(z) & I \end{bmatrix}$ has full row rank for all $z \in \mathbb{C}$, we infer $\bar{\Psi}(z) > 0$ for all $|z| = 1$. Completion of the squares yields

$$\Phi_\epsilon = \left(F_2 - \Phi_u P_2^H \bar{\Psi}^{-1}\right) \bar{\Psi} \left(F_2 - \Phi_u P_2^H \bar{\Psi}^{-1}\right)^H + \Phi_u - \Phi_u P_2^H \bar{\Psi}^{-1} P_2 \Phi_u. \tag{3.47}$$

This expression essentially divides Φ_ϵ into two parts, the first nonnegative and quadratic in F_2, the other part independent of F_2. It is thus obvious that the optimal $F_{0;2}^K$ is the one that makes the first term equal to zero:

$$F_{0;2}^K(z) = \Phi_u P_2^H \bar{\Psi}^{-1} \tag{3.48}$$

Hence,

$$F_0^K(z) = \Phi_u P_2^H \bar{\Psi}^{-1} \begin{bmatrix} -P_1 & I \end{bmatrix}. \qquad (3.49)$$

To see that $F_0^K(z)$ is analytic on the unit circle note that we can express

$$\bar{\Psi} = D_P^{-1}(N_{P,2}\Phi_u N_{P,2} + N_{P,1}\Phi_{v_1}N_{P,1} + \Phi_{v_1})D_P^{-H} \qquad (3.50)$$

$$= D_P^{-1}\psi D_P^{-H} \qquad (3.51)$$

where we denote $\psi = N_{P,2}\Phi_u N_{P,2} + N_{P,1}\Phi_{v_1}N_{P,1} + \Phi_{v_1}$. Note that ψ is positive definite on the unit circle. With this we can express F_0^K as

$$F_0(z) = \Phi_u N_{P,2}^H \psi^{-1} \begin{bmatrix} -N_{P,1} & D_P \end{bmatrix} \qquad (3.52)$$

from which the analyticity on the unit circle is immediate. The resulting estimation error spectrum follows directly from (3.47) for $F_2 = F_{0;2}^K$:

$$\Phi_{\epsilon,0}^K(z) = \Phi_u(z) - \Phi_u(z)P_2^H(z)\bar{\Psi}^{-1}(z)P_2(z)\Phi_u(z). \qquad \blacksquare \qquad (3.53)$$

It is noteworthy to remark that when P is stable, we could have arrived at expression (3.48) by observing that $F_{0;2}^K$ is the Wiener smoother for u_2 from the signal $\zeta = \begin{bmatrix} -P_1(z) & I \end{bmatrix} \text{col}(u_1, y)$, where $\bar{\Psi}(z)$ has the interpretation of being the spectrum of ζ. In the case P has unstable poles, the result in this proposition tells us that the expression of the optimal non-causal input estimator remains the same, although $\bar{\Psi}(z)$ then no longer has the interpretation of being the spectrum of a physical signal.

Additionally, note that in this proposition positive definiteness of $\Sigma(z)$ on the unit circle is not required, meaning that it also holds in case K has poles on the unit circle, or when P has modes on the unit circle that are uncontrollable from input u_2. We will comment on this property at the end of this section.

The solution for the K-independent N-causal input estimator is given in the following proposition:

Proposition 3.4.7 *Under assumption 3.3.1, the K-independent N-causal optimal input estimator, under the constraint that $F_1 = -F_2 P_1$, exists and is given by*

$$z^{-N}F_{0,N}^K = \{z^{-N}\Phi_u(z)P_2^H(z)\bar{M}^{-H}(z)\}_+\bar{M}^{-1}(z)\begin{bmatrix} -P_1(z) & I \end{bmatrix} \qquad (3.54)$$

where $\bar{M}(z)$ is a causally invertible spectral factor of

$$\bar{\Psi}(z) = P_2(z)\Phi_u(z)P_2^H(z) + P_1(z)\Phi_{v_1}(z)P_1^H(z) + \Phi_{v_2}(z). \qquad (3.55)$$

The estimation error spectrum is

$$\Phi_{\epsilon,0,N}^K = \Phi_u - \Phi_u P_2^H \bar{\Psi}^{-1} P_2 \Phi_u +$$
$$+ \left(\{z^{-N}\Phi_u P_2^H \bar{M}^{-H}\}_-\right)\left(\{z^{-N}\Phi_u P_2^H \bar{M}^{-H}\}_-\right)^H. \qquad (3.56)$$

Proof In the proof of proposition 3.4.6, a quadratic form for the error spectrum was derived in equation (3.47), which we will repeat here for convenience:

$$\Phi_\epsilon = \left(F_2 - \Phi_u P_2^H \bar{\Psi}^{-1}\right) \bar{\Psi} \left(F_2 - \Phi_u P_2^H \bar{\Psi}^{-1}\right)^H + \Phi_u - \Phi_u P_2^H \bar{\Psi}^{-1} P_2 \Phi_u. \qquad (3.57)$$

Positive definiteness of $\bar{\Psi}(z)$ on the unit circle ensures that it is possible to factor $\bar{\Psi}(z)$ as $\bar{\Psi}(z) = \bar{M}(z)\bar{M}^H(z)$ where $\bar{M}(z)$ is causal and $\bar{M}^{-1}(z)$ is analytic on and outside the unit circle. To find the filter $F_{0,N}$ that minimizes Φ_ϵ under the constraint that $z^{-N}F_2$ is causal, we rewrite $\Phi_\epsilon(z)$ as

$$\Phi_\epsilon = \left(z^{-N}F_2\bar{M} - z^{-N}\Phi_u P_2^H \bar{M}^{-H}\right)\left(z^{-N}F_2\bar{M} - z^{-N}\Phi_u P_2^H \bar{M}^{-H}\right)^H +$$
$$+ \Phi_u - \Phi_u P_2^H \bar{\Psi}^{-1} P_2 \Phi_u \qquad (3.58)$$
$$= \left(z^{-N}F_2\bar{M} - \{z^{-N}\Phi_u P_2^H \bar{M}^{-H}\}_+ - \{z^{-N}\Phi_u P_2^H \bar{M}^{-H}\}_-\right) \cdot$$
$$\cdot \left(z^{-N}F_2\bar{M} - \{z^{-N}\Phi_u P_2^H \bar{M}^{-H}\}_+ - \{z^{-N}\Phi_u P_2^H \bar{M}^{-H}\}_-\right)^H$$
$$+ \Phi_u - \Phi_u P_2^H \bar{\Psi}^{-1} P_2 \Phi_u. \qquad (3.59)$$

From this expression it is clear that the optimal causal $z^{-N}F_{0;2}\bar{M}$ is:

$$z^{-N}F_{0;2}\bar{M} = \{z^{-N}\Phi_u P_2^H \bar{M}^{-H}\}_+ \qquad (3.60)$$

and thus

$$z^{-N}F_{0,N}^K = \{z^{-N}\Phi_u P_2^H \bar{M}^{-H}\}_+ \bar{M}^{-1} \begin{bmatrix} -P_1 & I \end{bmatrix}. \qquad (3.61)$$

To see that $z^{-N}F_{0,N}^K$ is stable, first note that D_P is a left divisor of $\bar{M}$ (see equation (3.50)), hence we can express $\bar{M} = D_P^{-1}\check{M}$, with $\check{M}$ stable invertible. From this

$$z^{-N}F_{0,N}^K = \{z^{-N}\Phi_u N_{P,2}^H \check{M}^{-H}\}_+ \check{M}^{-1} \begin{bmatrix} -N_{P,1} & D_P \end{bmatrix} \qquad (3.62)$$

which is clearly stable. Substituting the optimal solution in equation (3.58), immediately yields the optimal estimation error spectrum:

$$\Phi_{\epsilon,0,N}^K = \Phi_u - \Phi_u P_2^H \bar{\Psi}^{-1} P_2 \Phi_u + \left(\{z^{-N}\Phi_u P_2^H \bar{M}^{-H}\}_-\right)\left(\{z^{-N}\Phi_u P_2^H \bar{M}^{-H}\}_-\right)^H. \quad \blacksquare$$
$$(3.63)$$

As indicated in the beginning of this section, there is a degree of conservatism in defining the optimal estimator for the case no controller knowledge is available, as the one that yields minimum estimation error for all possible controllers. Moreover, in our approach we have used a *sufficient* condition ensuring K-independence of the input estimator, constraining the solutions possibly more than strictly necessary, thereby increasing the conservatism. It is therefore worthwile to examine the applicability of the result. For that we compare the error obtained with

the input estimator using full controller knowledge in equation (3.11), and the error spectrum obtained with the non-causal K-independent optimal input estimator in (3.43). From this comparison, we conclude that when the noise spectrum $\Phi_{v_1}(e^{j\omega})$ is small enough so that spectral density of $P_1(e^{j\omega})\Phi_{v_1}(e^{j\omega})P_1^H(e^{j\omega})$ is negligible compared to $\Phi_{v_2}(e^{j\omega})$ for all $\omega \in [0, 2\pi)$, the loss of performance of the K-independent estimator is small compared to the situation exact knowledge of K is available. In many practical applications, the assumption that $\Phi_{v_1}(e^{j\omega})$ is small will be legitimate, for v_1 represents measurement noise of the control input only.

In some cases, exact knowledge will be available on u_1, e.g. when K and F are implemented in the same digital control environment. In these cases $\Phi_{v_1} = 0$, from which we conclude that the input estimation error obtained when designing the filter using knowledge of K will be equal to the input estimation error when designing it under the constraint $F_1 = -F_2 P_1$ and using u_1. However, the estimator that results from the constraint design has a number of advantages. It results in a lower order estimator and controller adaptation will not imply estimator adaptation. Moreover, it can also be used when K has poles on the unit circle, which is obviously the case when K has integral control. This can be understood by noting that Wiener filters are designed without knowledge of the initial condition of the system, meaning that they give optimal performance when the system has reached steady state. When K has poles on the unit circle, it will never reach this state. However, in the output of K also the response of it to the initial state is present. The K-independent input estimator uses this output signal of K as opposed to using knowledge of K alone.

3.4.3 Observability

The propositions in the previous sections formulate solutions to the optimal filter design problems for the causal and non-causal filter problem, both with and without controller knowledge. The performance of these estimators can be poor if one or more of the poles of $\Phi_u(z)$ are cancelled by the zeros of $P_2(z)$. When this is this case, the effect is that part of the dynamics of input process does not appear in the measured output. This is equivalent to saying that part of the states of the input process are unobservable in the measured output. Hence, having no information on these states in our measurements, this may lead to large errors in the estimates of these states.

In this section we will analyze the effects of this in some more detail. We will do this for the non-causal filter design problem with controller knowledge, and when $K(z)$ is stable. The analysis can be extended for the other problems. For convenience, we repeat the solution of the optimal non-causal filter with full controller knowledge, equation (3.10):

$$F_0(z) = \begin{bmatrix} 0 & \Phi_u(z)P_2^H(z)\Psi^{-1}(z)S(z)^{-1} \end{bmatrix} \qquad (3.64)$$

with $\Psi(z) = P_2(z)\Phi_u(z)P_2^H(z) + \Phi_{v_2}(z)$. To analyze possible pole-zero cancellations between Φ_u and P_2, let us define the spectral factorization $\Phi_u = M_u M_u^H$, and

let $M_u(z) = D_u(z^{-1})^{-1} N_u(z^{-1})$ be an irreducable left MFD of $M_u(z)$. With this we can write for the first factor of $F_{0;2}$

$$\Phi_u P_2^H = D_u^{-1} N_u N_u^H D_u^{-H} \cdot \bar{N}_{P,2}^H \bar{D}_P^{-H}.$$

The aforementioned pole-zero cancellation is equivalent to D_u and $\bar{N}_{P,2}$ having a common factor, that is, there exists a non-unimodular polynomial matrix $W_u(\xi)$ satisfying $D_u(\xi) = \check{D}_u(\xi) W_u(\xi)$ and $\bar{N}_{P,2}(\xi) = \check{N}_{P,2}(\xi) W_u(\xi)$ such that $\check{X}_u$ and $\check{N}_{P,2}$ are right-coprime. Substituting these identities in the optimal filter solution (3.64), it follows that this common factor W_u is a left divisor of the optimal filter:

$$F_{0;2} = W_u^{-1} \check{F}_{0;2} \tag{3.65}$$

where $\check{F}_{0;2}$ is defined as

$$\check{F}_{0;2} = \check{M}_u \check{M}_u^H \check{P}_2^H(z) \check{\Psi}^{-1}(z) S(z)^{-1}$$

with $\check{M}_u = \check{D}_u^{-1} N_u$, $\check{P}_2 = \bar{D}_P^{-1} \check{N}_{P,2}$, and $\check{\Psi} = \check{P}_2 \check{M}_u \check{M}_u^H \check{P}_2^H + \Phi_{v_2}$. Let us interpret this result. For this, define $\check{u}_2(t) = \check{M}_u(z) w(t)$, where $w(t)$ is a white noise process with unit covariance, so that $u_2(t) = W_u^{-1}(z) \check{u}(t)$. Due to the pole-zero cancellation, the states of $W_u^{-1}(z)$ are unobservable from y, making estimation of $u_2(t)$ directly from y impossible. The interpretation of equation (3.65) is that in such cases the optimal filter (3.64) obtains the optimal estimate of $\mathring{u}_2(t)$, and subsequently filters this by the inverse of the unobservable part of the input process, i.e. $W_u^{-1}(z)$. The result can be quite undesirable. Since at all frequencies both the signal and noise components of $\mathring{u}_2(t)$ are multiplied by the same gain, the effect can lead to unwanted amplification of the noise in the estimate.

To avoid the abovementioned pole-zero cancellation, the following additional assumption is made henceforth.

Assumption 3.4.8 *It is assumed that $\bar{N}_{P,2}(z)$ is right prime, i.e. it has full column rank for all $z \in \mathbb{C}$.*

Assumption 3.4.8 implies right-coprimeness of D_u and $\bar{N}_{P,2}$ for any possible input spectrum, meaning that no pole-zero cancellations between the input process and the dynamics of P_2 can happen. Observe that for certain given input spectrum $\Phi_u(z)$, this assumption is conservative. Necessary and sufficient for the states of $M_u(z)$ to be observable in $\mathrm{col}(u_1, y)$ is that $\bar{N}_{P,2}$ and D_u are right-coprime. Also observe that assumption 3.4.8 implies that $p \geq m_2$, the number of outputs of p should be larger than or equal to the number of unknown inputs.

By a similar argument we will assume no pole-zero cancellations will happen in the product $P_1 \Phi_{v_1} P_1^H$ that appears in the solution of the optimal filters for the case no controller knowledge case is used, see equation (3.42) and (3.95). Sufficient for this is that the following assumption holds:

Assumption 3.4.9 *It is assumed $N_{P,1}(z)$ is right prime, i.e. it has full column rank for all $z \in \mathbb{C}$.*

Similarly, this assumption implies that for any noise process v_1, no pole-zero cancellations between between noise process v_1 and the dynamics of P_1 will happen.

3.5 State-space solutions to the causal input estimation problems

In the preceding sections, general solutions for the N-causal optimal input estimators was formulated for the case full information on K is available, as well for the case that no explicit knowledge of K is at hand. Central in these solutions is the computation of a spectral factorization. The problem of spectral factorization has received considerable attention in the past 40 years. An overview of the resulting techniques is given in [140]. Most of the techniques are rather involved for the multivariable case that we consider. However, when the transfer matrices have a state-space structure, it is well known that the factorization can be computed efficiently by solving a Discrete-time Algebraic Ricatti Equation (DARE). In Section 3.5.1 we will recall this result and subsequently apply this to formulate explicit solutions of the N-causal optimal input filters in Sections 3.5.2 (controler knowledge available) and 3.5.3 (K-independent estimator).

3.5.1 Spectral factorization using state space structure

In this section we will give two relevant propositions, which allow us to construct the spectral factorization for systems in state-space form.

Proposition 3.5.1 *Consider a spectrum $\Phi(z)$ of the form*

$$\Phi_y(z) = \begin{bmatrix} C(zI-A)^{-1} & I \end{bmatrix} \begin{bmatrix} BQB^T & BS \\ S^TB^T & R_v \end{bmatrix} \begin{bmatrix} (z^{-1}I-A^T)^{-1}C^T \\ I \end{bmatrix} \qquad (3.66)$$

with (C,A) detectable,

$$\begin{bmatrix} BQB^T & BS \\ S^TB^T & R_v \end{bmatrix} \geq 0 \text{ and } R_v > 0. \qquad (3.67)$$

Define

$$A^s = A - BSR_v^{-1}C \text{ and } Q^s = Q - SR_v^{-1}S^T. \qquad (3.68)$$

Then $\Phi(e^{j\omega}) > 0$ for all $\omega \in [0, 2\pi)$ if, and only if, the matrix pair $(A^s, B(Q^s)^{1/2})$ is unit circle controllable, i.e $\begin{bmatrix} \lambda I - A^s & B(Q^s)^{1/2} \end{bmatrix}$ has full rank for all $|\lambda| = 1$.

3. Unknown input estimation from closed-loop data

Proof See [83]. ∎

Proposition 3.5.2 *Consider a spectrum $\Phi(z)$ of the form*

$$\Phi(z) = \left[\begin{array}{cc} C(zI - A)^{-1} & I \end{array}\right] \left[\begin{array}{cc} BQB^T & BS \\ S^T B^T & R_v \end{array}\right] \left[\begin{array}{c} (z^{-1}I - A^T)^{-1}C^T \\ I \end{array}\right] \tag{3.69}$$

with $\{A, B, C, Q, R_v, S\}$ satisfying the conditions as stated in proposition 3.5.1. Then a spectral factorization can be obtained as

$$\Phi_y(z) = M(z)RM^*(z), \quad M(\infty) = I, \quad R > 0 \tag{3.70}$$

where

$$M(z) = I + C(zI - A)^{-1}L \tag{3.71}$$

$$M^{-1}(z) = I - C(zI - A + LC)^{-1}L \tag{3.72}$$

$$L = (A\Pi C^T + BS)R^{-1} \tag{3.73}$$

$$R = R_v + C\Pi C^T \tag{3.74}$$

and Π is the unique positive semi definite solution to the discrete algebraic Ricatti equation (DARE)

$$\Pi = A\Pi A^T + BQB^T - LRL^T \tag{3.75}$$

such that $A - LC$ is stable, which will guarantee that $M(z)$ is minimum phase, i.e. $M^{-1}(z)$ is analytic for $z \notin \mathbb{C}^-$.

Proof See [83]. ∎

If in addition $(A^s, B(Q^S)^{1/2})$ is stabilizable, then Π is also the unique positive semi-definite solution of the DARE. Furthermore, if $(A^s, B(Q^S)^{1/2})$ is controllable, then Π is the unique positive definite solution of the DARE [83].

Observe that stability of A is not required in this theorem. If all conditions are met, but A is unstable, then the spectral factor M will also be unstable. Nevertheless, M^{-1} is stable and causal. In deriving the solutions for inthe input estimator, we have not used stability of the spectral factor. Hence, the above method for calculating a causally invertible spectral factor can be applied even in the case the matrix A is unstable. This is the case when dealing with the AMB spindle dynamics, as will become clear in the next section.

3.5.2 Solution of the N-causal input estimators in state space form, K known

In this section, an explicit solution for the N-causal input estimator as derived in proposition 3.4.2 will be given, using state space formulations for the transfer function matrices. We will do this for the case $K(z)$ is stable, in which case the optimal

State-space solutions to the causal input estimation problems

filter has the form in equation (3.27), which we repeat here for convenience:

$$z^{-N}F_{0,N}(z) = \begin{bmatrix} 0 & \{z^{-N}\Phi_u(z)P_2^H(z)M^{-H}(z)\}_+ M^{-1}(z)S(z)^{-1} \end{bmatrix} \qquad (3.76)$$

where $M(z)$ is a causally invertible spectral factor of $\Psi(z) = P_2(z)\Phi_u(z)P_2^H(z) + \Phi_{v_1}(z)$. The derivation will be performed in three main steps:

1. We will introduce state space realizations for the plant P, and for modeling filters for the input u_2 and the noise processes v_1 and v_2. This allows us to give an expression for $\Psi(z)$ using a state representation;

2. Using this state representation of $\Psi(z)$ and the factorization theorem in proposition 3.5.2, an explicit expression for $M^{-1}(z)$ is derived;

3. The causal part in equation (3.76) is calculated, yielding the desired result.

We start with step 1. Let us assume P admits the following state space realization

$$\begin{aligned} x(t+1) &= A_P x(t) + B_{P,1}u_{1,0}(t) + B_{P,2}u_2(t) \\ y_0(t) &= C_P x(t) \end{aligned} \qquad (3.77)$$

where $x \in \mathbb{R}^n$ is the state vector. Matrices A_P, $B_{P,1}$, $B_{P,2}$ and C_P are known matrices with appropriate dimensions. Matrix A_P is unstable, due the instrinsic instability of the AMB dynamics. Without loss of generality it is assumed that $\operatorname{rank}(B_{P,2}) = m_2$ and $\operatorname{rank}(C_P) = p$. Furthermore, it is assumed that (A_P, C_P) is observable and $(A_P, B_{P,1})$ is controllable[2]. The pair $(A_P, B_{P,2})$ is not necessarily controllable. It will be assumed however that $(A_P, B_{P,2})$ is controllable on the unit circle, i.e.

$$\operatorname{rank}\begin{bmatrix} A_P - \lambda I & B_{P,2} \end{bmatrix} = n \quad \forall |\lambda| = 1. \qquad (3.78)$$

From this it follows that there exists a similarity transformation such that we can equivalently express the state space system in equation (3.79) as

$$\begin{aligned} x(t+1) &= \begin{bmatrix} A_P^{11} & A_P^{12} \\ 0 & A_P^{22} \end{bmatrix} x(t) + \begin{bmatrix} B_{P,1}^1 \\ B_{P,2}^2 \end{bmatrix} u_{1,0}(t) + \begin{bmatrix} B_{P,2}^1 \\ 0 \end{bmatrix} u_2(t) \\ y_0(t) &= \begin{bmatrix} C_P^1 & C_P^2 \end{bmatrix} x(t) \end{aligned} \qquad (3.79)$$

with $(A_P^{11}, B_{P,2}^1)$ controllable and A_P^{22} no eigenvalues on the unit circle. Furthermore, it is assumed the triplets $(A_P^{11}, B_{P,2}^1, C_P^1)$ and $(A_P, B_{P,1}, C_P)$ have no invariant zeros, i.e.

$$\operatorname{rank}\begin{bmatrix} A_P^{11} - \lambda I & B_{P,2}^1 \\ C_P^1 & 0 \end{bmatrix} = n_1 + m_2 \quad \forall \lambda \in \mathbb{C} \qquad (3.80)$$

---

[2]Observe that requiring (A_P, C_P) is detectable and $(A_P, B_{P,1})$ is stabilizable would suffice. However, this would allow the realization to have stable unobservable and/or uncontrollable modes. Including the possiblity of such modes in the treatment would not add any added value, but would certainly complicate the analysis.

with n_1 the dimension of A_P^{11}, and

$$\text{rank} \begin{bmatrix} A_P - \lambda I & B_{P,1} \\ C_P & 0 \end{bmatrix} = n + m_1 \quad \forall \lambda \in \mathbb{C}. \tag{3.81}$$

Observe that (3.78) ensures that in equation (3.23) the common factor $W(z^{-1}) > 0$ for $|z| = 1$, while equations (3.80) and (3.81) are equivalent to assumptions 3.4.8 and 3.4.9 respectively.

Given assumption 3.3.1, we can obtain the spectral factorization of $\Phi_u(z)$ by a stable, minimal and strictly proper[3] LTI system P_u with state space realization given by

$$\begin{aligned} \xi(t+1) &= A_u \xi(t) + B_u w(t) \\ u_2(t) &= C_u \xi(t) \end{aligned} \tag{3.82}$$

with state vector $\xi \in \mathbb{R}^{n_u}$, and with $w(t)$ a white noise process with covariance $R_u > 0$, such that

$$\Phi_u(z) = P_u(z) R_u P_u^H(z). \tag{3.83}$$

Here B_u has full column rank, and C_u has full row rank. Similarly, for $i = 1, 2$, there exists a stable, minimal LTI system P_{v_i} with state space realization given by

$$\begin{aligned} \zeta_i(t+1) &= A_{v_i} \zeta_i(t) + B_{v_i} e_i(t) \\ v_i(t) &= C_{v_i} \zeta_i(t) + e_i(t) \end{aligned} \tag{3.84}$$

with state vector $\zeta_i \in \mathbb{R}^{n_{v_i}}$, and $e_i(t)$ a white noise process with covariance $R_{v_i} > 0$, such that

$$\Phi_{v_i}(z) = P_{v_i}(z) R_{v_i} P_{v_i}^H(z) \tag{3.85}$$

is the canonical spectral factorization of Φ_{v_i}. Here B_{v_i} has full column rank, and C_{v_i} has full row rank.

In order to write $\Psi(z)$ as defined in proposition 3.4.2 in the form of proposition 3.5.2, let $z = \text{col}(\xi, x, \zeta_1)$. Combining equation (3.79) with (3.82) and (3.84) yields the cascaded system P_c, that admits the state-space representation

$$\begin{aligned} z(t+1) &= A_c z(t) + B_{c,1} u_1(t) + B_{c,2} \begin{pmatrix} w(t) \\ e_2(t) \end{pmatrix} \\ y(t) &= C_c z(t) + D_{c,2} \begin{pmatrix} w(t) \\ e_2(t) \end{pmatrix} \end{aligned} \tag{3.86}$$

---

[3] We remark that it is more common to model the dynamics of a random process by a monic proper system driven by a white noise process. However, without any loss of generality it is always possible to obtain a model with a strictly proper system instead. The advantage of such a choice is that the output depends on the states of the system only and not on the driving white noise process. This choice simplifies the estimator.

with

$$A_c = \begin{bmatrix} A_u & 0 & 0 \\ B_{P,2}C_u & A_P & 0 \\ 0 & 0 & A_{v_2} \end{bmatrix}, \qquad B_{c,1} = \begin{bmatrix} 0 \\ B_{P,1} \\ 0 \end{bmatrix}, \qquad B_{c,2} = \begin{bmatrix} B_u & 0 \\ 0 & 0 \\ 0 & B_{v_2} \end{bmatrix}$$

$$C_c = \begin{bmatrix} 0 & C_P & C_{v_2} \end{bmatrix}, \qquad\qquad D_{c,2} = \begin{bmatrix} 0 & I \end{bmatrix}.$$

Define $P_{c,2}(z) = C_c(zI - A_c)^{-1}B_{c,2} + D_{c,2}$, and let

$$Q = \begin{bmatrix} R_u & 0 \\ 0 & R_{v_2} \end{bmatrix}. \tag{3.87}$$

Observe that $\Psi(z) = P_{c,2}(z)QP_{c,2}^H(z)$.

This brings us to step 2. With the obtained state space representation of $P_{c,2}$ we can compute a causally invertible spectral factor of Ψ, as expressed in the following proposition.

Proposition 3.5.3 *Given $\bar{A}_c$, $B_{c,2}$, C_c and Q as defined in (3.86) and (3.87). Let $M(z)$ satisfy*

$$M^{-1}(z) = I - C_c(zI - \bar{A}_c)^{-1}L, \ where \tag{3.88}$$

$$L = (A_c \Pi C_c^T + B_{c,2}S)R^{-1} \tag{3.89}$$

$$R = R_{v_2} + C_c \Pi C_c^T \tag{3.90}$$

and Π is the unique positive semi-definite solution to the discrete algebraic Ricatti equation (DARE)

$$\Pi = A_c \Pi A_c^T + B_{c,2}QB_{c,2}^T - LRL^T \tag{3.91}$$

such that $\bar{A}_c = A_c - LC_c$ is stable. Then $\Psi = MRM^H$.

Proof See appendix B.1.1. ∎

Observe that for convenience we use the normalized factorization $\Psi = MRM^H$ with $M(\infty) = I$. With this, the solution of the N-causal input estimation problem can be expressed as

$$z^{-N}F_{0,N} = \begin{bmatrix} 0 & \{z^{-N}\Phi_u(z)P_2^H(z)M^{-H}(z)\}_+ R^{-1}M^{-1}(z)S^{-1}(z) \end{bmatrix}. \tag{3.92}$$

The last step is to evaluate the causality operator in (3.92). The result, after some algebraic manipulation to simplify the result, is summarized the following proposition.

3. Unknown input estimation from closed-loop data

Proposition 3.5.4 *Given the state space representation of plant P in (3.79), the input model realization in (3.82), and the noise modeling filter realization of v_2 in (3.84). With this, let $B_{c,1}$ and C_c be defined as in (3.86), and let $\bar{A}_c$, R, Π and L follow from the spectral factorization in proposition 3.5.3. Then the N-causal optimal input estimator is given by*

$$z^{-N}F_{0,N}(z) = \begin{bmatrix} 0 & F_{0,N;2a}(z) + F_{0,N;2b}(z) \end{bmatrix} \tag{3.93}$$

with

$$F_{0,N;2a}(z) = z^{-N} \begin{bmatrix} C_u & 0 & 0 \end{bmatrix}(zI - \bar{A}_c)^{-1}(L + B_{c,1}K(z))$$

and

$$F_{0,N;2b}(z) = \begin{bmatrix} C_u & 0 & 0 \end{bmatrix}\Pi \sum_{m=1}^{N+1} z^{m-N-1}(\bar{A}_c^T)^{m-1}C_c^T R^{-1}\cdot$$
$$\cdot (I - C_c(zI - \bar{A}_c)^{-1}(L + B_{P,1}K(z)).$$

Proof See appendix B.1.2.

Observe that this filter has the structure of a smoothed state estimator of the augmented system P_c under feedback of K, where the input estimate (i.e. the output of the filter) is constructed from the estimated states of the input process.

3.5.3 Solution of the K-independent N-causal input estimator in state space form

Similar to the previous section, in this section an explicit solution for the K-independent N-causal input estimator as derived in proposition 3.4.7 will be given using state space formulations for the transfer function matrices. Again, for convenience, we repeat the optimal estimator for this case:

$$z^{-N}F_{0,N}^K = \{z^{-N}\Phi_u(z)P_2^H(z)\bar{M}^{-H}(z)\}_+ \bar{M}^{-1}(z) \begin{bmatrix} -P_1(z) & I \end{bmatrix} \tag{3.94}$$

where $\bar{M}(z)$ is a causally invertible spectral factor of

$$\bar{\Psi}(z) = P_2(z)\Phi_u(z)P_2^H(z) + P_1(z)\Phi_{v_1}(z)P_1^H(z) + \Phi_{v_2}(z). \tag{3.95}$$

The approach follows the same three steps as in the previous section. We begin with step 1., in which we express $\bar{\Psi}(z)$ using a state space realization. For that we introduce the compound system P_c, with realization given by

$$z(t+1) = A_c z(t) + B_{c,1}u_1(t) + B_{c,2}\begin{pmatrix} w(t) \\ e_1(t) \\ e_2(t) \end{pmatrix}$$

$$\bar{y}(t) = C_c z(t) + D_{c,2}\begin{pmatrix} w(t) \\ e_1(t) \\ e_2(t) \end{pmatrix} \tag{3.96}$$

where

$$A_c = \begin{bmatrix} A_u & 0 & 0 & 0 \\ B_{P,2}C_u & A_P & B_{P,1}C_{v_1} & 0 \\ 0 & 0 & A_{v_1} & 0 \\ 0 & 0 & 0 & A_{v_2} \end{bmatrix},$$

$$B_{c,1} = \begin{bmatrix} 0 \\ B_{P,1} \\ 0 \\ 0 \end{bmatrix}, \qquad\qquad B_{c,2} = \begin{bmatrix} B_u & 0 & 0 \\ 0 & B_{P,1} & 0 \\ 0 & B_{v_1} & 0 \\ 0 & 0 & B_{v_2} \end{bmatrix},$$

$$C_c = \begin{bmatrix} 0 & C_P & 0 & C_{v_2} \end{bmatrix}, \qquad\qquad D_{c,2} = \begin{bmatrix} 0 & 0 & I \end{bmatrix}.$$

In this expression the realizations (A_u, B_u, C_u), $(A_{v_i}, B_{v_i}, C_{v_i})$, $i = 1, 2$ and $(A_P, [B_{P,1}B_{B,2}], C_P)$ are as defined in the previous section. Observe that P_c models the output of P when instead of the true input $u_{1,0}$, the noise disturbed measurement u_1 is used. Define $P_{c,2}(z) = C_c(zI - A_c)^{-1}B_{c,2} + D_{c,2}$, and let

$$Q = \begin{bmatrix} R_u & 0 & 0 \\ 0 & R_{v_1} & 0 \\ 0 & 0 & R_{v_2} \end{bmatrix}, \quad S = \begin{bmatrix} 0 & 0 & R_{v_2} \end{bmatrix}^T. \tag{3.97}$$

With this we can express $\bar\Psi(z) = P_{c,2}(z) Q P_{c,2}^H(z)$.

Proceeding to step 2, we compute a causally invertible spectral factor of $\bar\Psi$, using the obtained state space representation of $P_{c,2}$:

Proposition 3.5.5 *Given* $\{A_c, B_{c,2}, C_c, Q, R_{v_2}, S\}$ *as defined in equations* (3.85), (3.96), *and* (3.97). *Let* $\bar M(z)$ *satisfy*

$$\bar M^{-1}(z) = I - C_c(zI - \bar A_c)^{-1}L, \text{ where} \tag{3.98}$$

$$L = (A_c \Pi C_c^T + B_{c,2}S)\bar R^{-1} \tag{3.99}$$

$$\bar R = R_{v_2} + C_c \Pi C_c^T \tag{3.100}$$

and Π *is the unique positive semi-definite solution to the discrete algebraic Ricatti equation (DARE)*

$$\Pi = A_c \Pi A_c^T + B_{c,2} Q B_{c,2}^T - L \bar R L^T \tag{3.101}$$

such that $\bar A_c = A_c - LC_c$ *is stable. Then* $\bar\Psi = \bar M \bar R \bar M^H$.

Proof See appendix B.1.3. ∎

The last step again involves computation of the causal part in equation (3.94), and the result of that is given the following proposition.

Proposition 3.5.6 *Given the state space representation of plant P in (3.79), the input model realization in (3.82), and the noise modeling filter realizations of v_1 and v_2 in (3.84). With this, let $B_{c,1}$ and C_c be defined as in (3.96), and let $\bar{A}_c$, $\bar{R}$, Π and L follow from the spectral factorization in proposition 3.5.5. Then the K-independent N-causal optimal input estimator is given by*

$$z^{-N} F_{0,N}^K(z) = \begin{bmatrix} 0 & F_{0,N;2a}^K(z) + F_{0,N;2b}^K(z) \end{bmatrix}. \tag{3.102}$$

with

$$F_{0,N;2a}^K(z) = z^{-N} \begin{bmatrix} C_u & 0 & 0 & 0 \end{bmatrix} (zI - \bar{A}_c)^{-1} \begin{bmatrix} B_{c,1} & L \end{bmatrix}$$

and

$$F_{0,N;2b}^K(z) = \begin{bmatrix} C_u & 0 & 0 & 0 \end{bmatrix} \Pi \cdot$$
$$\cdot \sum_{m=1}^{N+1} z^{m-N-1} (\bar{A}_c^T)^{m-1} C_c^T \bar{R}^{-1} \left(\begin{bmatrix} 0 & I \end{bmatrix} - C_c(zI - \bar{A}_c)^{-1} \begin{bmatrix} B_{c,1} & L \end{bmatrix} \right).$$

Proof See appendix B.1.4.

Observe that if $v_1 = 0$, then $u_{1,0} = Ky$. Using this in proposition 3.5.4 would immediately give the above result (observe that in this case $\dim(\zeta_1) = 0$).

3.6 Performance measures

In the previous sections we have derived explicit expressions for the input estimator $F(z)$. In this section we will discuss different measures that can be used to evaluate the performance of the estimators.

In line with Seron et al. [150], we will define the following sensitivities:

$$S_F : u_2 \to \epsilon \qquad \text{filtering sensitivity} \tag{3.103}$$
$$T_F : u_2 \to \hat{u}_2 \qquad \text{filtering complementary sensitivity} \tag{3.104}$$
$$S_V : v \to \epsilon \qquad \text{noise sensitivity} \tag{3.105}$$

with $v = \text{col}(v_1, v_2)$. With this, we can express

$$\epsilon(t) = \underbrace{S_F u_2(t)}_{:=\epsilon_u(t)} + \underbrace{S_V v(t)}_{:=\epsilon_v(t)}. \tag{3.106}$$

The filtering sensitivity represents the relative effect of the unknown input on the estimation error. On the other hand, the complementary sensitivity represents the relative effect of the unknown input on the estimate. The noise sensitivity function describes the relative effect of the noise disturbances on the estimate.

The amplitude response of the filtering complementary sensitivity T_F measures the transfer function from u_2 to its estimate $\hat{u}_2$. Hence, it is desired to achieve $|T_F(e^{j\omega})|$ close to identity in the frequency ranges where the spectral density of u_2 is large. In general (but depending on the spectral density of u_2) the filtering complementary sensitivity will have a low-pass characteristic, implying that the unknown input can only be estimated up to a certain frequency limit.

Studying the phase response of the filtering complementary sensitivity, particularly in the areas where it is close to the identity, will provide information regarding the phase distortion in the estimate. For that purpose it is common to calculate the *group delay*, being defined as the negative first order derivative of the argument of $T_F(e^{j\omega})$ with respect to the frequency ω [120],

$$\tau_F(\omega) := -\frac{d}{d\omega}\arg(T_F(e^{j\omega})) \tag{3.107}$$

with $\tau_F(\omega)$ in radians. When T_F has constant group delay in its pass band, this indicates a linear phase response, which in turn implies that T_F in this frequency range acts as a pure delay. By dividing τ_F by the sample time, a value for this delay as a (fractional) multiple of the sample time can be obtained.

3.7 Simulation example

The input estimation approach has been tested in simulation using Matlab/Simulink. A simulation model has been developed based on the properties of the EAAT AMB spindle setup discussed in Chapter 2. System P resulted from the therein described finite element modeling approach, which after balanced truncation [118] and discretization with $T = 50\mu s$ yielded a model with $n = 24$, $m_1 = 4$ (current inputs), $m_2 = 2$ (cutting force inputs in X and Y direction) and $p = 4$ (displacement outputs). Current measurement noise v_1 was assumed negligible, position measurement noise v_2 was modeled as a white noise with diagonal covariance matrix $R_v = \sigma_v^2 I$, with $\sigma_v = 10^{-7}m$. In the simulation, a PID controller was applied to stabilize each of the four magnetic bearings. The input estimators discussed in this section however, do not use the information on the controller, but were designed as described in Section 3.5.3. For the cutting forces, a waveform was chosen in accordance with a model describing the cutting forces when milling with a micro-endmill with two teeth [45].

We will discuss the simulation results for two cases. First, for a simple model for the input force, we will compare the results obtained with an estimator with and without delay. As the rotational speeds increase, estimation results improve if the spectral information on the cutting forces is used when designing estimators. This will be investigated in the second test.

3. UNKNOWN INPUT ESTIMATION FROM CLOSED-LOOP DATA
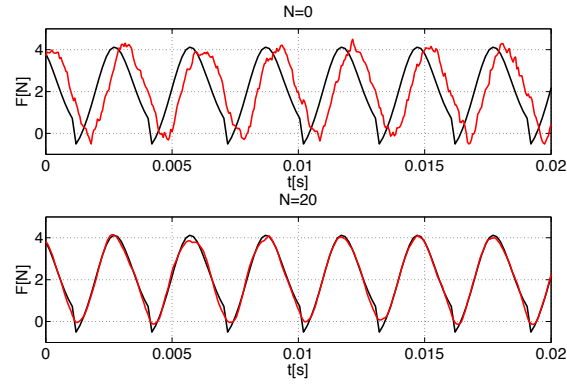
(a) X-direction (first element of u_1)
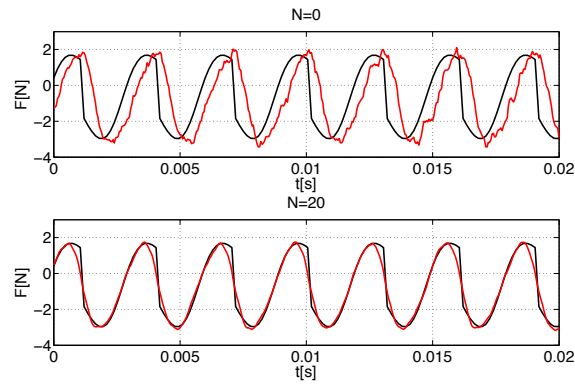
(b) Y-direction (second element of u_1)

Figure 3.2: Input estimation results for the cutting force for $N = 0$ and $N = 20$, in X and Y direction. Black: true input u_2, red: estimated input $\hat{u}_2$. The estimator with $N = 0$ shows a delay. The estimator $N = 20$ has no (additional) delay, and gives smaller estimation error.

(a) $|T_{F(1,1)}|$ (filtering complementary sensitivity)

(b) $|S_{F(1,1)}|$ (filtering sensitivity)

(c) $|S_{v(1,1)}|$ (noise sensitivity)
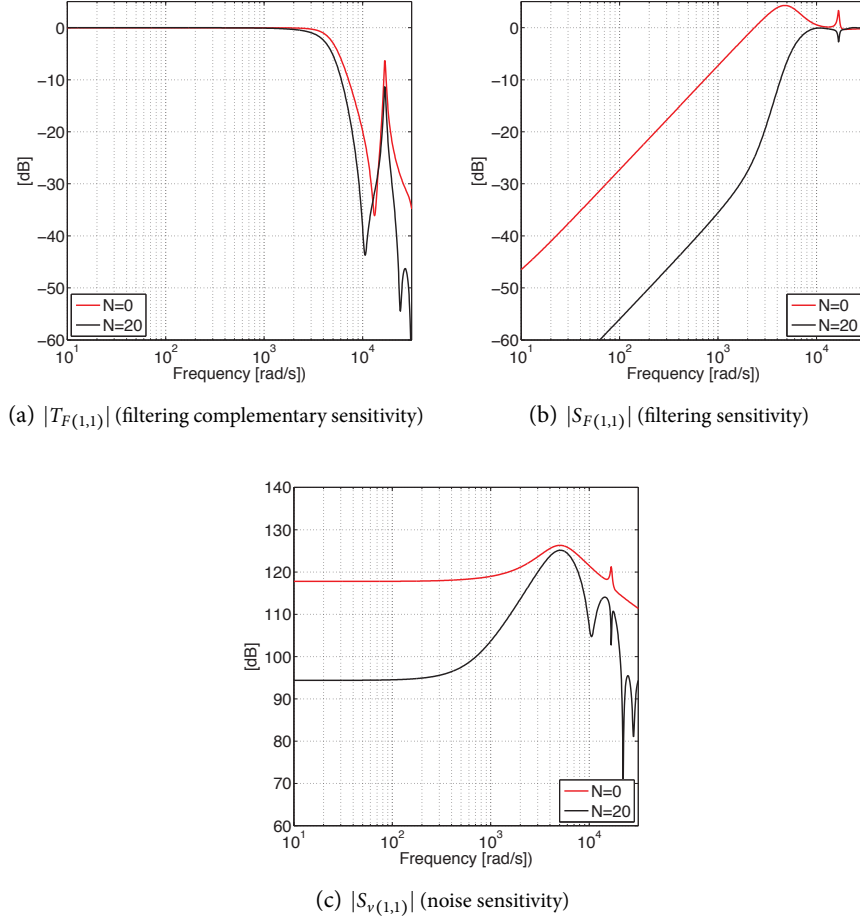
Figure 3.3: Amplitude frequency response of element $(1,1)$ of the filtering complementary sensitivity (a), filtering sensitivity (b), and the noise sensitivity (c) for $N = 0$ and $N = 20$. The delayed estimator performs better in the sense that the relative effect of the unknown input on the estimation error is smaller (b), while the relative effect of the noise disturbance on the estimate is also smaller (c). The filtering complementary sensitivities (a) of both estimators are very similar, although the delayed filter falls off at a slightly lower frequency.

3.7.1 Test 1: Input observers with $N = 0$ and with $N > 0$

In this test for the input a simple model is chosen, i.e. a random walk-like model given by

$$\xi(t+1) = 0.9999\xi(t) + w(t), \qquad u_2(t) = \xi(t) \tag{3.108}$$

so that $A_u = 0.9999 \cdot I$, $B_u = C_u = I$. We note that such models are often used in input estimation and all of the indirect cutting force estimation techniques referenced in Chapter 1 use this rather limited model. An estimator has been implemented with no delay ($N = 0$) and one with a delay of $N = 20$ time steps. The result of both estimators is depicted in Figure 3.2, where the result of the delayed estimator has been shifted by 20 time steps in order to compare it with the input signal.

We observe a number of things. A first difference is that the result of the estimator for $N = 0$ is more noisy than that of $N = 20$. This can be explained by Figure 3.3. For both $N = 0$ and $N = 20$, in this figure the amplitude frequency response of element (1,1) of the complementary filtering sensitivity T_F, the filtering sensitivity S_F and the noise sensitivity S_V are shown. These figures show that the delayed estimator performs better in the sense that the relative effect of the unknown input on the estimation error is smaller, while the relative effect of the noise disturbance on the estimate is also smaller. A more striking difference is that the estimator for $N = 0$ has a delay of around 9 time steps, whereas the estimator for $N = 20$ has no extra delay. It is well known that a filter representing a pure delay has a linear phase characteristic, and hence has constant group delay. In Figure 3.4 we have depicted the group delay of $T_{F(1,1)}$ for $N = 0$ for different levels of the position measurement noise disturbance. In this figure it appears that the group delay is close to constant over the frequency range in which $T_{F(1,1)}$ is close to 1, supporting the observation that there is a delay present in the input estimates. For $\sigma_v = 10^{-7}$, $\tau_F = 8.6$ time steps at low frequencies, which also corresponds with the observed delay in Figure 3.2.

It is remarkable that the filter that was designed not to have a lag (i.e. $N = 0$) still produces delayed estimates. In this perspective the method discussed in this chapter has the favorable property that a filter can be designed with a lag close to this delay, resulting in better estimates. This can be observed from Figure 3.5. In Figure 3.5(a) a plot is made of $\sigma_\epsilon = \sqrt{\mathbb{E}\epsilon^2}$ for increasing lag N, again for different levels of the noise disturbance on the position measurements. As expected, σ_ϵ descreases monotonely with N. The value of σ_ϵ at $N = 8$ is clearly much lower than at $N = 0$. To demonstrate that this decrease is not just the result of a smaller delay in the estimate (resulting in σ_u to be smaller), but also that a reduction of the noise in the estimate is reduced, plots of $\sigma_{\epsilon_u} = \sqrt{\mathbb{E}\epsilon_u^2}$ and $\sigma_{\epsilon_v} = \sqrt{\mathbb{E}\epsilon_v^2}$ are provided in Figures 3.5(b) and 3.5(c) respectively. From 3.5(c) we observe that $\sigma_{\epsilon_v} = 0.24$ N at $N = 0$, $\sigma_{\epsilon_v} = 0.19$ N ($= -19\%$) at $N = 8$. Increasing N further results in an even lower noise disturbance in the estimate, i.e. $\sigma_{\epsilon_v} = 0.11$ N ($= -54\%$) at $N = 20$.

In summary, from this simulation we conclude that with the method in this chapter a filter designed without delay, still has a delay. Designing a filter with the same delay results in better estimates. A filter with a delay that is slightly higher yields significant reduction of the estimation error.
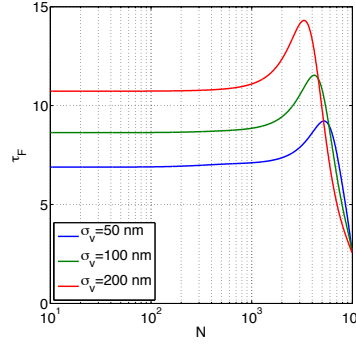
Figure 3.4: Group delay of $T_{F(1,1)}$ for $N = 0$ for different levels of the position measurement disturbance v_2. In the low frequency range the group delay is constant, supporting the observation that there is delay present in the transfer function from $u_2 \rightarrow \hat{u}_2$.

3.7.2 Test 2: Improved stochastic model for the input

The simulation described in the previous section is performed at a relatively low rotational speed. At higher rotational speeds, i.e. at $50,000$ rpm, it is observed that estimation errors made by estimator with $N = 0$ are quite large. Obviously increasing the rotational frequency of the AMB spindle results in cutting forces with higher frequencies. As with any mechanical system, the response of the AMB system decreases as the exciting frequencies increase, resulting in a decreased signal to noise ratio. Hence, we can improve the estimation result if we use the a priori information on the spectral content of the cutting force signal. To verify this, a spectral model for the input has been chosen that has high power in the low frequencies regions and incorporates peaks at the first three harmonics of the cutting force signal (i.e. at 1.7kHz, 3.3kHz and 5kHz). The resulting power spectral density is depicted in Figure 3.6. In Figure 3.7 the results obtained with an estimator with $N = 0$ and $N = 240$ are compared. As can be observed from these figures, good estimation results are obtained at these high speeds. Again, the estimation results obtained by the delayed estimator have a smaller error at the cost of time delay.

3.8 Summary and conclusions

For the application of micro-milling with an AMB spindle, minimum mean square error input estimators have been developed to estimate the cutting forces. As these

3. Unknown input estimation from closed-loop data
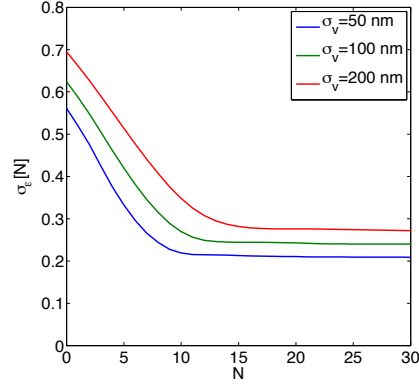
(a) Error σ_ϵ versus lag
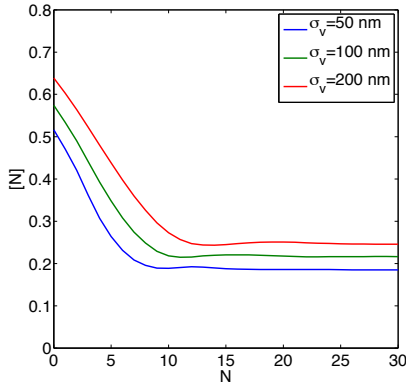
(b) Input error σ_{ϵ_u} versus lag
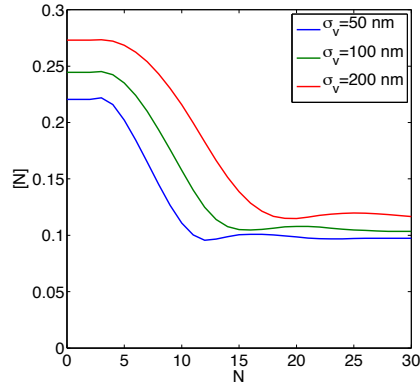
(c) Noise error σ_{ϵ_v} versus lag

Figure 3.5: Estimation error as function of the lag N for different levels of the position measurment disturbance v_2.

estimators use data already available in the closed-loop AMB system, no additional sensors are needed. Based on Wiener filter theory, conditions on the plant and controller are formulated, and estimators are developed to perform input estimation from closed-loop data. The estimators have an adjustable delay allowing to trade off the estimation error against the lag in the estimation results. Solutions are formulated for the case exact knowledge on the controller is available, and the case this knowledge is not available or cannot be used.
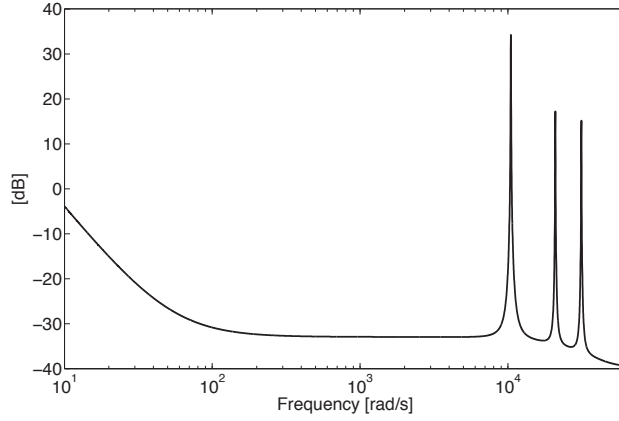
Figure 3.6: Spectral model for the cutting forces in simulation Section 3.7.2. High spectral density is modeled at the first three harmonics of the tooth pass frequency ($1.7 \cdot 2\pi$ krad/s).

For the case that no knowledge on the controller is available, an approach has been formulated that ensures equal performance for all possible controllers, making the estimator independent of the controller. This is achieved by a constraint optimal filter design problem, to which solutions are formulated. The conservatism to ensure equal performance for all possible controllers has been assessed, and will be small if the measurement noise on the control input of the plant is small. Moreover, if this measurement noise is zero, then it is shown that controller knowledge is equivalent to having full information on the controller output (provided the controller has no poles on the unit circle). The advantages of designing input estimators that use no controller knowledge are (1) a lower estimator order and (2) controller adaptation will not imply estimator adaptation.

Simulation results obtained with a realistic model of the EAAT spindle described in Chapter 2 were used to demonstrate the applicability of the given method. With the commonly used random walk model of the unknown input, an estimator designed without delay ($N = 0$) still produces delayed estimates. Designing an estimator with the same delay results in significantly better estimates. A filter with just a few time steps extra delay results, yields a reduction of error due to the measurement noise of up to 60%. Estimation results can be improved by using a priori information on the spectral content of the cutting forces to obtain a better model of the unknown input process.

3. UNKNOWN INPUT ESTIMATION FROM CLOSED-LOOP DATA
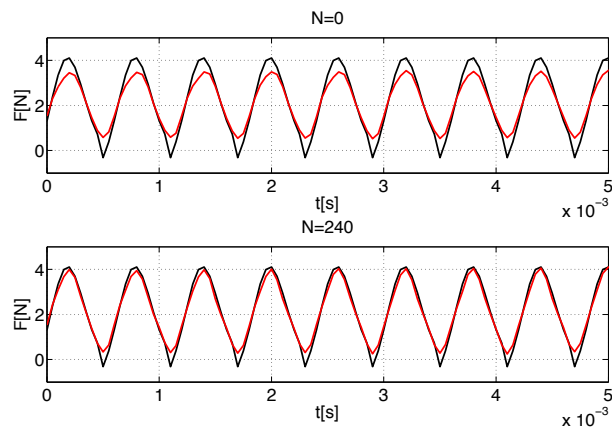
Figure 3.7: Cutting force estimation results (first element of $\hat{u}_2$: X-direction only) at 50,000 rpm obtained with the improved input force model. The delayed estimator again outperforms the estimator for $N = 0$ (black: true input u_2, red: estimated $\hat{u}_2$)

Chapter 4

Multivariable Frequency Domain Identification using IV-based Linear Regression

4.1   Introduction

In Chapter 3 an approach was given to estimate the micro-milling cutting forces from the signals of the AMBs. The success of this technique relies on the availability of an accurate model of the dynamics of the AMB spindle. In Chapter 2, it was discussed how such a model can be derived from first principles. However, the resulting models may still prove insufficiently accurate due to limited detailed knowledge on the system, or simplifications that need to be made in the modeling process. Under these conditions, identifying the system from measured data sequences can be a more efficacious alternative.

Chapter 4, 5 and 6 deal with the problem of identification of the dynamics of the AMB spindle. The content of these chapters is divided as follows. Identification of the AMB spindle can be separated into two identification problems. The first is the identification of the dynamics from the current input to the displacement of the rotor shaft at the bearings, which we refer to as the *bearing dynamics*. In Figure 2.11 in Chapter 2, the bearing dynamics are represented in the transfer function from $u_{1,0}$ to y. The second problem is the identification of the *tooltip dynamics*, which are the dynamics between the force on the tooltip and the displacement of the rotor shaft at the bearings. In Figure 2.11 in Chapter 2, the tooltip dynamics are represented in the transfer function from u_2 to y. In this thesis both identification problems are considered separately. This reason for this choice is as follows. It is straightforward to design experiments in which only the bearing dynamics are excited, while the tooltip dynamics are not. In such experiments, the AMB spindle will be run-

ning in free air, so that cutting forces are absent (i.e. $u_2 = 0$.). From the measured closed-loop data, the bearing dynamics can be identified. This is the main topic of Chapter 5. In contrast, designing experiments to excite the the tooltip dynamics are much more difficult, as will be further discussed in Chapter 6. As argued and shown in this chapter, solving the problem of identification of the tooltip dynamics can be simplified, if the bearing dynamics are known (viz. they have been identified already). Hence the choice to first consider the problem of identification of the bearing dynamics (Chapter 5), and then subsequently the problem of identification of the tooltip dynamics, with the assumption the bearing dynamics are known (Chapter 6).

In Chapter 5, system identification of the bearing dynamics will be approached by first making a non-parametric estimate of the multivariable frequency response function (FRF). Using this FRF estimate, a multivariable parametric model is estimated. Here, the main emphasis is on identification of a parametric model of the plant dynamics, leading to the choice of minimization of an output error OE criterion.

Minimizing a quadratic OE criterion with a model set in a fractional representation with a parametrized numerator, in general results in a nonconvex optimization problem. Gradient-based optimization can be employed to solve such problems, however the computational complexity has stimulated many authors to look for simpler alternatives. One approach is to replace the nonconvex optimization by a sequence of linear regression steps. A classical method to achieve this when estimating SISO models from frequency response function data is due to Sanathanan and Koerner [139]. This method forms the basis for several MIMO identification methods [23, 41, 54, 170], which differ in the model structure that is used to parametrize the multivariable system and the ability to incorporate frequency dependent weighting to improve the estimate. An important and well-known limitation of the algorithm proposed by Sanathanan and Koerner is that in general convergence of the iterations does not imply that the resulting parameter estimate minimizes the output error cost criterion, inevitably leading to a bias in the estimated model.

In literature, an alternative iterative linear regression procedure is available for time-domain data, based on an Instrumental Variable (IV) approach. This algorithm is known for time-domain identification of output error models as the Simplified Refined Instrumental Variable method (SRIV) [183, 184], and was recently formulated for frequency domain identification of SISO output error models [165]. This method has the property that upon convergence of the iterations a stationary point of the cost function is reached, also in the case the system is not in the model set.

This brings us to the main topic of this chapter, in which we generalize IV-based iterative linear regression to MIMO frequency domain identification of discrete-time and continuous-time models in matrix fraction description. Algorithms are derived for the case pre and post, or element-wise multivariable frequency weight-

ing of the output error is applied.

After introducing the identification setting in Section 4.2 and discussing the model sets in Section 4.3, these IV-based iterative linear regression algorithms are derived for identification of output error models in matrix fraction description (Section 4.4).

4.2 Identification setting

The central objective in this chapter is to find an LTI model P of a multivariable system with m inputs and p outputs using measured data. For this, the set of N noisy multivariable frequency response function observations $\{G(\omega_k) \in \mathbb{C}^{p \times m}\}_{k=1...N}$ is available. Here the model P is parametrized by either a left or right polynomial MFD, depending on the real valued parameter vector θ. Further details on the parametrization are discussed in the next section. To describe the discrepancy between the data and the model $P(\theta)$, we use the output error

$$E(\omega_k, \theta) = G(\omega_k) - P(\xi(\omega_k), \theta). \tag{4.1}$$

Here $\xi(\omega_k)$ is used to denote the frequency dependency of P, where $\xi(\omega_k) = i\omega_k$ or $\xi(\omega_k) = e^{i\omega_k}$ when P represents a continuous-time or discrete-time system respectively. With this, the aim is to solve the identification problem $\hat{\theta} = \arg\min_\theta V(\theta)$ with cost function

$$V(\theta) = \sum_{k=1}^{N} \|E(\omega_k, \theta)\|_F^2. \tag{4.2}$$

Application of frequency dependent weighting of the output error can be used to obtain estimates with smaller variance when the frequency response data have varying covariance, or to estimate e.g. control-relevant models. Cost functions with weighted errors can be obtained by substituting for $E(\omega_k, \theta)$ in (4.2) either the *input/output weighted OE*

$$E_{i/o}(\omega_k, \theta) = W_o(\omega_k)[G(\omega_k) - P(\xi(\omega_k), \theta)]W_i(\omega_k), \tag{4.3}$$

with $W_o(\omega_k) \in \mathbb{R}^{p \times p}$ and $W_i(\omega_k) \in \mathbb{R}^{m \times m}$, or the *Schur-weighted OE*

$$E_s(\omega_k, \theta) = W_s(\omega_k) .* [G(\omega_k) - P(\xi(\omega_k), \theta)], \tag{4.4}$$

with $W_s(\omega_k) \in \mathbb{R}^{p \times m}$, and where $.*$ is used to denote the Schur matrix product (i.e. element-wise multiplication).

4.3 Model structure

The MIMO model is represented in a left or right matrix fraction description:

$$P(\xi, \theta) = B(\xi, \theta)A^{-1}(\xi, \theta) \quad \text{(R-MFD)}$$
$$P(\xi, \theta) = A^{-1}(\xi, \theta)B(\xi, \theta) \quad \text{(L-MFD)}$$

with $B(\xi, \theta) = B_{n_b}\xi^{n_b} + B_{n_b-1}\xi^{n_b-1} + \cdots + B_0$ and $A(\xi, \theta) = \xi^{n_a} + A_{n_a-1}\xi^{n_a-1} + \cdots + A_0$, where $B_i \in \mathbb{R}^{p \times m}$, $i = 0 \ldots n_b$ and $A_i \in \mathbb{R}^{m \times m}$ (for models in R-MFD), or $A_i \in \mathbb{R}^{p \times p}$ (for models in L-MFD), for $i = 0 \ldots n_a - 1$. The parameter vector θ is constructed by accumulating all elements of the matrices A_i, $i = 0 \ldots n_a - 1$ and B_i, $i = 0 \ldots n_b$. As will become clear in the sequel, it is convenient to choose an ordering of these elements that depends on the choice for a left or right matrix fraction description of the model set. To avoid unnecessary notational complexity, it is assumed here that the order of the polynomials of all elements of A are n_a and those of B are all n_b. However, the approach presented in this chapter is equally suited for model sets where the matrix fractions have elements with varying polynomial orders. We will return to this in Section 4.4.6.

4.4 An IV-based iterative method to solve a multivariable OE identification problem

The identification problem that was posed in Section 4.2, generally results in a non-convex optimization problem. In this section we give an iterative linear regression algorithm to solve this optimization problem for both selected model sets, having the property that convergence implies (local) optimality. This algorithm is a MIMO extension of the frequency domain formulation of the SRIV method as given by Van den Hof and Douma in [165]. We first give the result for the unweighted OE cost function. To that end, first in Section 4.4.1, a criterion for optimality is derived. Using this, in Sections 4.4.2 and 4.4.3 the algorithms are derived for models in L-MFD and R-MFD respectively. In Section 4.4.4 we will extend these algorithms for the weighted OE cost functions.

4.4.1 A criterion for optimality

Similar as with the SRIV method, the starting notion is that for all $\hat{\theta}$ that locally minimize $V(\theta)$, it holds that $\frac{\partial}{\partial \theta}V(\theta)\big|_{\theta=\hat{\theta}} = 0$. From the definition of $V(\theta)$ in (4.2), it follows that

$$
\begin{aligned}
\frac{\partial}{\partial \theta}V(\theta) &= \frac{\partial}{\partial \theta}\sum_{k=1}^{N}\mathrm{vec}\left[E(\omega_k, \theta)\right]^H \mathrm{vec}\left[E(\omega_k, \theta)\right] \\
&= \sum_{k=1}^{N} -2Re\left\{\mathrm{vec}\left[E(\omega_k, \theta)\right]^H M_k(\theta)\right\}
\end{aligned}
$$

where $(\cdot)^H$ denotes the complex conjugate transpose, $\mathrm{vec}(\cdot)$ the vectorization operator, and $M_k(\theta) = \frac{\partial}{\partial \theta}\mathrm{vec}[P(\xi(\omega_k), \theta)]$. Hence, for all $\hat{\theta}$ for which $\frac{\partial}{\partial \theta}V(\hat{\theta}) = 0$,

An IV-based iterative method to solve a multivariable OE identification problem

the following equality holds:

$$\sum_{k=1}^{N} Re \left\{ M_k^H(\hat{\theta}) \text{vec}[E(\omega_k, \hat{\theta})] \right\} = 0. \tag{4.5}$$

4.4.2 Iterative procedure for models in L-MFD

In this section, we will use equation (4.5) to arrive at an iterative linear regression algorithm to estimate $\hat{\theta}$ for models in L-MFD. For that, we will rewrite equation (4.5) in a regression format. Let us therefore introduce the notation

$$\Theta = \begin{bmatrix} A_{n_a-1} & \dots & A_0 & B_{n_b} & \dots & B_0 \end{bmatrix}, \qquad \theta = \text{vec}(\Theta). \tag{4.6}$$

We give the following two propositions:

Proposition 4.4.1 *With $E(\xi(\omega_k), \theta)$ as defined in equation (4.1), where the model is represented in L-MFD, and with θ as defined in (4.6), the following identity holds*

$$\text{vec}\left[E(\omega_k, \theta)\right] = Y_k(\theta) - X_k(\theta)\theta \tag{4.7}$$

with

$$Y_k(\theta) = \left[I \otimes A^{-1}(\xi(\omega_k), \theta)\right] \text{vec}[\xi(\omega_k)^{n_a} G(\omega_k)]$$
$$X_k(\theta) = \left[\Omega^T(\omega_k) \otimes A^{-1}(\xi(\omega_k), \theta)\right]$$

where $\otimes$ is the Kronecker product, and

$$\Omega(\omega_k) = \begin{bmatrix} -\xi(\omega_k)^{n_a-1} G(\omega_k) \\ \vdots \\ -\xi(\omega_k)^0 G(\omega_k) \\ \xi(\omega_k)^{n_b} I_{m\times m} \\ \vdots \\ \xi(\omega_k)^0 I_{m\times m} \end{bmatrix}.$$

Proof For the given model parametrization, and using the given definition of $\Omega(\omega_k)$, observe that we can express $E(\xi(\omega_k), \theta)$ as

$$E(\xi(\omega_k), \theta) = G(\omega_k) - A^{-1}(\xi(\omega_k), \theta) B(\xi(\omega_k), \theta)$$
$$= A^{-1}(\xi(\omega_k), \theta) \left[A(\xi(\omega_k), \theta) G(\omega_k) - B(\xi(\omega_k), \theta)\right]$$
$$= A^{-1}(\xi(\omega_k), \theta) \left[\xi(\omega_k)^{n_a} G(\omega_k) - \Theta\Omega(\omega_k)\right].$$

To proceed, we need the following two identities:

$$\text{vec}(AB) = (I \otimes A)\text{vec}(B) = (B^T \otimes I)\text{vec}(A) \tag{4.8a}$$
$$(A \otimes B)(C \otimes D) = AB \otimes DB. \tag{4.8b}$$

Using (4.8a) we can write

$$\text{vec}\left[E(\omega_k,\theta)\right] = \left[I \otimes A^{-1}(\xi(\omega_k),\theta)\right]\text{vec}\left[\xi(\omega_k)^{n_a}G(\omega_k) - \Theta\Omega(\omega_k)\right]$$

which by applying (4.8a) and subsequently (4.8b), we can rewrite to

$$\begin{aligned}
\text{vec}\left[E(\omega_k,\theta)\right] &= \left[I \otimes A^{-1}(\xi(\omega_k),\theta)\right]\cdot \\
&\quad \cdot \left(\text{vec}[\xi(\omega_k)^{n_a}G(\omega_k)] - (\Omega^T(\omega_k) \otimes I)\text{vec}(\Theta)\right) \\
&= \left[I \otimes A^{-1}(\xi(\omega_k),\theta)\right]\text{vec}[\xi(\omega_k)^{n_a}G(\omega_k)] + \\
&\quad\quad\quad - \left[\Omega^T(\omega_k) \otimes A^{-1}(\xi(\omega_k),\theta)\right]\theta
\end{aligned}$$

which is the claimed result. ∎

Proposition 4.4.2 *For models parametrized in L-MFD,*

$$M_k(\theta) = \Phi_k(\theta)^T \otimes A^{-1}(\xi(\omega_k),\theta) \tag{4.9}$$

with

$$\Phi_k(\theta) = \begin{bmatrix} -\xi(\omega_k)^{n_a-1}P(\xi(\omega_k),\theta) \\ \vdots \\ -\xi(\omega_k)^0 P(\xi(\omega_k),\theta) \\ \xi(\omega_k)^{n_b}I_{m\times m} \\ \vdots \\ \xi(\omega_k)^0 I_{m\times m} \end{bmatrix}.$$

Proof See appendix B.2.1. ∎

With the results in equations (4.7) and (4.9), we can recast (4.5) into

$$\sum_{k=1}^N Re\left\{M_k^H(\hat\theta)(Y_k(\hat\theta) - X_k(\hat\theta)\hat\theta)\right\} = 0 \tag{4.10}$$

or equivalently

$$\sum_{k=1}^N \begin{bmatrix} Re\{M_k^T(\hat\theta)\} & Im\{M_k^T(\hat\theta)\} \end{bmatrix}\cdot$$
$$\cdot\left(\begin{bmatrix} Re\{Y_k(\hat\theta)\} \\ Im\{Y_k(\hat\theta)\} \end{bmatrix} - \begin{bmatrix} Re\{X_k(\hat\theta)\} \\ Im\{X_k(\hat\theta)\} \end{bmatrix}\hat\theta\right) = 0.$$

With the notation

$$\mathbf{M}^T(\theta) := \begin{bmatrix} Re\{M_1^T(\theta)\} & Im\{M_1^T(\theta)\} & \dots \\ & \dots & Re\{M_N^T(\theta)\} & Im\{M_N^T(\theta)\} \end{bmatrix}$$

An IV-based iterative method to solve a multivariable OE identification problem

$$\mathbf{X}(\theta) := \begin{bmatrix} Re\{X_1(\theta)\} \\ Im\{X_1(\theta)\} \\ \vdots \\ Re\{X_N(\theta)\} \\ Im\{X_N(\theta)\} \end{bmatrix}, \qquad \mathbf{Y}(\theta) := \begin{bmatrix} Re\{Y_1(\theta)\} \\ Im\{Y_1(\theta)\} \\ \vdots \\ Re\{Y_N(\theta)\} \\ Im\{Y_N(\theta)\} \end{bmatrix}$$

it follows that the solution of (4.5) is characterized by

$$\mathbf{M}^T(\hat{\theta})(\mathbf{Y}(\hat{\theta}) - \mathbf{X}(\hat{\theta})\hat{\theta}) = 0. \tag{4.11}$$

From this, a natural iterative identification algorithm follows:

$$\text{sol}_{\hat{\theta}_{j+1}} \left\{ \mathbf{M}^T(\hat{\theta}_j)(\mathbf{Y}(\hat{\theta}_j) - \mathbf{X}(\hat{\theta}_j)\hat{\theta}_{j+1}) = 0 \right\} \tag{4.12}$$

with solution

$$\hat{\theta}_{j+1} = \left[\mathbf{M}^T(\hat{\theta}_j)\mathbf{X}(\hat{\theta}_j) \right]^{-1} \mathbf{M}^T(\hat{\theta}_j)\mathbf{Y}(\hat{\theta}_j). \tag{4.13}$$

When this algorithm converges, necessarily $V'(\hat{\theta}) = 0$, ensuring that $\hat{\theta}$ is a stationary point of the cost function. Observe that (4.13) has the structure of an IV estimator. Also, note that replacing $\mathbf{M}$ by $\mathbf{X}$ would give the Sanathanan-Koerner iteration for the given problem.

4.4.3 Iterative procedure for models in R-MFD

Analogous to the analysis in the previous section, we will now use equation (4.5) to derive an iterative linear regression algorithm for systems in R-MFD . To that end, we introduce a different notation for Θ, i.e.

$$\Theta^T = \begin{bmatrix} A_{n_a-1}^T & \dots & A_0^T & B_{n_b}^T & \dots & B_0^T \end{bmatrix}$$
$$\theta = \text{vec}(\Theta). \tag{4.14}$$

With this, we give the following two propositions.

Proposition 4.4.3 *With $E(\xi(\omega_k), \theta)$ as defined in (4.1), where the model is represented in R-MFD, and with θ as in defined (4.14), the following identity holds*

$$\text{vec}\left[E(\omega_k, \theta)\right] = Y_k(\theta) - X_k(\theta)\theta \tag{4.15}$$

with

$$Y_k(\theta) = \left[A^{-T}(\xi(\omega_k), \theta) \otimes I \right] \text{vec}\left(\xi(\omega_k)^{n_a} G(\omega_k) \right)$$
$$X_k(\theta) = \left[A^{-T}(\xi(\omega_k), \theta) \otimes \Omega(\omega_k) \right]$$

where

$$\Omega(\omega_k) = \begin{bmatrix} -\xi(\omega_k)^{n_a-1}G(\omega_k) & \dots & -\xi(\omega_k)^0 G(\omega_k) \\ & & \xi(\omega_k)^{n_b}I_{p\times p} & \dots & \xi(\omega_k)^0 I_{p\times p} \end{bmatrix}.$$

4. Multivariable Frequency Domain Identification using IV-based Linear Regression

Proof For the given model parameterization, and using the above definition of $\Omega(\omega_k)$, observe that

$$E(\xi(\omega_k),\theta) = G(j\omega_k) - B(\xi(\omega_k),\theta)A^{-1}(\xi(\omega_k),\theta)$$
$$= \left[G(j\omega_k)A(\xi(\omega_k),\theta) - B(\xi(\omega_k),\theta) \right] A^{-1}(\xi(\omega_k),\theta)$$
$$= \left[\xi(\omega_k)^{n_a}G(j\omega_k) - \Omega(\omega_k)\Theta \right] A^{-1}(\xi(\omega_k),\theta).$$

Applying (4.8a) and (4.8b) yields

$$\text{vec}\left[E(\xi(\omega_k),\theta)\right] = \left[A^{-T}(\xi(\omega_k),\theta) \otimes I \right] \text{vec}\left(\xi(\omega_k)^{n_a}G(j\omega_k) \right) +$$
$$- \left[A^{-T}(\xi(\omega_k),\theta) \otimes \Omega(\omega_k) \right]\theta,$$

which proves the claim. ∎

Proposition 4.4.4 *For models parametrized in R-MFD,*

$$M_k(\theta) = A^{-T}(\xi(\omega_k),\theta) \otimes \Phi_k(\theta) \tag{4.16}$$

with

$$\Phi_k(\theta) = \left[\begin{array}{ccc} -\xi(\omega_k)^{n_a-1}P(\xi(\omega_k),\theta) & \dots & -\xi(\omega_k)^0 P(\xi(\omega_k),\theta) \\ & \xi(\omega_k)^{n_b}I_{p\times p} & \dots & \xi(\omega_k)^0 I_{p\times p} \end{array} \right].$$

Proof See appendix B.2.2. ∎

With the results in (4.15) and (4.16), we can rewrite (4.5) also for models in R-MFD to (4.10), albeit with different definitions of the matrices Y_k, X_k and M_k. Hence, by applying these definitions, a similar iterative identification algorithm can be followed as the one derived for systems in L-MFD in the previous section.

4.4.4 Minimization of weighted OE cost criteria

The algorithms in the previous sections were derived for the cost function based on the unweighted error. Here we will show how these results can be generalized for the case input/output weighting or Schur weighting is applied.

Input-output weighting

Observe that with the input/output weighted error $E_{i/o}(\omega_k,\theta)$ as defined in (4.3), setting the first derivative of the cost function to zero yields the equality

$$\sum_{k=1}^{N} Re\left\{ M_{i/o,k}^{H}(\hat{\theta})\text{vec}\left[W_o(\omega_k)E(\hat{\omega}_k,\theta)W_i(\omega) \right] \right\} = 0 \tag{4.17}$$

An IV-based iterative method to solve a multivariable OE identification problem

where $M_{i/o,k}(\theta) = \frac{\partial}{\partial\theta}\text{vec}[W_o(\omega_k)P(\xi(\omega_k),\theta)W_i(\omega_k)]$. Using the identity $\text{vec}(ABC) = (C^T \otimes A)\text{vec}(B)$, we derive that

$$\text{vec}[W_o(\omega_k)E(\omega_k,\theta)W_i(\omega_k)] = [W_i^T(\omega_k) \otimes W_o(\omega_k)]\text{vec}[E(\omega_k,\theta)]$$

and

$$M_{i/o,k}(\theta) = \frac{\partial}{\partial\theta}[W_i^T(\omega_k) \otimes W_o(\omega_k)]\text{vec}[P(\xi(\omega_k),\theta)]$$
$$= [W_i^T(\omega_k) \otimes W_o(\omega_k)]M_k(\theta).$$

By substituting these identities in (4.17), and using the expressions derived for $\text{vec}[E(\omega_k,\theta)]$ and $M_k(\omega_k,\theta)$, iterative algorithms to minimize the input/output weighted cost can be obtained in the same fashion as derived in the previous sections.

Schur weighting

In a similar fashion as for input/output weighting, note that with the Schur-weighted error $E_s(\omega_k,\theta)$ as defined in (4.4), $\frac{\partial}{\partial\theta}V(\theta) = 0$ implies

$$\sum_{k=1}^{N} Re\left\{M_{s,k}^H(\hat{\theta})\text{vec}\left[W_s(\omega_k).*E(\omega_k,\hat{\theta})\right]\right\} = 0 \qquad (4.18)$$

where $M_{s,k}(\theta) = \frac{\partial}{\partial\theta}\text{vec}[W_s(\omega_k).*P(\xi(\omega_k),\theta)]$. We derive that

$$\text{vec}[W_s(\omega_k).*E(\omega_k,\theta)] = \text{vec}[W_s(\omega_k)].*\text{vec}[E(\omega_k,\theta)]$$

and

$$M_{s,k}(\theta) = \frac{\partial}{\partial\theta}\text{vec}[W_s(\omega_k)].*\text{vec}[P(\xi(\omega_k),\theta)]$$
$$= \text{vec}[W_s(\omega_k)].*M_k(\theta).$$

Again, substitution of these identities in (4.17), in conjuction with the derived expressions for $\text{vec}[E(\omega_k,\theta)]$ and $M_k(\omega_k,\theta)$, allows to derive iterative algorithms that minimize the Schur weighted cost upon convergence.

4.4.5 Estimation of common denominator models

The algorithm that is described in this chapter ensures that converging iterations imply that an optimal estimate of the parameters in a matrix fraction representation is obtained. Here we will demonstrate that this property can also be obtained for model representations with a common denominator.

For this, let the model set be defined by $P(\xi, \theta) = B(\xi, \theta)A^{-1}(\xi, \theta)$, where $B(\xi, \theta)$ is as defined before, and $A(\xi, \theta) = I \cdot a(\xi, \theta)$ with $a(\xi, \theta)$ a scalar polynomial. Estimation of the parameters in this represention can be reformulated to estimation of the parameters in a representation that matches the fully parametrized matrix fraction representation of Section 4.3[1]. Indeed, there exists $G_v(\omega_k)$, $B_v(\xi, \theta)$ and $A_v(\xi, \theta)$, with $B_v(\xi, \theta)$ and $A_v(\xi, \theta)$ fully parametrized, such that

$$\|E(\omega_k, \theta)\|_F^2 = \|E_v(\omega_k, \theta)\|_F^2 \tag{4.19}$$

where $E_v(\omega_k, \theta) = G_v(\omega_k) - B_v(\xi(\omega_k), \theta)A_v^{-1}(\xi(\omega_k), \theta)$. To see this, note that for models with a common denominator, we can write

$$\|E(\omega_k, \theta)\|_F^2 = \left| \mathrm{vec}[G(\omega_k)] - \mathrm{vec}[B(\xi(\omega_k), \theta)] \cdot a^{-1}(\xi(\omega_k), \theta) \right|^2.$$

With this it follows that if $G_v(\omega_k) = \mathrm{vec}[G(\omega_k)]$, $B_v(\xi, \theta) = B_{v,n_b}\xi^{n_b} + \cdots + B_{v,0}$, where $B_{v,i} \in \mathbb{R}^{p \cdot m \times 1}$, $i = 0 \ldots n_b$ and $B_{v,i} = \mathrm{vec}(B_i)$, and $A_v(\xi, \theta) = a(\xi, \theta)$, identity (4.19) will hold. Once having obtained estimates $\hat{B}_{v,i}$, we can directly construct estimates $\hat{B}_i$ using the identity $B_{v,i} = \mathrm{vec}(B_i)$.

4.4.6 Extension for model sets with non-full parametrization

Until now, it was assumed that polynomial matrices $A(\xi, \theta)$ and $B(\xi, \theta)$ are fully parametrized. However, it is straightforward to deal with model descriptions for non-full parametrizations. Observe that following the approach in the previous sections for a non-full parameterization would result in a parameter vector θ with one or more zero elements. Removal of these elements from θ, as well as deletion of the corresponding columns from the matrices M_k and X_k, yields the desired result.

4.5 Conclusions

Iterative linear regression algorithms are given for estimation of output error models in left or right matrix fraction description from frequency response data. These algorithms are extensions of the SISO IV-based linear regression algorithm, which has the property that convergence implies a stationary point of the cost function is reached. This property, in combination with the freedom in the definition of the model set and the possibility to incorporate pre, post or element-wise multivariable frequency weighting, make this an attractive approach for MIMO frequency domain identification of output error models. It is not claimed that the method discussed here can outperform gradient-based optimization methods. However, it appears to be a favorable alternative for the classically applied SK-iterations.

---

[1] A polynomial matrix $A(\xi, \theta) = \xi^{n_a} + A_{n_a-1}\xi^{n_a-1} + \cdots + A_0$ is said to be fully parametrized none of the coefficient matrices A_i, $i = 0 \ldots n_a - 1$ contains zeros., and all elements are independently parametrized

Chapter 5

System identification of the AMB spindle, part I: Bearing dynamics

5.1   Introduction

As indicated in the introduction of the previous chapter, identification of the AMB spindle can be split into two identification problems, i.e. identification of the *bearing dynamics* and the identification of the *tooltip dynamics*. This chapter deals with the first.

For a number of reasons, identification of the bearing dynamics is a complex problem. First, the dynamics are intrinsically unstable, which implies that identification needs to be performed in a stabilizing closed-loop. The second reason pertains to the multivariable character of the dynamics, where due to gyroscopic effects the coupling between the various channels increases with rising rotational frequency. A third reason is the high order of the system. As discussed in Chapter 2, an AMB milling spindle typically exhibits high-frequent and very ill-damped resonances, which originate from the flexural modes of the spindle rotor. The last reason is that the electromagnetic actuators of AMBs are nonlinear by nature, implying that only small excitation signals can be used if nonlinear distortions are to be avoided. Hence the signal-to-noise ratio (SNR) of the measurement data is limited, requiring careful experiment design.

System identification of AMB spindles has been studied for many years, mostly for the purpose of control design. The majority of this work is approached using frequency domain techniques. Non-parametric estimates of the frequency response function (FRF) are made, to which grey-box or black-box parametric models are fit. Grey-box models are estimated in the approaches of [53], [5], [107], and [56]. The approach by Gähler et al. [53] is based on a modal decomposition of an analytical model and involves identification of the modal frequencies. This approach

was extended by Lösch [107] to include estimation of the gyroscopic coupling at high rotational speeds. The approach by Ghersin et al. [56] uses a model class described by state space systems with a prescribed structure, which follows from analytical modeling of the setup at stand-still. In the approach of Aeschlimann [5], the predicted frequency response from an analytic model is fit to the measured frequency response by variation of the uncertain parameters in the model. Black-box approaches have also been reported, among them the subspace approach by Mohd-mokhtar et al. [115], and the approach by Ahn et al. [6] in which low-order control-relevant models in matrix fractional description are fit to the estimated FRF.

Common in many of these approaches is that the above described complexity of the identification problem is reduced in various ways. A frequently made simplification is to perform grey-box identification, which imposes a model structure on the system with reduced complexity. Another simplification involves ignoring the gyroscopic coupling, which reduces the dimensions of the problem. Although the resulting models might be appropriate for (robust) control design objectives, the obtained accuracy is insufficient when pursuing more demanding applications like process monitoring and control of the micro-milling process.

Furthermore, all of these approaches have in common that linearity of the AMB spindle system is assumed, while it is in fact known to be intrinsically nonlinear. Still, in many applications, a linear model is desired and may also be sufficient to meet the performance specifications. However, presence of nonlinear distortions requires a careful experiment design to make a good trade-off between keeping the excitation small to minimize nonlinear distortion and maintaining appropriate SNR to obtain sufficiently low variance error in the FRF estimate.

The objective of this chapter is to identify a linear parametric model of the bearing dynamics that is suitable for application to cutting force estimation. This application has a number implications for the identification problem, making it more involved than the identification of the bearing dynamics for the application of control design. These implications can be summarized as follows:

- The resulting model must be accurate in the frequency range in which the cutting forces are estimated, which is in general larger than the frequency range of interest for control design of the bearings. The larger frequency range of interest implies that not only rigid body dynamics, but also flexible body dynamics come into play. For example, for the EAAT AMB spindle this range of interest is 0 to 5 kHz (see Chapter 3). This range includes the first and second bending mode frequency of the rotor. A further complication lies in the fact that at the high rotational speeds used in micro-milling, the effects of gyroscopy become relevant (see Chapter 2). This means that the dynamics of motion in the radial direction become increasingly more coupled, while in addition the bending moments are split. In summary, the aforementioned simplifications cannot be justified, and instead high order models are needed in order to describe the bearing dynamics. At present, literature is lacking an

AMB identification procedure able to deal with this complexity.

- The application for estimation sets tight constraints on the accuracy of the identified model. The obvious reason is that any errors in the model dynamics of the AMB spindle will result in errors in the cutting force estimate. Both measurement noise, as well as nonlinear distortions give rise to errors in the estimated model, where evidently the choice of the excitation signal plays an important role. Large excitation yields a good signal to noise ratio, but also large nonlinear distortions. On the other hand, a small excitation might avoid distortions from nonlinearities, but still yield poor results due to an inferior signal to noise ratio. The problem of designing identification experiments allowing to make a good trade-off between the errors from these two sources, has not received much attention in the literature on AMB spindle identification.

In order to deal with these complications and to meet the aforementioned objective to identify a model of the bearing dynamics suitable for application to cutting force estimation, an approach has been selected consisting of two stages:

1. Stage one involves estimation of the multivariable Frequency Response Function (FRF) of the bearing dynamics. In contrast to many of the contributions in the literature on AMB identification, careful attention is given to minimizing both bias and variance of the FRF estimate. To that end, an experiment design is given that incorporates recent results regarding estimation of the FRF of MIMO systems from closed-loop data.

   Furthermore, a procedure is given to detect the level of nonlinear distortion in the FRF estimate of the AMB spindle. This approach is based on the variance analysis method for detection of nonlinearities in FRF estimates as available in literature, and is extended to detection of nonlinear distortion in MIMO FRF estimates obtained using an instrumental variables approach. Being able to detect the level of nonlinear distortion, the problem is addressed how the experimental conditions can be modified so, as to reduce the effect of the nonlinear distorions in the FRF estimate, particularly when dealing with closed-loop experiments.

   The last aspect of stage one involves non-parametric identification of the noise disturbances in the closed-loop, which is addressed as well by a variance analysis using the periodic nature of the excitation.

2. In stage two, the objective is to identify a model for the MIMO plant in matrix fraction description (MFD), using the estimated FRF from stage one. Here, the main emphasis is on identification of a parametric model of the bearing dynamics, and not so much on obtaining a parametric model of the measurement noise. This leads to the choice of minimization of an output error criterion. In order to reduce the variance error of the estimated parameters,

a Schur-weighted OE cost criterion is formulated, where an estimate of the variance of the FRF is used for the weighting. The resulting identification problem fits the framework of Chapter 4, and the IV-based iterative linear regression approach is used to solve this.

A complication of the large dynamic range is that the IV-based iterative linear regression algorithm suffers from poor numerical conditioning (similar to the Sanathanan and Koerner algorithm). The multi-band technique of Bayard [22] is adopted to deal with this. The key idea here is to divide the frequency range of interest into smaller subbands and estimate submodels that describe the system on each band. When combined, these submodels produce a model for the entire system.

The choice for a frequency-domain approach for the problem of identifying the bearing dynamics is justified as follows:

- **Data/noise reduction**: in stage one, large amounts of data in the time-domain can be aggregated to compute a reduced-size data set in the frequency domain with improved SNR.

- **Closed-loop effects**: The closed-loop issues are dealt with in stage one, where the FRF of the bearing dynamics are estimated from the closed-loop data. The parametric modeling in stage two is then simply an open-loop problem.

- **Instability of the bearing dynamics**: The parametric modeling step in stage two is performed in the frequency domain. In this domain instability of the bearing dynamics imposes no limitations on the model set (contrary to time-domain prediction error identification of unstable systems).

This chapter is structured as follows. After discussing the configuration in Section 5.2, we will address the approach to estimation of the FRF of the bearing dynamics in Section 5.3. Subsequently in Section 5.4 we discuss how nonlinear distortion in the estimate of the FRF can be detected, and reduced. From the experimental data, non-parametric estimates of the noise disturbances in the system can be obtained as well. An approach to do this is given in Section 5.5. Estimation of a parametric model using the estimated FRF (and its covariance) is discussed in Section 5.6.

5.2   Configuration and identification problem

In relation to Figure 2.11 of Chapter 2, identification of the bearing dynamics involves identification of P_1 alone, while $u_2 = 0$ (the cutting forces are zero during identification of the bearing dynamics). For reasons of notational efficiency, in this chapter we will therefore ignore existence of P_2 and simply write P instead of P_1 and u instead of u_1. With this, we obtain the modified block diagram of Figure 5.1.
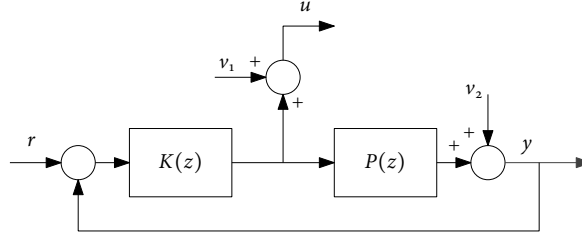
Figure 5.1: Block diagram of the AMB spindle for the purpose of identifying the bearing dynamics.

The reference of the controller r is available to excite the system, and measurements of the input and output of P are taken. The general objective of the identification procedure is to use the obtained data to find a parametric model of the unstable plant $P(z)$. In this procedure we will assume that the noise disturbances on the measurements of the currents v_1 are negligible (we will justify this assumption for the EAAT spindle in Section 5.5). The rotational speed of the spindle is assumed to be constant. For the identification, the runout disturbances in the system are not treated separately and are included in v_1 and v_2.

5.3 Multivariable FRF Estimation of the AMB spindle

In this section, estimation of the multivariable FRF of the AMB spindle system is performed. We will focus on obtaining a non-parametric FRF estimate from the closed-loop data with low bias and variance for given excitation levels. First, the topic of experiment design is discussed in Section 5.3.1, after which we address the issue of non-parametric FRF estimation of multivariable systems in closed-loop (Section 5.3.2). Estimation of the covariance of the FRF estimate is dealt with in Section 5.3.3.

5.3.1 Experiment design

With the purpose of estimating the FRF of the bearing dynamics, the following experiment design is chosen. Excitation signals are applied to the reference input r. To avoid leakage effects when calculating the Discrete Fourier Transform (DFT), r is assumed to be N_K-periodic: $r(t + N_K T_s) = r(t)$ with T_s the sample time. When the system has reached its steady state, K periods of $u(t)$ and $y(t)$ are collected, resulting in $N = KN_K$ samples per experiment. The DFTs of the reference r, the

input u and the output y, averaged over the K periods, are given by

$$R(\omega_k) = \frac{1}{K\sqrt{N}} \sum_{t=1}^{N} r(tT_s) e^{-j\omega_k t T_s} \tag{5.1a}$$

$$U(\omega_k) = \frac{1}{K\sqrt{N}} \sum_{t=1}^{N} u(tT_s) e^{-j\omega_k t T_s} \tag{5.1b}$$

$$Y(\omega_k) = \frac{1}{K\sqrt{N}} \sum_{t=1}^{N} y(tT_s) e^{-j\omega_k t T_s} \tag{5.1c}$$

where only the frequencies $\omega_k = k\frac{2\pi}{N_K T_s}$, $k = 1 \ldots N_K$ are considered. In order to estimate $P(\omega_k)$ from data, at least m experiments with independent excitation signals are needed. Let M represent the number of experiments and let $\mathbf{R}(\omega_k)$ be the matrix that is formed by collecting the DFTs of the reference r (each column corresponding to one experiment):

$$\mathbf{R}(\omega_k) = \begin{bmatrix} R^1(\omega_k) & R^2(\omega_k) & \ldots & R^M(\omega_k) \end{bmatrix} \in \mathbb{C}^{m \times M} \tag{5.2}$$

where the superscript is used to indicate the experiment number. Matrices $\mathbf{U}(\omega_k)$ and $\mathbf{Y}(\omega_k)$ are formed accordingly from the DFTs of u and y respectively. To facilitate the analysis, it is assumed that the total set of experiments M is an integer multiple of the number of inputs m, i.e. $M = M_B \cdot m$. In what follows, we will consider one *experiment block* to consist of m experiments, and so M_B is the number of experiment blocks. Following from this, the matrix $\mathbf{R}(\omega_k)$ is partitioned into M_B blocks of size $m \times m$ as

$$\mathbf{R}(\omega_k) = \begin{bmatrix} \mathbf{R}^{[1]}(\omega_k) & \ldots & \mathbf{R}^{[M_B]}(\omega_k) \end{bmatrix} \tag{5.3}$$

where we use the superscripted index in square brackets to denote the experiment block number. Partionings for matrices $\mathbf{U}(\omega_k)$ and $\mathbf{Y}(\omega_k)$ are defined conformably.

For the excitation, orthogonal random phase multi-sine signals are selected. For given excitation power, these excitation signals are known to yield the least variance of the FRF estimate compared to other common excitation signals [43]. The orthogonal random phase multisine signal is defined as follows. The blocks $\mathbf{R}^{[i]}$, $i = 1 \ldots M_B$ satisfy

$$\mathbf{R}^{[i]}(\omega_k) = \mathbf{R}^{[i]}_{\text{diag}}(\omega_k)\mathbf{W} \tag{5.4}$$

with $\mathbf{R}^{[i]}_{\text{diag}}(\omega_k) = \text{diag}\left\{ R^{[i]}_l(\omega_k) \right\}_{l=1}^{m}$ a diagonal matrix, where each entry on the diagonal is the DFT of a random phase multisine signal, i.e.

$$r_l(t) = \sum_{k=1}^{N_K/2} A_{k,l} \cos(\omega_k t + \phi_{k,l}), \quad l = 1 \ldots m \tag{5.5}$$

with amplitudes $A_{k,l}$, frequencies $\omega_k \in \left\{ \frac{2\pi l}{N_K T_s}, l = 1, \ldots, \frac{N_K}{2} - 1 \right\}$ (for N_K even), and random phases $\phi_{k,l}$ uniformly distributed on the interval $[0, 2\pi)$. The matrix $\mathbf{W}$ is an arbitrary, deterministic unitary (or orthogonal) matrix, satisfying $\mathbf{W}^H \mathbf{W} = \mathbf{W}\mathbf{W}^H = m I_m$. For arbitrary input dimension, a possible choice for $\mathbf{W}$ is given by the DFT matrix:

$$\mathbf{W}_{pq} = e^{\frac{j 2\pi}{m}(p-1)(q-1)} \tag{5.6}$$

However, since $m = 4$ in the case of the AMB spindle, a much simpler choice is the 4×4 Hadamard matrix, given by

$$\mathbf{W} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}. \tag{5.7}$$

With this, the resulting procedure for one experiment block becomes as follows:

- Experiment 1: for all four channels of the reference a multisine signal is generated, each with a different realization of the random phases, resulting in the signal r. The system is excited with this signal and $K N_K$ samples of the steady state response of u and y are measured.

- Experiment 2: experiment 1 is repeated, but now with channel 2 and 4 multiplied with -1.

- Experiment 3: experiment 1 is repeated, but now with channel 3 and 4 multiplied with -1.

- Experiment 4: experiment 1 is repeated, but now with channel 2 and 3 multiplied with -1.

5.3.2 Estimation of the multivariable FRF

Non-parametric estimation of the FRF of dynamical systems has been studied extensively in literature [175, 133, 63, 105, 128, 127]. Originally, most of these methods were developed for SISO systems, but extensions are available for MIMO systems [64, 169, 62].

Here a closed-loop identification problem is considered. As reviewed by Van den Hof in [164], in time-domain identification two main approaches can be followed to deal with the closed-loop effects: (1) direct identification, and (2) indirect identification. In direct identification, standard (open-loop) identification techniques are applied, without taking into account the presence of a feedback controller. The main difference of indirect methods compared to direct methods is that a measurable external excitation signal is available. Indirect methods vary in

the way that the plant model is parametrized and the way they deal with (removing) the noise contribution at the input of the plant.

For non-parametric frequency-domain closed-loop identification of multivariable systems that we consider, several estimators have been proposed in literature [176]. We will first introduce three main approaches and then comment on their bias and variance properties.

- **The H_1 estimator**. This estimator uses the direct approach [105]: it is based on the data of u and y only, and presence of the feedback is ignored.

$$\hat{P}^{H_1}(\omega_k) = \mathbf{Y}(\omega_k)\mathbf{U}^H(\omega_k)(\mathbf{U}(\omega_k)\mathbf{U}^H(\omega_k))^{-1} \qquad (5.8)$$

- **The arithmetic mean estimator (ARI)**. Like the H_1 estimator, the ARI estimator is based on a direct approach. The main idea behind this estimator is that FRF estimates obtained with each block of excitations $\mathbf{R}^{[i]}(\omega_k)$ are averaged:

$$\hat{P}^{\text{ARI}}(\omega_k) = \frac{1}{M_B}\sum_{i=1}^{M_B}\mathbf{Y}^{[i]}(\omega_k)(\mathbf{U}^{[i]}(\omega_k))^{-1}. \qquad (5.9)$$

- **The joint input-output (JIO) estimator**. When the reference signal is also recorded, a joint input-output approach [105] can be followed. In this indirect method, the input and output of the plant are modeled using the external excitation and noise signal. Derived from this, the JIO estimator is given by:

$$\hat{P}^{\text{JIO}}(\omega_k) = \mathbf{Y}(\omega_k)\mathbf{R}^H(\omega_k)(\mathbf{U}(\omega_k)\mathbf{R}^H(\omega_k))^{-1}. \qquad (5.10)$$

The JIO estimator can also be thought of the combination of two H_1 estimators

$$\hat{T}_{ur}^{H_1}(\omega_k) = \mathbf{U}(\omega_k)\mathbf{R}^H(\omega_k)(\mathbf{R}(\omega_k)\mathbf{R}^H(\omega_k))^{-1} \qquad (5.11)$$

$$\hat{T}_{yr}^{H_1}(\omega_k) = \mathbf{Y}(\omega_k)\mathbf{R}^H(\omega_k)(\mathbf{R}(\omega_k)\mathbf{R}^H(\omega_k))^{-1} \qquad (5.12)$$

where $\hat{T}_{ur}^{H_1}$ is an estimate of the transfer function of the reference r to the plant input u and $\hat{T}_{yr}^{H_1}$ is the estimate of the transfer function of the reference r to the plant input y. With this,

$$\hat{P}^{\text{JIO}}(\omega_k) = \hat{T}_{yr}^{H_1}(\omega_k)(\hat{T}_{ur}^{H_1}(\omega_k))^{-1}. \qquad (5.13)$$

As equation (5.10) also has the structure of an instrumental variable estimator with $\mathbf{R}$ the instrumental variable, the JIO estimator is also referred to as such in literature (see e.g. [169]).

In [176], Wernholt and Gunnarsson analyze the properties of the aforementioned estimators when using orthogonal random phase multisine excitation signals, by means of both analysis and simulation. It is shown that in general the H_1 and ARI estimators are asymptotically biased. For the H_1 estimator an asymptotic relative bias expression is given, showing that this bias is approximately proportional to the signal-to-noise ratio. In contrast, the JIO estimator is shown to be asymptotically unbiased (i.e. for $M_B \to \infty$). Furthermore, all estimators give approximately the same variance for sufficiently large SNR. Based on these characteristics, the JIO estimator is considered the estimator of preference for our application.

5.3.3 Estimation of the covariance of JIO estimator

Under specific conditions, including explicit knowledge of the noise disturbance dynamics and the controller, the covariance of the JIO estimator can be computed, as was shown by Wernholt et al. in [176]. Frequently this information is not available. An extensive treatment of the problem of estimating the covariance matrix of FRF estimates of open loop MIMO systems from experimental data under different experimental conditions is given by Pintelon et al. in [129, 130]. Literature however lacks results to estimate the covariance of the FRF obtained using the JIO estimator. In this section we give an approach to this for the case that orthogonal random phase multisine excitation is used. We will first show that under orthogonal excitation the JIO estimator can be reformulated using sample arithmetic mean estimate expressions of T_{ur} and T_{yr}. Next, by computing the sample covariance of T_{ur} and T_{yr}, which is done in a similar way as performed by Pintelon, an estimate of the covariance in the JIO estimate of P is obtained by means of a first order approximation.

An alternative expression for the JIO estimator

As was noted in Section 5.3.2, the JIO estimator can be interpreted as the right division of the H_1 FRF estimator of T_{yr} by the H_1 FRF estimate of T_{ur}. However, under specific conditions on the excitation as expressed in the following theorem, the JIO estimator can also be interpreted as the right matrix division of the arithmetic mean estimate of the FRF of T_{yr} by the arithmetic mean estimate of the FRF of T_{ur}, where both arithmetic mean estimates are taken over a sample set of $M_B \times K$ FRF estimates. For this, first note from equation (5.1) that $\mathbf{U}^{[i]}(\omega_k)$ and $\mathbf{Y}^{[i]}(\omega_k)$ are averaged over multiple periods of the multisine. Hence we can express

$$\mathbf{U}^{[i]}(\omega_k) = \frac{1}{K} \sum_{j=1}^{K} \mathbf{U}^{[i,j]}(\omega_k) \tag{5.14}$$

$$\mathbf{Y}^{[i]}(\omega_k) = \frac{1}{K} \sum_{j=1}^{K} \mathbf{Y}^{[i,j]}(\omega_k) \tag{5.15}$$

where $\mathbf{U}^{[i,j]}$ and $\mathbf{Y}^{[i,j]}$ are used to denote the DFT of the jth period of the ith experiment block of the input and output of P respectively. Let $\hat{T}_{ur}^{[i,j]}$ and $\hat{T}_{yr}^{[i,j]}$ denote the H$_1$ estimate of T_{ur} and T_{yr} from $\mathbf{U}^{[i,j]}$ and $\mathbf{Y}^{[i,j]}$ respectively. With this we give the following proposition.

Proposition 5.3.1 *Let the estimates $\hat{T}_{ur}(\omega_k)$ and $\hat{T}_{yr}(\omega_k)$ be obtained by averaging $\hat{T}_{ur}^{[i,j]}$ and $\hat{T}_{yr}^{[i,j]}$ for all periods K and all experiment blocks M_B:*

$$\hat{T}_{ur}(\omega_k) = \frac{1}{M_B K} \sum_{i=1}^{M_B} \sum_{j=1}^{K} \hat{T}_{ur}^{[i,j]}(\omega_k) \tag{5.16}$$

$$\hat{T}_{yr}(\omega_k) = \frac{1}{M_B K} \sum_{i=1}^{M_B} \sum_{j=1}^{K} \hat{T}_{yr}^{[i,j]}(\omega_k). \tag{5.17}$$

Then if the excitation satisfies

$$\mathbf{R}^{[i]}(\omega_k) = \mathbf{R}_{\text{diag}}^{[i]}(\omega_k)\mathbf{W}, \text{ with } \mathbf{W}\mathbf{W}^H = mI, \text{ for all } i = 1 \ldots M_B \tag{5.18a}$$

$$(\mathbf{R}_{\text{diag}}^{[i]}(\omega_k))(\mathbf{R}_{\text{diag}}^{[i]}(\omega_k))^H =: \mathcal{A}(\omega_k), \text{ for all } i = 1 \ldots M_B \tag{5.18b}$$

with $\mathbf{R}_{\text{diag}}^{[i]}$ defined as in (5.4) and $\mathcal{A}$ an invertible real diagonal matrix, then

$$\hat{P}_{\text{JIO}}(\omega_k) = \hat{T}_{yr}(\omega_k)\hat{T}_{ur}^{-1}(\omega_k) \tag{5.19}$$

Proof We show that under the given conditions, the estimate $\hat{T}_{ur}(\omega_k)$ in equation (5.16) is equivalent to the H$_1$ estimator of T_{ur}:

$$\hat{T}_{ur}(\omega_k) = \frac{1}{M_B K} \sum_{i=1}^{M_B} \sum_{j=1}^{K} \hat{T}_{ur}^{[i,j]}(\omega_k) \tag{5.20}$$

$$= \frac{1}{M_B K} \sum_{i=1}^{M_B} \sum_{j=1}^{K} \mathbf{U}^{[i,j]}(\omega_k)(\mathbf{R}^{[i]}(\omega_k))^H (\mathbf{R}^{[i]}(\omega_k)(\mathbf{R}^{[i]}(\omega_k))^H)^{-1} \tag{5.21}$$

$$= \frac{1}{m M_B K} \sum_{i=1}^{M_B} \sum_{j=1}^{K} \mathbf{U}^{[i,j]}(\omega_k)(\mathbf{R}^{[i]}(\omega_k))^H \mathcal{A}^{-1}(\omega_k) \tag{5.22}$$

$$= \frac{1}{m M_B} \sum_{i=1}^{M_B} \mathbf{U}^{[i]}(\omega_k)(\mathbf{R}^{[i]}(\omega_k))^H \mathcal{A}^{-1}(\omega_k) \tag{5.23}$$

$$= \mathbf{U}(\omega_k)\mathbf{R}^H(\omega_k)(\mathbf{R}(\omega_k)\mathbf{R}^H(\omega_k))^{-1} \tag{5.24}$$

$$= \hat{T}_{ur}^{\text{H}_1}(\omega_k). \tag{5.25}$$

Analogously, it follows that $\hat{T}_{yr}(\omega_k)$ in equation (5.17) is equivalent to the H$_1$ estimator of T_{yr}. Then from equation (5.13), equation (5.19) is immediate. ∎

Observe that equation (5.18) implies that the excitation is orthogonal with constant excitation amplitude in all experiment blocks. We remark that although this proposition is formulated for the case orthogonal excitation is used, orthogonality is not necessary. As long as $\mathbf{R}^{[i]}(\omega_k)(\mathbf{R}^{[i]}(\omega_k))^H$ is the same in every experiment block, the same result is obtained.

This result can be used to estimate to covariance of $\hat{P}^{\text{JIO}}$. Indeed, since the estimators $\hat{T}_{ur}$ and $\hat{T}_{yr}$ are sample average type expressions, their covariances can be estimated easily using sample covariance expressions, as will be discussed in the next subsection. Subsequently, it will be shown how these sample covariances can be used to obtain an estimate of the covariance of $\hat{P}^{\text{JIO}}$.

Sample covariance of $\hat{T}_{ur}$ and $\hat{T}_{yr}$

To estimate the joint covariance of $\hat{T}_{ur}(\omega_k)$ and $\hat{T}_{yr}(\omega_k)$, we introduce the following notation. Let A and B be complex stochastic matrix variables of arbitrary dimensions. We denote the mean and covariance of these matrices as

$$\bar{A} := \mathbb{E}A, \quad \bar{B} := \mathbb{E}B$$

$$\begin{bmatrix} \text{COV}\{A\} & \text{COV}\{A, B\} \\ \text{COV}\{B, A\} & \text{COV}\{B\} \end{bmatrix} := \mathbb{E}\begin{pmatrix} \text{vec}[A - \bar{A}] \\ \text{vec}[B - \bar{B}] \end{pmatrix}\begin{pmatrix} \text{vec}[A - \bar{A}] \\ \text{vec}[B - \bar{B}] \end{pmatrix}^H.$$

With this we define the joint covariance matrix of $\hat{T}_{ur}(\omega_k)$ and $\hat{T}_{yr}(\omega_k)$ as follows:

Definition 5.3.2 *The joint covariance matrix of estimates $\hat{T}_{ur}(\omega_k)$ and $\hat{T}_{yr}(\omega_k)$ is defined as*

$$C_T(\omega_k) := \begin{bmatrix} \text{COV}\{\hat{T}_{ur}\} & \text{COV}\{\hat{T}_{ur}, \hat{T}_{yr}\} \\ \text{COV}\{\hat{T}_{yr}, \hat{T}_{ur}\} & \text{COV}\{\hat{T}_{yr}\} \end{bmatrix} \tag{5.26}$$

For each experiment block, the matrix $C_T(\omega_k)$ can be estimated using a sample covariance expression over all periods. Omitting the frequency argument for brevity, this expression is

$$\hat{C}_T^{[i]} = \frac{1}{M_B K(K-1)} \sum_{j=1}^{K} \begin{pmatrix} \text{vec}[\hat{T}_{ur}^{[i,j]} - \hat{T}_{ur}^{[i]}] \\ \text{vec}[\hat{T}_{yr}^{[i,j]} - \hat{T}_{yr}^{[i]}] \end{pmatrix}\begin{pmatrix} \text{vec}[\hat{T}_{ur}^{[i,j]} - \hat{T}_{ur}^{[i]}] \\ \text{vec}[\hat{T}_{yr}^{[i,j]} - \hat{T}_{yr}^{[i]}] \end{pmatrix}^H \tag{5.27}$$

where

$$\hat{T}_{ur}^{[i]} = \frac{1}{K}\sum_{j=1}^{K}\hat{T}_{ur}^{[i,j]} \qquad\qquad \hat{T}_{yr}^{[i]} = \frac{1}{K}\sum_{j=1}^{K}\hat{T}_{yr}^{[i,j]}. \tag{5.28}$$

Averaging over all experiment blocks yields

$$\hat{C}_T = \frac{1}{M_B} \sum_{i=1}^{M_B} C_T^{[i]} \tag{5.29}$$

$$= \frac{1}{M_B^2 K(K-1)} \sum_{i=1}^{M_B} \sum_{j=1}^{K} \begin{pmatrix} \mathrm{vec}[\hat{T}_{ur}^{[i,j]} - \hat{T}_{ur}^{[i]}] \\ \mathrm{vec}[\hat{T}_{yr}^{[i,j]} - \hat{T}_{yr}^{[i]}] \end{pmatrix} \begin{pmatrix} \mathrm{vec}[\hat{T}_{ur}^{[i,j]} - \hat{T}_{ur}^{[i]}] \\ \mathrm{vec}[\hat{T}_{yr}^{[i,j]} - \hat{T}_{yr}^{[i]}] \end{pmatrix}^H \tag{5.30}$$

An expression for the covariance of the JIO estimator

Starting from expression (5.19), the purpose of this section is to derive an expression for the covariance matrix of the JIO estimate of P. Since $\hat{P}^{\mathrm{JIO}}$ is a nonlinear function of $\hat{T}_{ur}$ and $\hat{T}_{yr}$, it is in general unfeasible to calculate the covariance matrix of $\hat{P}^{\mathrm{JIO}}$ from the joint covariance of $\hat{T}_{ur}$ and $\hat{T}_{yr}$. An approximation can be calculated through a first order linearization of the expression of the JIO estimator. To formulate that, we first need the following lemma.

Lemma 5.3.3 *Let A and B be two correlated complex stochastic matrices and let the variation of A and B be small in comparison to the mean value $\bar{A}$ and $\bar{B}$. Given is the map $C = f(A, B)$, which is analytical in the point $A = \bar{A}$, $B = \bar{B}$. Then an approximation of $\mathrm{COV}\{C\}$ through first order linearization of $f(A, B)$ around $\bar{A}$ and $\bar{B}$ is given by*

$$\mathrm{COV}\{C\} \approx \begin{bmatrix} \left.\frac{\partial \mathrm{vec}[f(A,B)]}{\partial \mathrm{vec}[A]}\right|_{A=\bar{A}} & \left.\frac{\partial \mathrm{vec}[f(A,B)]}{\partial \mathrm{vec}[B]}\right|_{B=\bar{B}} \end{bmatrix} \cdot$$
$$\cdot \begin{bmatrix} \mathrm{COV}\{A\} & \mathrm{COV}\{A,B\} \\ \mathrm{COV}\{B,A\} & \mathrm{COV}\{B\} \end{bmatrix} \begin{bmatrix} \left(\left.\frac{\partial \mathrm{vec}[f(A,B)]}{\partial \mathrm{vec}[A]}\right|_{A=\bar{A}}\right)^H \\ \left(\left.\frac{\partial \mathrm{vec}[f(A,B)]}{\partial \mathrm{vec}[B]}\right|_{B=\bar{B}}\right)^H \end{bmatrix} \tag{5.31}$$

Proof A proof for the case A and B are real random variables is given in [86], Section 10.6, for the case A and B are real random vectors in [133], Section 2.13.1, and for the case A and B are complex vectors in [128], Section 14.2. This lemma is the generalization for the case A and B are complex matrices and is proved similarly as in the aforementioned references. ■

We can now formulate the following proposition:

Proposition 5.3.4 *Through first order linearization of the expression of the JIO estimator, an approximation of the covariance of the JIO estimate of P is given by*

$$C_P(\omega_k) := \mathrm{COV}\{\hat{P}^{\mathrm{JIO}}(\omega_k)\}$$
$$\approx M_P(\omega_k) C_T(\omega_k) M_P^H(\omega_k) \tag{5.32}$$

with

$$M_P(\omega_k) = \begin{bmatrix} -\hat{T}_{ur}^{-T} \otimes \hat{P}^{\mathrm{JIO}}(\omega_k) & \hat{T}_{ur}^{-T}(\omega_k) \otimes I \end{bmatrix} \tag{5.33}$$

Proof First define

$$h_1(\omega_k) = \text{vec}[T_{ur}(\omega_k)] \qquad\qquad h_2(\omega_k) = \text{vec}[T_{yr}(\omega_k)]$$
$$h(\omega_k) = \text{col}(h_1(\omega_k), h_2(\omega_k)).$$

Application of lemma 5.3.3 immediately yields (5.32), where $M_P(\omega_k)$ is given by

$$M_P(\omega_k) := \frac{\partial}{\partial \hat{h}} \text{vec}[\hat{P}^{\text{JIO}}(\omega_k)]$$
$$= \left[\begin{array}{cc} \frac{\partial}{\partial \hat{h}_1} \text{vec}[\hat{P}^{\text{JIO}}(\omega_k)] & \frac{\partial}{\partial \hat{h}_2} \text{vec}[\hat{P}^{\text{JIO}}(\omega_k)] \end{array} \right]. \qquad (5.34)$$

Using equation (5.19) and the identities $\text{vec}(AB) = (I \otimes A)\text{vec}B = (B^T \otimes I)\text{vec}A$, we obtain

$$\text{vec}[\hat{P}^{\text{JIO}}(\omega_k)] = [\hat{T}_{ur}^{-T}(\omega_k) \otimes I]\hat{h}_2(\omega_k) \qquad (5.35a)$$
$$= [I \otimes \hat{T}_{yr}(\omega_k)]\text{vec}[\hat{T}_{ur}^{-1}(\omega_k)]. \qquad (5.35b)$$

The first element of M_P in equation (5.34) can be derived using equation (5.35b) and lemma B.2.1, resulting in

$$\frac{\partial}{\partial \hat{h}_1} \text{vec}[\hat{P}^{\text{JIO}}(\omega_k)] = [I \otimes \hat{T}_{yr}(\omega_k)] \frac{\partial}{\partial h_1} \text{vec}[\hat{T}_{ur}^{-1}(\omega_k)]$$
$$= [I \otimes \hat{T}_{yr}(\omega_k)][-\hat{T}_{ur}^{-T}(\omega_k) \otimes \hat{T}_{ur}^{-1}(\omega_k)]$$
$$= [-\hat{T}_{ur}^{-T}(\omega_k) \otimes \hat{P}^{\text{JIO}}(\omega_k)].$$

The second element of M_P in equation (5.34) directly follows taking the derivative of equation (5.35a) to $\hat{h}_2$, yielding

$$\frac{\partial}{\partial \hat{h}_2} \text{vec}[\hat{P}^{\text{JIO}}(\omega_k)] = [\hat{T}_{ur}^{-T}(\omega_k) \otimes I],$$

which completes the proof. ∎

We summarize the obtained results so far. In the previous Section 5.3.2 an approach to estimation of the multivariable FRF of the AMB spindle from closed-loop data was described. This approach has the attractive properties of being asymptotically unbiased and yielding low covariance. The above results allow to estimate the covariance in this FRF estimate. Having this covariance estimates available will prove valuable in stage 2, when a parametric model is estimated using the estimated FRF. As will become clear in Section 5.6, the estimate of the covariance of the FRF can be used as a weighting function to reduce the variance error of the model parameter estimates.

5.4   Reduction of nonlinear distortion in the FRF estimate

We recall that the key problem of this chapter is to estimate a parametric linear model of P from sequences of measured data. However, as was discussed in Chapter 2, plant P has several sources of nonlinearities. Still, as a linear model of P in its operating point would suffice, a more precise formulation of the objective is to estimate a model of the first order linearization of the AMB spindle system in its operating point. In order to achieve this goal, in Section 5.4.1 an approach is given that allows to detect the level of nonlinear distortion in the estimate of the FRF of the first order linearization. The given approach is based on the variance analysis method of Pintelon et al. [131, 145], which detects nonlinear distortions in *direct* FRF estimates. The approach in this section extends the variance analysis method for detection of the level of nonlinear distortions in a JIO estimate of the FRF of a multivariable plant. Being able to quantify the level of nonlinear distortion, the next step is to improve the experimental conditions, so that the effect of the nonlinearities on the FRF estimates is small. We will discuss in particular what steps can be taken when dealing with closed-loop experiments (Section 5.4.2).

5.4.1   Detection of the level of nonlinearity in the JIO estimate

In literature, numerous approaches are available to detect distortions orginating from nonlinearities in a system (for an overview, see [167]). When estimating the FRF using periodic excitation signals, two basic strategies are available. The first approach is to use a special multisine signal, so that selected even and odd frequences are not excited. At these frequencies the nonlinearities can be detected by measuring the output levels (see [145] for more details). The second approach is the variance analysis approach of Pintelon et al. [131]. In this approach for detection of nonlinear distortions in a SISO system, a random phase multisine excitation signal is generated and multiple periods of this signal are applied to the system. For each period of the multisine, an estimate of the FRF is calculated. This procedure is then repeated for different realizations of the multisine. Detection of nonlinear distortions is achieved by comparing the variations of the estimated FRF for different periods of the same realization of the multisine to the variations of the estimated FRF for different realizations of the random phases of the excitation. The former variations only depend on the measurement noise, whereas the latter depend on both measurement noise and distortions, due to nonlinearities. Hence, if both variances are the same, no significant nonlinearities are present. However, if the latter variance level is significantly larger, this indicates the presence of nonlinear distortions in the FRF estimate. Wernholt [177] extended this method for detection of nonlinearities of MIMO systems in closed loop. In this work, the ARI estimator was used (which essentially ignores the closed-loop effects), and variance calculations were performed for each element of the FRF matrix estimate individually. Here we further extend the approach to estimating the level of nonlinear distortion in the

JIO estimate of P. Instead of element-wise variance calculation, we perform variance analysis by estimating the full covariance matrix of $\hat{P}^{JIO}$, using the results of the previous section.

Pintelon and Schoukens [128, 145] showed that when a nonlinear SISO system satisfying some formal, but rather unrestrictive conditions[1] is excited with a random phase multisine excitation signal, the nonlinearity appears in the DFT of the output as a systematic bias, and as a noise-like contribution that varies with the realization of the random phases of the sinusoids of the excitation. Dobrowiecki et al. extended this in [42] to nonlinear MIMO systems. They showed that the output of a nonlinear system G satisfying similar conditions as in the SISO case, excited with a random phase multisine signal, can be expressed as

$$Y(\omega_k) = (G_0(\omega_k) + G_B(\omega_k))U(\omega_k) + Y_S(\omega_k) + N_G(\omega_k) \qquad (5.36)$$

where G_0 is the first order linearization of G in the selected operating point, G_B represents the bias due to the nonlinear distortions, Y_S is the noise-like nonlinear contribution, and N_G is a noise term due to measurement noise. As shown by Dobrowiecki et al., G_B depends on the spectrum of the input only, while the nonlinear contribution $Y_S(\omega_k)$ has the character of a noise source: it has zero mean ($\mathbb{E}_\phi Y_S(\omega_k) = 0$ — where $\mathbb{E}_\phi$ is used to denote the expectation operator over the realizations of random phases of the multisine), for every frequency ω_k it is uncorrelated with the input ($\mathbb{E}_\phi Y_S(\omega_k)U^H(\omega_k) = 0$), and it has random phase.

We will apply this result to the closed-loop configuration of the AMB spindle. Nonlinearity of P implies nonlinearity of the mappings from $r \to u$ and $r \to y$. In the JIO estimator, the FRF of the first order linearization of these mappings is estimated and combined according to equation (5.19) to make an estimate of the FRF of P. Hence, errors in the estimates of T_{ur} and T_{ur} due to nonlinear distortion, propagate in the JIO estimate of the FRF of P. In what follows, we will give a measure of the nonlinear distortion in $\hat{T}_{ur}$ and $\hat{T}_{yr}$, and use this to detect the level of nonlinear distortion in the JIO estimate of the FRF of P.

As shown in the previous section, the FRF estimates $\hat{T}_{ur}$ and $\hat{T}_{yr}$ are obtained by averaging over the sample sets $\hat{T}_{ur}^{[i,j]}(\omega_k)$, $\hat{T}_{yr}^{[i,j]}(\omega_k)$, with $i = 1 \ldots M_B$ and $j = 1 \ldots K$. Applying the result in equation (5.36), we can express for $\hat{T}_{ur}^{[i,j]}(\omega_k)$ and $\hat{T}_{yr}^{[i,j]}(\omega_k)$

$$\hat{T}_{ur}^{[i,j]}(\omega_k) = T_{ur,0}(\omega_k) + T_{ur,B}(\omega_k) + T_{ur,S}^{[i]}(\omega_k) + N_{T_{ur}}^{[i,j]}(\omega_k) \qquad (5.37a)$$

$$\hat{T}_{yr}^{[i,j]}(\omega_k) = T_{yr,0}(\omega_k) + T_{yr,B}(\omega_k) + T_{yr,S}^{[i]}(\omega_k) + N_{T_{yr}}^{[i,j]}(\omega_k) \qquad (5.37b)$$

---

[1]The analysis by Pintelon and Schoukens in [128, 145] is performed for systems for which there exists a Volterra series representation that converges to the output in the mean square sense with probability 1 for all excitations in the class of allowable inputs. This class of input signals is defined as the set of signals with a Gaussian amplitude distribution and a user-defined power spectrum, which includes the sets of Gaussian noise, periodic noise and random phase multisine signals.

where $T_{ur,0}$ is the first order linearization of T_{ur}, $T_{ur,B}$ is the nonlinear bias, $T_{ur,S}^{[i]}$ is the nonlinear disturbance term (which only depends on the realization of the random phase, hence the superscript $[i]$), and $N_{T_{ur}}^{[i,j]}$ is the error term due to measurement noise (which varies with each period and experiment block, hence the superscript $[i,j]$). The terms $T_{yr,0}$, $T_{yr,B}$, $T_{yr,S}^{[i]}$, and $N_{T_{yr}}^{[i,j]}$ are defined similarly.

The variance analysis approach of Pintelon et al. in [131] entails the following. From (5.37a), we see that $\hat{T}_{ur}^{[i,j]}(\omega_k)$ has two disturbance sources, where $T_{ur,S}^{[i]}(\omega_k)$ is independent over i, but equal over j, and $N_{T_{ur}}^{[i,j]}(\omega_k)$ is independent over both i and j. The same holds for $\hat{T}_{ur}^{[i,j]}(\omega_k)$. This implies that under presence of $T_{ur,S}^{[i]}$ and $T_{yr,S}^{[i]}$, the covariance of T_{ur} and T_{yr} taken over the multiple periods of the excitation only, will differ from the covariance taken over both the periods of the excitation and the realizations of the random phases of the excitation. This can be quantified as follows. For convenience, we repeat the sample joint covariance expression of $\hat{T}_{ur}$ and $\hat{T}_{yr}$ as defined in equation (5.29) in the previous section:

$$\hat{C}_{T,I} = \frac{1}{M_B^2 K(K-1)} \sum_{i=1}^{M_B} \sum_{j=1}^{K} \begin{pmatrix} \text{vec}[\hat{T}_{ur}^{[i,j]} - \hat{T}_{ur}^{[i]}] \\ \text{vec}[\hat{T}_{yr}^{[i,j]} - \hat{T}_{yr}^{[i]}] \end{pmatrix} \begin{pmatrix} \text{vec}[\hat{T}_{ur}^{[i,j]} - \hat{T}_{ur}^{[i]}] \\ \text{vec}[\hat{T}_{yr}^{[i,j]} - \hat{T}_{yr}^{[i]}] \end{pmatrix}^H .$$

(5.38)

As this covariance expression incorporates only variations over the periods, observe that it will not depend on $T_{ur,S}^{[i]}$ and $T_{yr,S}^{[i]}$. In contrast to this, the covariance computed over all experiment blocks and periods, i.e.

$$\hat{C}_{T,II} = \frac{1}{M_B K(M_B K-1)} \sum_{i=1}^{M_B} \sum_{j=1}^{K} \begin{pmatrix} \text{vec}[\hat{T}_{ur}^{[i,j]} - \hat{T}_{ur}] \\ \text{vec}[\hat{T}_{yr}^{[i,j]} - \hat{T}_{yr}] \end{pmatrix} \begin{pmatrix} \text{vec}[\hat{T}_{ur}^{[i,j]} - \hat{T}_{ur}] \\ \text{vec}[\hat{T}_{yr}^{[i,j]} - \hat{T}_{yr}] \end{pmatrix}^H$$

(5.39)

will depend on $T_{ur,S}^{[i]}$ and $T_{yr,S}^{[i]}$. Hence nonlinear distortion in $\hat{T}_{ur}$ and $\hat{T}_{yr}$ can be detected by verifying if the elements of $\hat{C}_{T,II}$ are significantly larger than those of $\hat{C}_{T,I}$.

Now, using the estimates $C_{T,I}$ and $C_{T,II}$ and applying proposition 5.3.4, we can also estimate the covariance of $\hat{P}^{\text{JIO}}$ over periods of the same realization of the excitation (denoted as $\hat{C}_{P,I}$), as well as the covariance over different realizations of the excitation (denoted as $\hat{C}_{P,II}$):

$$\hat{C}_{P,I}(\omega_k) = M_P(\omega_k)\hat{C}_{T,I}(\omega_k)M_P^H(\omega_k) \qquad (5.40)$$

$$\hat{C}_{P,II}(\omega_k) = M_P(\omega_k)\hat{C}_{T,II}(\omega_k)M_P^H(\omega_k). \qquad (5.41)$$

Again, if both variances are the same, we can conclude that no significant nonlinear distortion is present in the estimate P^{JIO}. However, if the elements of $\hat{C}_{P,II}$ are

significantly larger than those of $\hat{C}_{P,I}$, this indicates the presence of a nonlinear distortion in the FRF estimate $\hat{P}^{JIO}$.

For the sake of completeness, we remark that strictly speaking, a statistical test is necessary to determine if $\hat{C}_{P,II}$ is indeed significantly larger than $\hat{C}_{P,I}$. However, in practice we will be interested to detect when the nonlinear distortion dominates over the measurement noise, in which cases $\hat{C}_{P,II}$ is larger than $\hat{C}_{P,I}$ by several times. Then carrying out such a statistical test will be superfluous.

We conclude by recalling from equation (5.36) that nonlinearities in a system, excited by a random-phase multisine, arise in the DFT of its output in two ways, i.e. in the form of a nonlinear bias (G_B) and a noise-like nonlinear contribution (Y_S). The approach in this section focuses only on detecting the level of the latter. Additional research is needed to perform detection of the nonlinear bias. Literature is scarse on approaches to achieve this.

5.4.2 Approach to minimizing the level of nonlinear distortion in the FRF estimate

In literature several recommendations are available to minimize the level of nonlinear distortions on the FRF estimate (see e.g. [128, 145]). These recommendations are summarized as follows:

A. Reduce the level of the nonlinear distortion itself. This can be achieved by choosing a multisine excitation signal with odd frequencies only, so that only distortions from the odd nonlinearities are measured (even nonlinearities cause the nonlinear distortions of odd excitation frequencies to appear at even frequencies only, at which the FRF is not measured [128]). The use of orthogonal multisine excitation has also been shown to reduce the level of nonlinear distortion [42]. In general, reduction of the crest factor of the input signal is also recommended, although for closed-loop experiments it is not straightforward to design excitation signals that minimize the crest factor of the input of the plant.

B. Use the fact that Y_S has zero mean (with expectation taken over experiments): perform experiment repetition with different realizations of the random phases of the multisine.

Here we add a third approach to reduce the level of nonlinear distortions on the FRF, particular when dealing with closed-loop identification:

C. Improve the signal to nonlinear noise ratio. Observe that in equation (5.37a) $T_{ur,S}^{[i]}$ and $T_{yr,S}^{[i]}$ are obtained by

$$T_{ur,S}^{[i]} = \mathbf{U}_S^{[i]}(\mathbf{R}^{[i]})^{-1} \tag{5.42}$$

$$T_{yr,S}^{[i]} = \mathbf{Y}_S^{[i]}(\mathbf{R}^{[i]})^{-1} \tag{5.43}$$

where $\mathbf{U}_S^{[i]}$ and $\mathbf{Y}_S^{[i]}$ represent the stochastic nonlinear noise in the DFT of u and y respectively in experiment block i. From this we infer that reduction of the level of nonlinear distortions in the estimated FRF of T_{ur} and T_{yr}, and consequently also of P, can be achieved by choosing the experimental conditions such, that the above quotients are reduced. This is in particular a feasible strategy in the frequency regions where T_{ur} is small and T_{yr} is close to the identity. This is as seen as follows. In the region where T_{yr} is close to the identity, the level of nonlinear distortion is predominantly related to the amplitude of the input signal u, as any nonlinear behavior at the output is corrected for by the controller. As T_{ur} is small, increasing the amplitude of the excitation in these frequency ranges will not result in a significant increase of the nonlinear noise terms $\mathbf{U}_S^{[i]}$ and $\mathbf{Y}_S^{[i]}$. However, $T_{ur,S}^{[i]}$ and $T_{yr,S}^{[i]}$ are reduced due to the increased size of $\mathbf{R}^{[i]}$, yielding the reduced nonlinear distortion in the FRF estimates.

We remark that in some cases better estimates can be obtained by choosing a different FRF estimator. As was pointed out by Wernholt et al. in [178], when dealing with nonlinearities in the loop, the JIO estimator might not necessarily yield the smallest bias and variance. As it relies on estimates of closed-loop transfer functions, additional errors might be introduced, in particular when dealing with a setup with nonlinearities in the controller. As shown by Wernholt et al., in such cases the ARI estimator or variants with nonlinear averaging might give better results.

5.4.3 Results

The procedure that was described in the previous sections has been applied to the AMB spindle from EAAT that was discussed in Chapter 2. For the experiments the following settings were selected:

- Number of experiment blocks $M_B = 20$, i.e. the total number of experiments is 80;

- Number of periods per experiment $K = 12$;

- Period length: $N_K = 16384$;

- Sampling frequency: 20 kHz, hence $T_s = 50\mu s$;

- The frequency range of interest is 0 to 5 kHz. With odd excitation, this implies that amplitudes of the multisine A_k (we will omit the subscript l for the input number from hereon) are nonzero only for $k = 1, 3, \ldots \frac{N_K}{4}$;

- The rotational speed in these experiments was 0 rpm[2].

---

[2]These experiments were carried out at standstill to eliminate any runout distortions. In later sections we will show results obtained at other rotational speeds.

The first set of three tests involved excitation with a flat spectrum, with increased amplitude level for subsequent tests: $A_k = \{0.025, 0.100, 0.200\}$ for all $k = 1, 3, \ldots \frac{N_k}{4}$. To detect the level of the nonlinear behavior of the AMB spindle in the FRF obtained with each of these datasets, $C_{P,I}$ and $C_{P,II}$ were estimated. For reasons of space — the dimensions of these covariance matrices are 16×16 — we only compare the square root of the $(1,1)$ element in Figure 5.2. These plots confirm that for increased excitation amplitude, the covariance at higher frequencies decreases. Indeed, noting that $C_{P,I(1,1)}$ only depends on the signal to measurement noise ratio, it is logical that $C_{P,I(1,1)}$ decreases for increased excitation amplitude. At high frequencies, no significant differences between $C_{P,I(1,1)}$ and $C_{P,II(1,1)}$ are observed either (except in the frequency ranges around the resonance modes of the spindle). In contrast, at lower frequencies it is observed that $C_{P,II(1,1)}$ is larger than $C_{P,I(1,1)}$, where the difference increases for increased excitation amplitude. It is seen that also at low frequencies $C_{P,I(1,1)}$ decreases due to the improved signal to measurement noise ratio, but in contrast $C_{P,II(1,1)}$ remains on the same level (in fact, closer inspection shows a slight increase). This is explained with a nonlinear distortion that grows proportionally with the excitation amplitude.

From these results, we conclude that the error in $\hat{P}^{JIO}$ in the lower frequency range is dominated by the nonlinearities. Hence, as indicated in Section 5.4.2, experiment repetition with different realizations of the random phases is advisable to reduce these errors. This might take a high number of experiment repetitions to obtain an acceptable error level. Indeed, observe that the error that is depicted in Figure 5.2 is the result of averaging over 20 experiment blocks (hence 80 experiments were performed, each of which involved measurement of 12 periods of 16384 samples). Even with this large amount of data, the standard deviation of the elements of $\hat{P}^{JIO}$ is in the 0-10 dB range for lower frequencies. For the off-diagonal elements of P this means that the estimation error is in the same size order as the magnitude of P (in fact, for some elements of P, the estimation error is even an order of magnitude larger).

Based on these outcomes, we apply a second test where we change the excitation spectrum according to approach C. described in Section 5.4.2, to reduce the level of nonlinear distortion in the FRF estimate in closed-loop identification. This yields the excitation spectrum in Figure 5.3(a). Compared to the flat excitation spectrum with $A_k = 0.10$, the excitation amplitude at low frequencies is increased up to 25 times. The other experimental conditions are kept the same.

In Figure 5.3(b) we still observe that $C_{P,II(1,1)}$ is significantly larger than $C_{P,I(1,1)}$, again indicating presence of nonlinear disturbances. However, the level of $\sqrt{C_{P,II(1,1)}}$ is reduced to around -20 dB. Indeed, as a result of the higher excitation amplitude at low frequencies, $\sqrt{C_{P,II(1,1)}}$ has been reduced by up to 25 dB. In Figure 5.4, the FRF obtained with the improved excitation spectrum is compared with the FRF obtained with the data from the experiment with constant excitation with $A_k = 0.1$.

(a) $A_k = 0.025\ \mu$m for $k = 1, 3, 5, \dots, 4095$

(b) $A_k = 0.1\ \mu$m for $k = 1, 3, 5, \dots, 4095$

(c) $A_k = 0.20\ \mu$m for $k = 1, 3, 5, \dots, 4095$
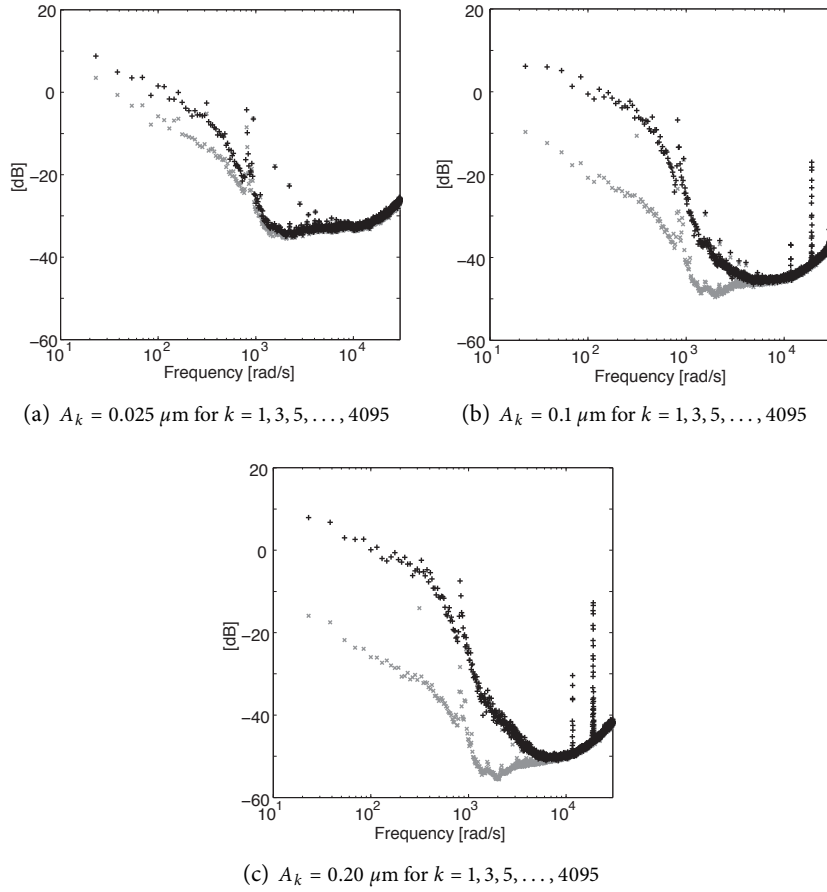
Figure 5.2: Results with excitation with constant amplitude spectrum: comparison of the square root of element (1,1) of $C_{P,I}(\omega_k)$ (grey x), and $C_{P,II}(\omega_k)$ (black +) for different excitation levels.

5.5 Estimation of non-parametric noise models

In the previous section we used the data obtained by periodic excitation of the closed-loop to make a non-parametric estimate of the FRF of the plant. In this section we will use the periodic nature of the excitation to obtain non-parametric estimates of the spectra of v_1 and v_2 and their cross spectrum. This is an extension of the approach for non-parametric estimation of the noise covariance in Pintelon and Schoukens, Section 2.5.1 [128], for the multivariable closed-loop configuration that we consider. For the EAAT spindle, the results of this procedure are used to

(a) Comparison of the flat excitation spectrum with $A_k = 0.100$ for $k = 1, 3, \ldots 4095$ (grey x) and the improved excitation spectrum with increased amplitude at the lower frequencies (black +).

(b) Results with excitation amplitude spectrum as depicted in Figure 5.3(a): comparison of the square root of element (1,1) of $C_{P,I}(\omega_k)$ (grey x), and $C_{P,II}(\omega_k)$ (black +).
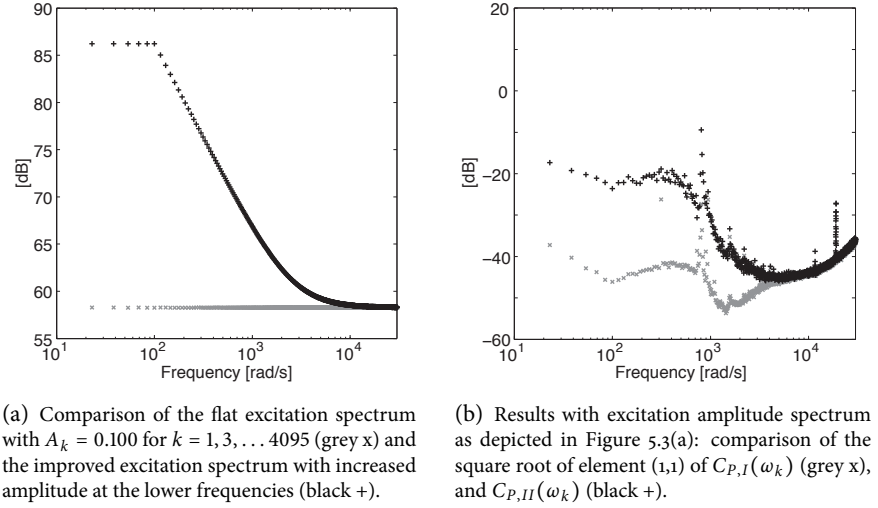
Figure 5.3: Reduction of the nonlinear distortion in the JIO FRF estimate.

justify that measurement noise v_1 is negligible, as well as to demonstrate that v_1 and v_2 are uncorrelated.

5.5.1 Approach

Let us denote the spectral density of the disturbance on the measured signals $u(t)$ and $y(t)$ as Φ_U and Φ_Y, and the cross spectral density by Φ_{UY}. From Figure 5.1 it is straightforward to derive that

$$\begin{pmatrix} \Phi_U & \Phi_{UY} \\ \Phi_{UY}^H & \Phi_Y \end{pmatrix} = \begin{bmatrix} I & KS \\ 0 & S \end{bmatrix} \begin{pmatrix} \Phi_{v_1} & \Phi_{v_1 v_2} \\ \Phi_{v_1 v_2}^H & \Phi_{v_2} \end{pmatrix} \begin{bmatrix} I & KS \\ 0 & S \end{bmatrix}^H \tag{5.44}$$

where S is the sensitivity function as defined in equation (3.7) in Chapter 3, Φ_{v_1}, Φ_{v_2} are the spectral density of v_1 and v_2 respectively, and $\Phi_{v_1 v_2}$ is the cross spectral density of v_1 and v_2. From this it is immediate that

$$\begin{pmatrix} \Phi_{v_1} & \Phi_{v_1 v_2} \\ \Phi_{v_1 v_2}^H & \Phi_{v_2} \end{pmatrix} = \begin{bmatrix} I & -K \\ 0 & S^{-1} \end{bmatrix} \begin{pmatrix} \Phi_U & \Phi_{UY} \\ \Phi_{UY}^H & \Phi_Y \end{pmatrix} \begin{bmatrix} I & -K \\ 0 & S^{-1} \end{bmatrix}^H . \tag{5.45}$$

Using this expression, we will estimate Φ_{v_1}, Φ_{v_2} and $\Phi_{v_1 v_2}$ using estimates of the matrices on the right-hand side. We will start with the center matrix and use the periodic nature of the data to estimate Φ_U, Φ_Y and Φ_{UY}. Recall from Section 5.3.1 that the excitation signal is periodic and that K periods of u and y are recorded,

119
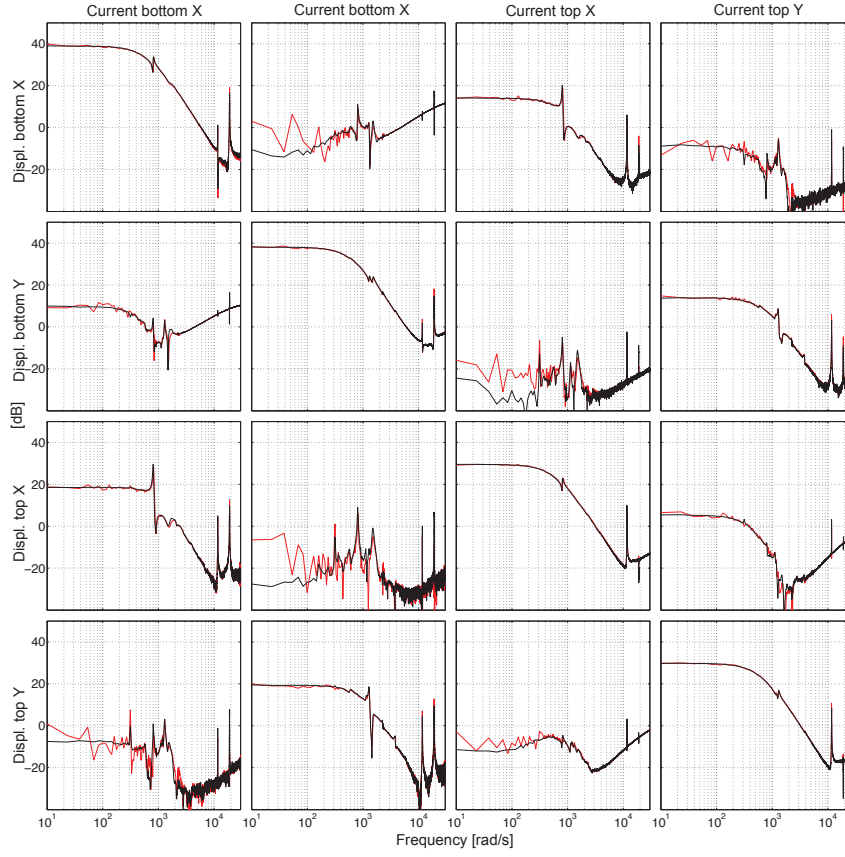
Figure 5.4: Estimated FRF of the EAAT spindle (in [μm/A]) with flat excitation spectrum ($A_k = 0.10$: red), and with increased excitation amplitude in the lower frequency range (black).

which is repeated M times. Let the DFT of u and y in experiment i and period j be denoted by $U^{i,j}(\omega_k)$ and $Y^{i,j}(\omega_k)$ respectively, and let $\bar{U}^i(\omega_k)$ and $\bar{Y}^i(\omega_k)$ represent the noise free part of the DFT of u and y in experiment i, i.e.

$$\bar{U}^i(\omega_k) = \left(T_{ur,0}\left(e^{j\omega_k}\right) + T_{ur,B}\left(e^{j\omega_k}\right) \right) R^i(\omega_k) + U_S^i(\omega_k) \tag{5.46}$$

$$\bar{Y}^i(\omega_k) = \left(T_{yr,0}\left(e^{j\omega_k}\right) + T_{yr,B}\left(e^{j\omega_k}\right) \right) R^i(\omega_k) + Y_S^i(\omega_k). \tag{5.47}$$

With this, we can express

$$U^{i,j}(\omega_k) = \bar{U}^i(\omega_k) + \check{U}^{i,j}(\omega_k) \tag{5.48}$$

$$Y^{i,j}(\omega_k) = \bar{Y}^i(\omega_k) + \check{Y}^{i,j}(\omega_k) \tag{5.49}$$

where $\check{U}^{i,j}(\omega_k)$ and $\check{Y}^{i,j}(\omega_k)$ denote the DFT of the noise in measurements of u and y. Observe that the expressions for $U(\omega_k)$ and $Y(\omega_k)$ in equation (5.1) are in fact sample mean estimates of $\bar{U}(\omega_k)$ and $\bar{Y}(\omega_k)$ respectively over all periods. Similarly, estimates of disturbance spectra Φ_U and Φ_Y at frequencies ω_k can be obtained by calculating the sample covariance matrices of $U(\omega_k)$ and $Y(\omega_k)$, which are given by

$$\hat{\Phi}_U^i(\omega_k) = \frac{1}{K-1} \sum_{j=1}^{K} (U^{i,j}(\omega_k) - U^i(\omega_k))(U^{i,j}(\omega_k) - U^i(\omega_k))^H \tag{5.50}$$

$$\hat{\Phi}_Y^i(\omega_k) = \frac{1}{K-1} \sum_{j=1}^{K} (Y^{i,j}(\omega_k) - Y^i(\omega_k))(Y^{i,j}(\omega_k) - Y^i(\omega_k))^H \tag{5.51}$$

$$\hat{\Phi}_{UY}^i(\omega_k) = \frac{1}{K-1} \sum_{j=1}^{K} (U^{i,j}(\omega_k) - U^i(\omega_k))(Y^{i,j}(\omega_k) - Y^i(\omega_k))^H \tag{5.52}$$

Averaging over the M experiment repetitions yields for $\hat{\Phi}_U(\omega_k)$

$$\hat{\Phi}_U(\omega_k) = \frac{1}{M(K-1)} \sum_{i=1}^{M} \sum_{j=1}^{K} (U^{i,j}(\omega_k) - U^i(\omega_k))(U^{i,j}(\omega_k) - U^i(\omega_k))^H. \tag{5.53}$$

Similar expressions for $\hat{\Phi}_Y(\omega_k)$ and $\hat{\Phi}_{UY}(\omega_k)$ can be given. By application of the JIO estimator we can find estimates for $K(e^{j\omega_k})$ and $S^{-1}(e^{j\omega_k})$, which are given by

$$\hat{S}^{-1}(e^{j\omega_k}) = \mathbf{R}(\omega_k)\mathbf{R}^H(\omega_k)[(\mathbf{Y}(\omega_k) + \mathbf{R}(\omega_k))\mathbf{R}^H(\omega_k)]^{-1} \tag{5.54}$$

$$\hat{K}(e^{j\omega_k}) = \mathbf{U}(\omega_k)\mathbf{R}^H(\omega_k)[(\mathbf{Y}(\omega_k) + \mathbf{R}(\omega_k))\mathbf{R}^H(\omega_k)]^{-1}. \tag{5.55}$$

Substitution of the estimates $\hat{\Phi}_U(\omega_k)$, $\hat{\Phi}_Y(\omega_k)$, $\hat{\Phi}_{UY}(\omega_k)$, $\hat{S}^{-1}(e^{j\omega_k})$ and $\hat{K}(e^{j\omega_k})$ in equation (5.45) yields the desired estimates of Φ_{v_1}, Φ_{v_2} and $\Phi_{v_1 v_2}$.

5.5.2 Results

The procedure in the previous section has been applied to the same data obtained with the EAAT spindle as described in Section 5.4.3. For reasons of space, we only show the results for the bottom X-bearing. In Figure 5.5 a comparison is made between the squared magnitude of the FRF of the current measurement of the X bearing (i.e. $|U_1(\omega_k)|^2$, the subscript 1 indicating the first element of the vector) and the estimated spectral density of the noise on the current measurements (i.e.

5. System identification of the AMB spindle, part I: Bearing dynamics

Figure 5.5: Comparison between the squared magnitude of the FRF of the current measurement of the X bearing (i.e. $|U_1(\omega_k)|^2 [A^2]$) and the estimated spectral density of the noise on the current measurements (i.e. $\hat{\Phi}_{v_1(1,1)}(\omega_k)$). The SNR is well above 40 dB for the almost the entire frequency range.

$\hat{\Phi}_{v_1(1,1)}(\omega_k)$. As can be observed, the SNR is well above 40 dB for almost the entire frequency range. However, around 4.1 krad/s, and multiplies of that frequency, $\hat{\Phi}_{v_1(1,1)}(\omega_k)$ shows a peak. Most likely this is a disturbance frequency of the current amplifier, and indicates an imperfection of its design. For reference, remark that a white noise disturbance with a standard deviation of 5 mA would correspond to a flat spectral density of -46 dB. From this plot, we may conclude that for the EAAT spindle the assumption is justified that the noise on the current measurements is negligible, except at a limited number of frequencies.

Similarly, in Figure 5.6 a comparison is made between the squared magnitude of the FRF of the position measurement of the X bearing (i.e. $|Y_1(\omega_k)|^2$) and the estimated spectral density of the noise on the position measurements (i.e. $\hat{\Phi}_{v_2(1,1)}(\omega_k)$). In this plot we see that the spectral density of the noise on the position measurement is somewhat concentrated at the lower frequencies. Also, peaks can be observed at $50 \cdot 2\pi$ rad/s, and multiples of this frequency, indicating that the 50 Hz of the mains is present as a disturbance in the position measurement.

The spectra of the noise on the current and position measurement of the other bearings show a similar picture. Furthermore, it was observed that the off-diagonal
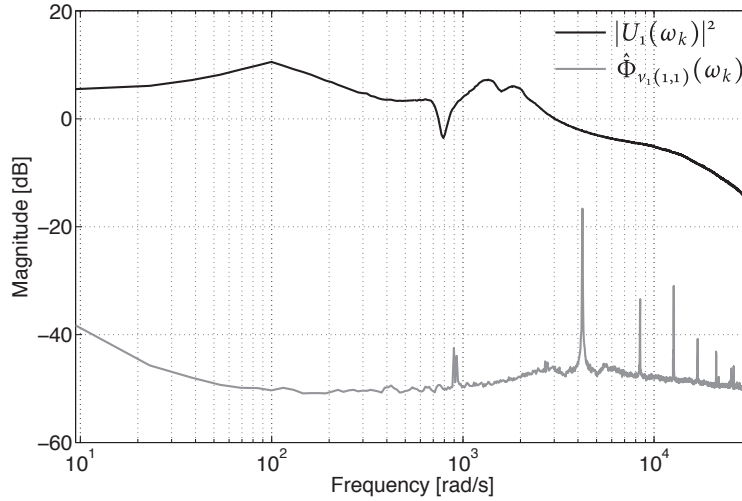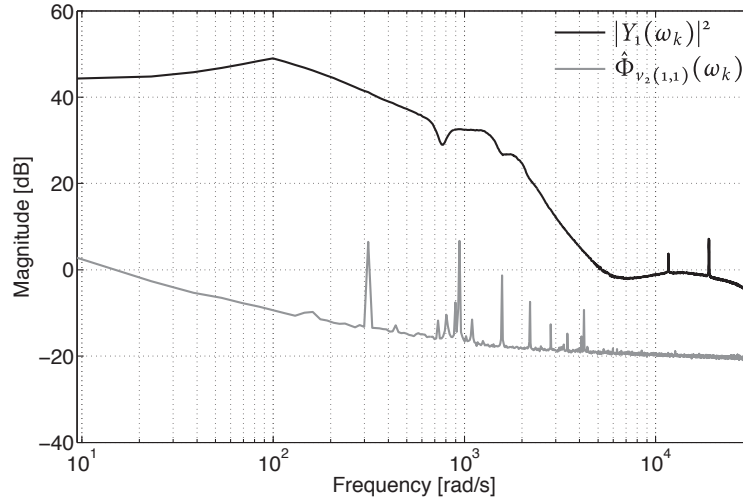
Figure 5.6: Comparison between the squared magnitude of the FRF of the position measurement of the X bearing (i.e. $|Y_1(\omega_k)|^2 [\mu m^2]$) and the estimated spectral density of the noise on the position measurements (i.e. $\hat{\Phi}_{v_2(1,1)}(\omega_k)$). The spectral density of the noise on the position measurement is somewhat concentrated at the lower frequencies.

elements of $\hat{\Phi}_{v_1}$ and $\hat{\Phi}_{v_2}$, as well as by the elements of $\hat{\Phi}_{v_1v_2}$ are small (except at the abovementioned disturbance frequencies), justifying the assumption that the noise in the different measurements is mutually uncorrelated.

We conclude by noting that the estimated spectra can be used to perform parametric noise modeling of the noise on the current and position measurements, which can be used in the cutting force estimation method of Chapter 3. For the EAAT spindle, the above results demonstrate the limited added value of that. Indeed, as discussed, the noise on the current measurements can be neglected. Furthermore, ignoring the mains disturbance, Figure 5.6 shows that the noise is not exactly white, but nonetheless the spectral density only decays slowly with increasing frequency: from 100 rad/s to 10 krad/s the drop is ±10 dB. Hence, it would be reasonable to model this disturbance as a white noise. This assumption implies that parametric noise modeling can reasonably be avoided in the case of the EAAT spindle. We remark that a white noise disturbance with a standard deviation of 0.1μm would correspond to a flat spectral density of -20 dB. The oscillatory distortions in

the measurements can be dealt with separately, e.g. by prefiltering[3].

5.6 Estimation of a parametric model of the bearing dynamics

In summary, the first stage of the identification procedure of the bearing dynamics has resulted in the following:

- A procedure has been described that using closed-loop AMB spindle data, yields a non-paramaric estimate of the FRF of the bearing dynamics with low bias and variance. In addition, an estimate of the covariance of the FRF estimate has been obtained. A method to quantifying the level of non-linear distortion in the FRF estimate has been given, allowing to tune the experimental conditions to reduce this distortion in the FRF estimate.

- A procedure has been given that with the closed-loop AMB spindle data allows to compute a non-parametric estimate of the noise spectra in the setup.

The objective of stage two is to estimate a parametric model, using the non-parametric estimates obtained in stage one. Here, the main emphasis is on identification of a parametric model of the plant dynamics, and not so much on obtaining a parametric model of the noise disturbances. This leads to the choice of minimization of an output error criterion. In Section 5.6.1 the resulting identification problem is defined more precisely. As will be shown, a Schur-weighted OE cost criterion is formulated, where an estimate of the covariance of the FRF is used for the weighting. The resulting identification problem fits the framework of Chapter 4, and the IV-based iterative linear regression approach is used to solve this. This algorithm suffers from poor numerical conditioning due to the large frequency range, and the large dynamic range for which the identification of the bearing dynamics is carried out. An approach to deal with this is is discussed in Section 5.6.2. In Section 5.6.3 the approach is illustrated using experimental results.

5.6.1 Parametric modeling by minimization of an weighted OE criterion

In this section it is discussed how a parametric model of the bearing dynamics of the AMB spindle can be obtained by a solving a weighted Output Error minimalization problem. To that end, assume estimates of the FRF $\{\hat{P}(\omega_k)\}_{k=1\ldots N_K}$ are available.

For the parametric modeling of the bearing dynamics, the polynomial matrix fraction description is adopted. As was discussed in Chapter 4, the MFD has a left

---

[3]Of course a better strategy would be to eliminate them altogether, but this could imply redesign of the device.

and a right form given by

$$P(\xi, \theta) = B(\xi, \theta)A^{-1}(\xi, \theta) \quad \text{(R-MFD)}$$
$$P(\xi, \theta) = A^{-1}(\xi, \theta)B(\xi, \theta) \quad \text{(L-MFD)}$$

with $B(\xi, \theta) = B_{n_b}\xi^{n_b} + B_{n_b-1}\xi^{n_b-1} + \cdots + B_0$ and $A(\xi, \theta) = \xi^{n_a} + A_{n_a-1}\xi^{n_a-1} + \cdots + A_0$, where $B_i \in \mathbb{R}^{p \times m}$, $i = 0 \ldots n_b$ and $A_i \in \mathbb{R}^{p \times m}$ (for models in R-MFD), or $A_i \in \mathbb{R}^{p \times m}$ (for models in L-MFD), for $i = 0 \ldots n_a - 1$. The parameter vector θ is constructed by accumulating all elements of the matrices A_i, $i = 0 \ldots n_a - 1$ and B_i, $i = 0 \ldots n_b$. As the input and output dimension of the bearing dynamics are identical (input dimension $m = 4$ and output dimension $p = 4$), there is no distinct reason to prefer one of the representations over the other [6] (see also Chapter 4).

The procedure to estimate the parameters θ given the data $\{\hat{P}(\omega_k)\}_{k=1\ldots N_K}$ is as follows. Under the assumptions that the estimation error in the FRF is circular complex Gaussian distributed with zero mean, and that the estimation error at frequency ω_k is independent of the estimation error at frequency ω_l for all $k \neq l$ [4], the maximum likelihood estimate of the parameters can be formulated as[5]

$$\hat{\theta}_{\text{ML}} = \arg \min_\theta \sum_{k=1}^{N_K} \text{vec}[E(\omega_k, \theta)]C_P^{-1}(\omega_k)\text{vec}[E(\omega_k, \theta)]^H \tag{5.56}$$

with

$$E(\omega_k, \theta) = \hat{P}(\omega_k) - P(e^{j\omega_k}, \theta) \tag{5.57}$$

and C_P the covariance matrix of $\hat{P}(\omega_k)$. Often information on the (non-parametric) covariance matrix C_P is not available, and identification schemes in which C_P is replaced by an estimate $\hat{C}_P$ have been proposed [147, 128, 146, 108]. With the experiment design described in Section 5.3.1, estimation of the covariance matrix of the JIO estimate of the FRF of the bearing dynamics can be carried out as described in Section 5.3.3. The performance of the estimator where C_P is replaced by an estimate $\hat{C}_P$ obtained with multisine excitation has been studied by Mahata et al. in [108]. It was shown that the resulting estimator is consistent when at least $p + 1$ periods of the excitation are used to estimate the covariance matrix. Furthermore, expressions of the asymptotic parameter covariance (i.e. for $N_K \to \infty$) are derived and these are shown to be proportional to the covariance of the maximum likelihood estimate. The proportionality constant δ depends on the number of periods K used to estimate the covariance matrix, with $\delta \to 1$ for $K \to \infty$.

---

[4]Note that the characterization of the uncertainty in $\{\hat{P}(\omega_k)\}_{k=1\ldots N_K}$ is here described in the frequency domain. Under mild conditions, these properties can be related to the noise disturbances on the time data that were used to estimate the FRF, see [128].

[5]See Ljung [105] for the SISO case, and Mahata et al. [108] for the MIMO case.

When the covariance matrix C_P is diagonal, a simplification of equation (5.56) can be made as follows. Define W as

$$W_{(i,j)}(\omega_k) = \left[C_{P(p(i-1)+j,\,p(i-1)+j)}(\omega_k) \right]^{-1/2} \tag{5.58}$$

that is, each element in $W(\omega_k)$ is the inverse of the standard deviation of the corresponding element in $\hat{P}(\omega_k)$. It is easily verified that the estimator

$$\hat{\theta} = \arg\min_\theta \sum_{k=1}^{N_K} \| E_s(\omega_k, \theta) \|_F^2 . \tag{5.59}$$

with

$$E_s(\omega_k, \theta) = W(\omega_k) . * \left[\hat{P}(\omega_k) - P(e^{j\omega_k}, \theta) \right] \tag{5.60}$$

is identical to the maximum likelihood estimator when the off-diagonal elements of C_P are zero. When C_P is not diagonal, but the off-diagonal elements are small, the loss of statistical efficiency of the estimator might well trade-off against the numerical advantages. Indeed, in equation (5.59), inversion of the $mp \times mp$ covariance matrix is eliminated and replaced by mp scalar inversions.

Observe that in equation (5.60) we have obtained a Schur-weighted Output Error, where an estimate of the variance of the elements of the estimated FRF is used for the weighting. The resulting nonlinear optimization problem in equation (5.59) can be solved using the IV-based linear regression approach that is discussed in Chapter 4.

The above described procedure in which the estimated variance is used as weighting is quite attractive for AMB spindle modeling. An important reason is that the variance of the estimated FRF can vary strongly over different frequencies. This occurs in particular when the identification is performed while the spindle is rotating. In that case periodic disturbances synchronous with the rotation of the spindle are experienced, resulting in estimation errors in the FRF estimate around the spindle frequency and its harmonics. The choice to weigh the output error with the inverse of the standard deviation of the FRF, allows to give less emphasis to FRF estimates that are less accurate and avoids a complicated procedure of making user-defined weighting functions (which would need to be different for each rotational speed then).

5.6.2 Multiband modeling

It is well known that the Sanathanan and Koerner iterations suffer from poor numerical conditioning when they are applied to high-order lightly damped systems with a dataset $\{\hat{P}(\omega_k)\}$ extending over a large frequency range. The IV-based linear regression approach detailed in Chapter 4 appears to exhibit problems with numerical condition in similar situations. This shortcoming of the Sanathanan and Koerner approach motivated Bayard in [22] to formulate a multi-band approach. The

key idea here is to divide the frequency range of interest into smaller subbands and to estimate submodels that describe the system within each band. When combined, these submodels produce a model for the entire system.

The approach for two subbands is as follows (generalization for more subbands is trivial). Define complementary weighting functions W_l and W_h by

$$W^l(\omega_k) = \begin{cases} 1 & \omega_k < \omega_B \\ 0 & \text{else} \end{cases} \tag{5.61}$$

with ω_B the frequency separating the two subbands, and let $W^h(\omega_k) = 1 - W^l(\omega_k)$. Define low-band and high-band parametric models $P^l(\xi, \theta_l)$ and $P^h(\xi, \theta_h)$, parametrized by the parameter vectors θ_l and θ_h respectively. With this, the Schur weighted OE is defined as

$$E_s(\omega_k, \theta_l, \theta_h) = W(\omega_k). * [\hat{P}(\omega_k) - P^l(e^{j\omega_k}, \theta_l) - P^h(e^{j\omega_k}, \theta_h)]. \tag{5.62}$$

The estimate of the low-band model is defined by

$$\hat{\theta}_l = \arg\min_\zeta \sum_{k=1}^{N_K} \left\| W^l(\omega_k) E_s(\omega_k, \zeta, \hat{\theta}_h) \right\|_F^2 \tag{5.63}$$

and the estimate of the high-band model by

$$\hat{\theta}_h = \arg\min_\zeta \sum_{k=1}^{N_K} \left\| W^h(\omega_k) E_s(\omega_k, \hat{\theta}_l, \zeta) \right\|_F^2. \tag{5.64}$$

The proposed procedure is to initialize $\theta_h = 0$, and then iterate between estimating the low-band model by equation (5.63) and estimating the high-band model by equation (5.64). Upon convergence, $\hat{P} = \hat{P}^l + \hat{P}^h$ gives the desired wide-band model.

We remark that for the Sanathanan-Koerer algorithm alternative solutions have been formulated to circumvent the mentioned conditioning problems. The key idea is to use an orthogonal polynomial basis for the numerator and denominator of the fractional model, leading to an (almost) optimal numerical conditioning of the normal equations in the linear regression steps. Solutions have been formulated for SISO systems [137, 128], as well as MIMO systems in MFD [126] (using vector orthogonal polynomials). An extension of such an approach for the IV-based linear regression method has not been developed yet.

5.6.3 Results

The approach in the previous sections has been applied for estimation of a parametric model of the EAAT spindle. The FRF and its covariance were estimated using the procedure described in Section 5.3 and 5.4, where the rotational speed is set to $80,000$ rpm. The FRF data, available on the frequency range from 7.6 rad/s to 27.6

5. SYSTEM IDENTIFICATION OF THE AMB SPINDLE, PART I: BEARING DYNAMICS
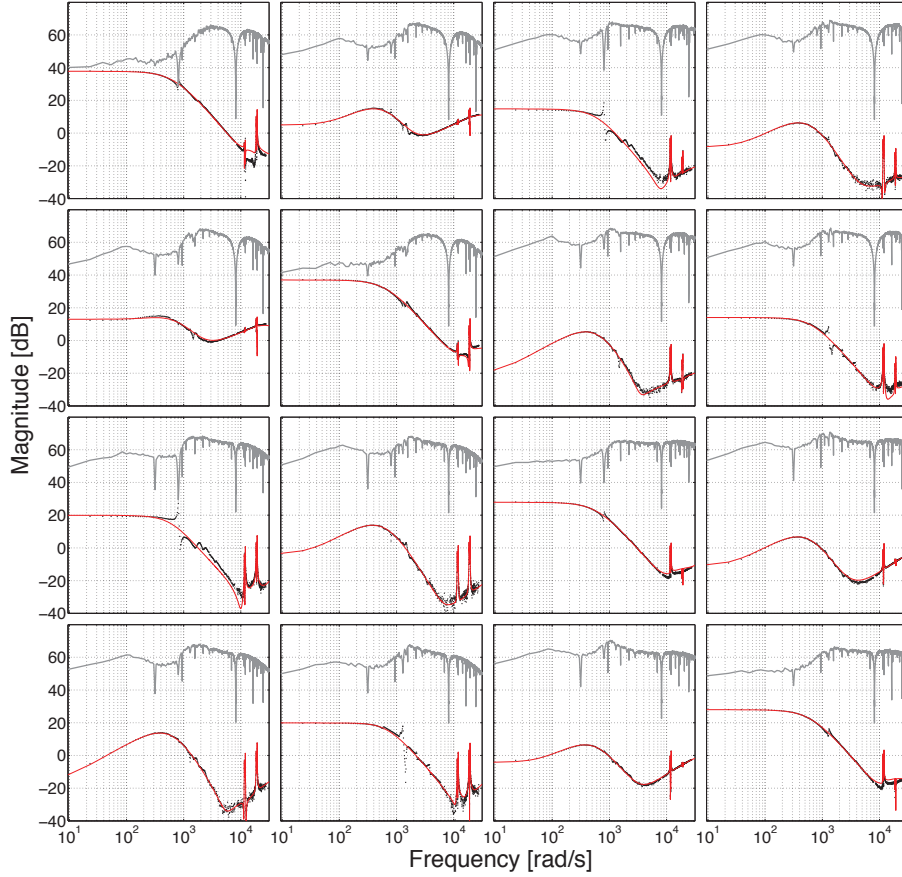
Figure 5.7: Estimation of a parametric model of the EAAT spindle: comparison of the FRF data $\hat{P}(\omega_k)$ (black dots) and the estimated parametric model $\hat{P} = P^l(\omega_k, \hat{\omega}_l) + P^l(\omega_k, \hat{\omega}_l)$ (red line). Weighting $W(\omega_k)$ is depicted in grey. The parametric model $\hat{P}$ closely fits the estimated FRF $\hat{P}(\omega_k)$ in the regions where the weighting W is large.

krad/s was divided into two subbands, with $\omega_B = 7.7$ krad/s[6]. The lower band held 250 FRF data points, the higher band 458 data points. Points with a high variance (and thus very low weighting) were excluded a priori to make the procedure more efficient.

Without multi-band modeling, and model orders with $n_a \geq 4$ and $n_b \geq 3$, the

---

[6]Observe that with this separation the low band essentially captures the rigid body behavior, while the high band model models the flexural behavior of the AMB spindle

Figure 5.8: Estimation of a parametric model of the EAAT spindle: comparison of element (1,1) of the FRF data $\hat{P}(\omega_k)$ (black dots) and element (1,1) of the estimated parametric model $\hat{P} = P^l(\omega_k, \hat{\theta}_l) + P^h(\omega_k, \hat{\theta}_h)$ (red line). Weighting $W(\omega_k)$ is depicted in grey.

IV-based iterative linear regression diverges due to poor numerical conditioning. With multi-band modeling, these issues can be avoided. Results are shown for the case that a model set in R-MFD is selected for both the lower band and higer band submodel, where for the former the orders are $n_a = 3$ and $n_b = 2$ (64 parameters), and the latter $n_a = 5$ and $n_b = 4$ (128 parameters). The IV-based linear regression method decribed in Chapter 4 is applied to solve (5.63) and (5.64). This results in a nested iteration: the multi-band technique involves iteratively solving an optimization problem for the lower and upperband, while each optimization step is solved iteratively using the IV-based linear regression method. Hence, the following algorithm is obtained:

Initiate: $\hat{\theta}_{l;0} = 0$, $\hat{\theta}_{h;0} = 0$;

Step m, lower band Solve (5.63), with $\theta_h = \hat{\theta}_{h;m}$, use $\hat{\theta}_{l;m}$ to initiate IV-based iterative linear regression. Upon convergence of iterations, this yields $\hat{\theta}_{l;m+1}$;

Step m, higher band Solve (5.64), with $\theta_l = \hat{\theta}_{l;m+1}$, use $\hat{\theta}_{h;m}$ to initiate IV-based iterative linear regression. Upon convergence of iterations, yields $\theta_{h;m+1}$;

Continue to step $m + 1$, until convergence.

In step $m = 1$, convergence of the IV-based iterative linear regression was obtained after 3 (lower band) and 5 steps (higher band). With these first estimates of $\hat{\theta}_l$ and $\hat{\theta}_h$, the iterations are continued in step $m = 2$ and $m = 3$. After $m = 4$, no further significant reduction of the overall output error was obtained.

A comparison between the FRF data and the estimated parametric model is depicted in Figure 5.7. The frequency reponse of the estimated model shows very high correspondence to the dataset. The dynamics including the ill-damped resonances are estimated correctly. To see this, a zoomed-in plot of element (1,1) of the amplitude response function matrix is given in Figure 5.8. As can be seen in this figure, the parametric model well captures the bending mode vibrations, which appear as twin of close resonances due to the gyroscopic effects. Moreover, the applied weighting (also plot in Figures 5.7 and 5.8) effectively avoids modeling errors due to large variance errors in the FRF data. Here such variance errors are particularly present around the harmonics of the rotational frequency of the spindle ($80,000$ rpm = 8.4 krad/s). Similarly, in the frequency range from $800 - 1500$ rad/s some additional modes seem to be present. It was verified that these are related to the fixation of the spindle to the Z-stage in the milling machine[7]. The standard deviation of the FRF data at these modes is up to 10 times larger than at other frequencies, and consequently the data is weighted less while estimating a model of the spindle. As a result, these modes are not included in the estimated model.

To further assess the fit of the estimated model to the FRF data, a comparison is made between the singular values of the FRF data and the estimated model. The result is depicted in Figure 5.9. Also this figure shows the high correspondence between the parametric model and the FRF data. Furthermore, to validate that the estimated model is correctly representing the actual dynamics, is verified that the estimated model was stabilized by the existing controller. This is indeed the case.

As a last verification, a comparison of the obtained results with a first principles modeling was made. For this, we recall from Chapter 2 that the EAAT spindle satisfies the therein described configuration I with direct measurement of the inputs. For this configuration, under the given conditions, the FRF of P represents the FRF of the (true) continuous-time AMB system. Hence, an additional identification was performed with a continuous-time model set having the same properties. The resulting estimated model is compared to a model derived from first principles according to the procedure discussed in Chapter 2 (which is also in continuous-time). In Figure 5.10 the poles of the estimated model are compared to the analytic model of the EAAT spindle. This figure demonstrates that the pole locations of the estimated model are close to the pole locations of the model derived from first principles. This is a favorable result, as it demonstrates that both from first principles, as well as using the system identification approach in this chapter, essentially the same dynamical behavior is modeled.

---

[7]In another test the spindle was rotated 90 degrees around its Z-axis. As a result the modes shifted from the X-plane to the Y-plane, and vice versa.
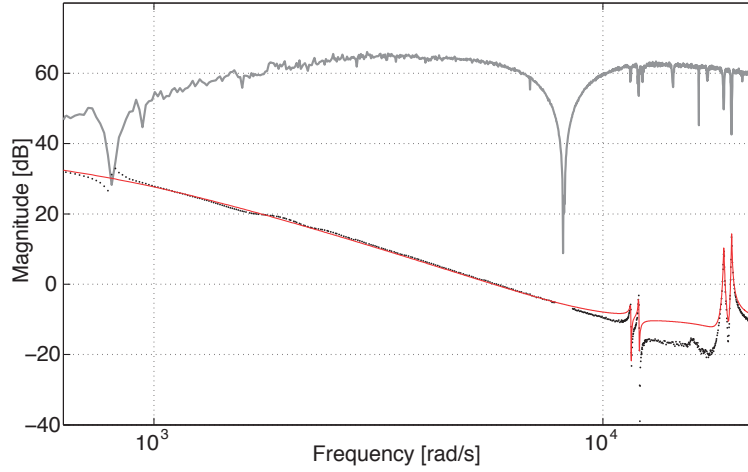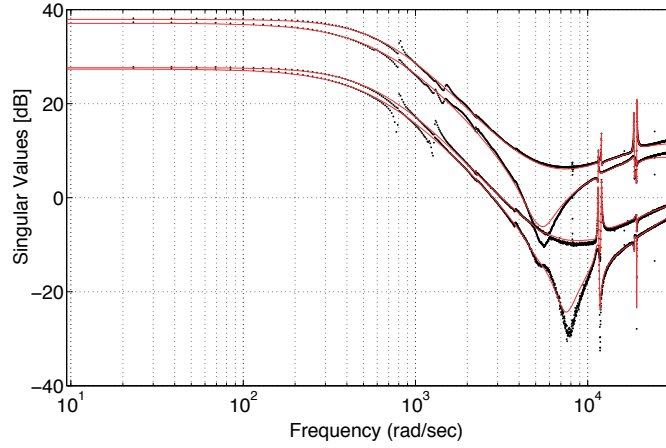
Figure 5.9: Estimation of a parametric model of the EAAT spindle: comparison of the sigular values of the FRF data $\hat{P}(\omega_k)$ (black dots) and of the the estimated parametric model $\hat{P} = P^l(\omega_k, \hat{\omega}_l) + P^h(\omega_k, \hat{\omega}_l)$ (red line). The estimated model captures well the flexural modes of the spindle at high rotational speeds

5.7 Summary and conclusions

In this chapter we have considered the problem of obtaining accurate models of the bearing dynamics of the AMB spindle in the frequency range relevant for process monitoring and control. The complexity of this problem is found in the instability of the dynamics, its high order and multivariable character, the parametrically dependency of the dynamics on the rotational speed, and in the presence of nonlinearities in the system. Current results in literature reduce the complexity of the problem by ignoring one or more of these issues. In order to obtain accurate models that can be used for process monitoring and control, in particular for force estimation, an approach has been given that addresses the identification problem in full.

A frequency domain approach has been taken, consisting of two stages. In the first stage, accurate estimates of the multivariable FRF of the bearing dynamics are made by excitation of the closed-loop with orthogonal random phase multisine signals and application of the JIO FRF estimator. In addition, an approach has been given to estimate the covariance in this estimate. Furthermore, a procedure has been given to detect the level of nonlinear distortion in the FRF estimate of the AMB spindle. This approach is based on the variance analysis method for detection of nonlinearities in FRF estimates as available in literature, and is extended to detection of nonlinear distortion in MIMO FRF estimates obtained using the JIO estimator. A strategy has been formulated to reduce the nonlinear distortion in the JIO FRF estimate, and it has been shown that this results in a reduction of the non-

5. SYSTEM IDENTIFICATION OF THE AMB SPINDLE, PART I: BEARING DYNAMICS
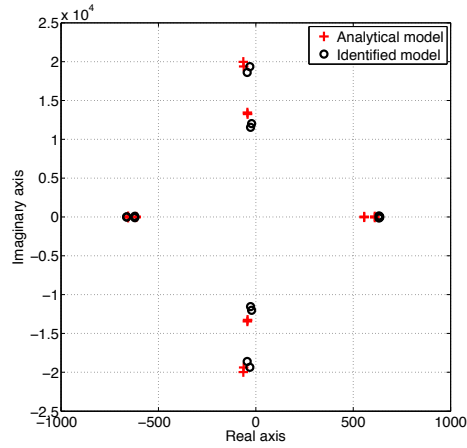
Figure 5.10: Estimation of a parametric model of the EAAT spindle: comparison of the poles of the estimated continuous-time model and the poles of an analytic model of the EAAT spindle at 80,000 rpm.

linear distortion of up to 25 dB. The key here is to increase the excitation amplitude in the lower frequency region, causing a better signal to nonlinear disturbance ratio.

Using the same experiment design, non-parametric noise modeling is performed by variance analysis, demonstrating for the experimental AMB setup that the disturbance on the current measurement is negligible. In addition, this test revealed some imperfections of the design of the experimental AMB spindle setup, in the form of strong oscillations present in the signals.

In the second stage, a parametric model of the AMB spindle is estimated using the estimated FRF and its covariance. A Schur-weighted Output Error criterion has been formulated, which has been shown to approximate the maximum likelihood cost criterion under mild assumptions. Minimization of the criterion function is achieved by application of the IV-based iterative linear regression method of Chapter 4. Multi-band modeling is performed to avoid the numerical condition problems experienced with the iterative linear regression method when dealing with a FRF data set extending of a large frequency range and dynamic range. Experimental results show the high correspondence of the estimated parametric model to the estimated FRF data. It is verified that the poles of the identified model and a model derived from first principles are closely together, illustrating the ability of the given identification approach to capture the dynamics as would be predicted from physical insights.

132

Chapter 6

System identification of the AMB spindle, part II: Tooltip dynamics

6.1 Introduction

In the previous chapter, identification of the bearing dynamics of the AMB spindle was covered. In this chapter we will address the problem of identifying the tooltip dynamics, i.e. the dynamics from the force on the tooltip to the position of the rotor at the location of the displacement sensors. In Figure 2.11 in Chapter 2, the tooltip dynamics are represented in the transfer function from u_2 to y. A principal challenge in the identification of the tooltip dynamics is the generation of data. This is in great contrast with the bearing dynamics identification problem, where excitation of the system can easily be realized by adding signals to the closed-loop, and collection of large amounts of data does not impose significant challenges. An essential prerequisite to identifying the tooltip dynamics from measured data, is that the tooltip is excited with a known (c.q. measured) force. The fragility of the tooltip of a micro-end mill rules out many standard contact excitation techniques that make use of shakers or impact hammers, as these methods would unavoidably result in catastrophic tool breakage. A more cautious approach could be to apply step relaxation methods, where a static preload is applied to the tooltip which is released instantaneously. However, the excitation power that can be applied in this way is very limited, making this approach less attractive. Moreover, such a technique would only be useful at standstill, where we are interested in identifying the dynamics at high rotational speeds. Contactless methods using electromagnetic excitation [3], laser excitation (thermal excitation) [29], or acoustic excitation[12] have as common drawbacks that the excitation force cannot be measured and that the excitation level is limited.

In a way to circumvent the problem of exciting small and fragile tools, a sub-

133

structure coupling approach known as *receptance coupling* [136, 144] could be applied. By definition, the receptance of a mechanical system is the transfer function from force to displacement. Receptance coupling involves the procedure of determining the receptance of a complex mechanical system from models of its subsystems by appropriate modeling of the joints. The benefit of this technique is that once models of substructures and the joint are obtained, a model of the assembled structure can be calculated without any further experiments. Park et al. applied this technique to model the tooltip dynamics for a spindle with conventional bearings [123, 121]. Here separate models are obtained for the tool and the spindle, where analytic FE models for tool are used and the spindle dynamics are identified by adding a short cylindrical insert and applying excitation forces to that. The disadvantage of this approach lies in the difficulty to identify the coupling and the limited accuracy of the FE models of the tool.

In all of the above methods, data generation happens by applying a user-defined excitation to the system. In this thesis, we follow a different route and identify the tooltip dynamics in-process. This means that for the purpose of identifying the tooltip dynamics, we add a force sensor to the machine and during a milling experiment measure the cutting forces, as well as the signals of the AMB. From the resulting data set, the objective is to estimate the tooltip dynamics. The clear advantage of this approach is that the data generation for identification is performed under conditions very similar to the conditions under which the resulting model is to be used. However, these experimental conditions also impose two constraints. First of all, the amount of data that can be obtained in this fashion is limited. Second, since the milling process itself is the source of the excitation, we have no direct influence on the spectral content of the excitation signal. In an attempt to deal with these restrictions, we apply prior knowledge of the system, allowing us to significantly reduce the complexity of the identification problem. This is done as follows. The key step is to make use of the assumed property of the system discussed in Chapter 2, stating that within the frequency range of interest all observable modes are controllable through the control input of the bearings. For convenience, we repeat the consequence of this that a minimal state-space representation for P exists, given by

$$x(t+1) = A_P x(t) + B_{P,1} u_1(t) + B_{P,2} u_2(t) \qquad (6.1a)$$

$$y_0(t) = C_P x(t) \qquad (6.1b)$$

with $(A_P, B_{P,1})$ controllable, and (C_P, A_P) observable. The bearing dynamics are described by the matrices A_P, $B_{B,1}$ and C_P, the tooltip dynamics by the matrices A_P, $B_{B,2}$ and C_P. Hence, once having identified the bearing dynamics, the only remaining matrix to identify in the tooltip dynamics is matrix $B_{P,2}$. Therefore, the problem addressed in this chapter is the identification of this matrix from data obtained during a milling experiment, for given A_P, $B_{B,1}$ and C_P.

In Section 6.2 it is discussed how this problem can be formulated as a prediction
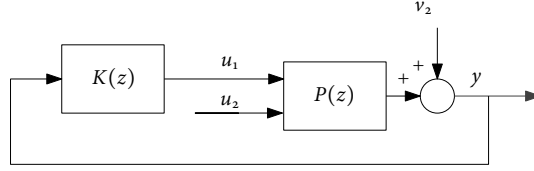
Figure 6.1: Configuration of the tooltip identification problem

error identification problem with a specific model parametrization following from the closed-loop configuration, and a fixed noise model that is non-minimum phase. The solution to this problem can be obtained via linear regression, as is presented in Section 6.3. The LS estimator becomes computationally unattractive for growing data sets, and we will extend it to a recursive estimator in Section 6.4. The methods are applied to a realistic simulation model of the the EAAT spindle and results are given in Section 6.5. Last, we give a brief discussion of the method in Section 6.6.

6.2   Configuration and identification problem

The configuration that we consider is depicted in Figure 6.1. The stable closed loop consists of a plant P and controller K. Let plant P assume the state space realization as given in equation (6.1), where $A_P \in \mathbb{R}^{n \times n}$, $B_{P,1} \in \mathbb{R}^{n \times m_1}$, $B_{P,2} \in \mathbb{R}^{n \times m_2}$ and $C_P \in \mathbb{R}^{p \times n}$. Plant P is unstable, meaning that A_P has unstable poles. Without loss of generality, we will assume $B_{P,1}$ and C_P to have full rank. Furthermore, the output of the plant P is disturbed with noise[1] v: $y(t) = y_0(t) + v(t)$. It is at first assumed this noise is white, with $\mathbb{E} v v^T = \sigma_v^2 I$. Finally, let the controller have the minimal state space realization

$$x_K(t+1) = A_K x_K(t) + B_K y(t) \tag{6.2a}$$
$$u_{1,0}(t) = C_K x_K(t) + D_K y(t) \tag{6.2b}$$

where $A_K \in \mathbb{R}^{n_K \times n_K}$, $B_K \in \mathbb{R}^{n_K \times p}$, $C_K \in \mathbb{R}^{m_1 \times n_K}$ and $D_K \in \mathbb{R}^{m_1 \times p}$. Consider A_P, $B_{P,1}$, C_P, A_K, B_K, C_K and D_K are known. It is easily verified that the closed loop

---

[1] For notational efficiency we have dropped the subscript '2' and write v instead of v_2 in this chapter.

admits the state space representation

$$
\begin{pmatrix} x(t+1) \\ x_K(t+1) \\ \hline y(t) \end{pmatrix} = \left[\begin{array}{cc|cc} A_P + B_{P,1}D_KC_P & B_{P,1}C_K & B_{P,2} & B_{P,1}D_K \\ B_KC_P & A_K & 0 & B_K \\ \hline C_P & 0 & 0 & I \end{array} \right] \begin{pmatrix} x(t) \\ x_K(t) \\ u_2(t) \\ v(t) \end{pmatrix}
$$

$$
= \left[\begin{array}{c|cc} \mathcal{A} & \mathcal{B}_1 & \mathcal{B}_2 \\ \hline \mathcal{C} & \mathcal{D}_1 & \mathcal{D}_2 \end{array} \right] \begin{pmatrix} x(t) \\ x_K(t) \\ u_2(t) \\ v(t) \end{pmatrix}.
$$

$$(6.3)$$

With this, the data-generating system is assumed to obey

$$
y(t) = T_0(q)u_2(t) + S_0(q)v(t) \tag{6.4}
$$

with transfer functions[2] $T_0(q) = \mathcal{C}(qI - \mathcal{A})^{-1}\mathcal{B}_1$ and $S_0(q) = \mathcal{C}(qI - \mathcal{A})^{-1}\mathcal{B}_2 + I$. Observe that $S_0(q)$ is known and that $T_0(q)$ is linear in the unknown matrix $B_{P,2}$. Identification of this matrix can be formulated in the prediction error identification framework of Ljung [105] as follows.

First note that $S_0(q)$ is non-minimum phase, since A_P is unstable. This is a well-known result, which is easily verified: for $S_0^{-1}(q)$ we can write

$$
S_0^{-1}(q) = I - \mathcal{C}(qI - \mathcal{A} + \mathcal{B}_2\mathcal{C})^{-1}\mathcal{B}_2 \tag{6.5}
$$

and

$$
\mathcal{A} - \mathcal{B}_2\mathcal{C} = \left[\begin{array}{cc} A_P & B_{P,1}C_K \\ 0 & A_K \end{array} \right].
$$

From this, we immediately observe that the eigenvalues of $S_0^{-1}(q)$ are the union of the eigenvalues of the plant P and the controller C. In order to define a stable predictor, we define the factorization

$$
S_0(q)S_0(q)^H = M(q)M(q)^H \tag{6.6}
$$

with $M(q)$ minimum phase (and causal). Using the state space representation of $S(q)$, this factorization can be computed (see proposition 3.5.2 in Chapter 3). The data-generating system can now equivalently be represented as

$$
y(t) = T_0(q)u_2(t) + M(q)v_M(t) \tag{6.7}
$$

---

[2] Again, note that we dropped the subscript yu_2 here and write T instead of T_{yu_2}

with $v_M(t)$ a white noise process with the same variance as $v(t)$. A predictor model is composed as the transfer function $T(q, \theta)$ and $M(q)$, yielding the one-step predictor

$$\hat{y}_M(t|t-1;\theta) = [M^{-1}(q) - I]y(t) - M^{-1}(q)T(q, \theta)u_2(t) \tag{6.8}$$

yielding prediction error

$$\epsilon_M(t, \theta) = M^{-1}(q)[y(t) - T(q, \theta)u_2(t)]. \tag{6.9}$$

Specific in this problem is the particular parametrization of the model $T(q, \theta)$ according to the above closed-loop state space structure, with parameter vector θ representing the unknown elements of the matrix $B_{P,2}$. This vector is estimated on the basis of input-output data $Z^N := \{u_2(t), y(t)\}_{t=1\ldots N}$, according to a quadratic cost criterion

$$V_N(\theta, Z^N) = \frac{1}{N}\sum_{t=1}^{N}\epsilon_M(t, \theta)^T \epsilon_M(t, \theta). \tag{6.10}$$

by

$$\hat{\theta}_N = \arg\min_{\theta} V_N(\theta, Z^N). \tag{6.11}$$

Since $\epsilon_M(t, \theta)$ is affine in the parameters, this identification problem can be solved via linear regression, as will be done in the next section.

6.3 Estimation of $B_{P,2}$ by linear regression

In this section we will derive a solution for the identification problem formulated previously. First a linear regression approach to this problem is given, followed by a bias and variance analysis for the resulting estimate $\hat{B}_{P,2}$, as well as of the estimated frequency reponse $\hat{P}_2(e^{j\omega})$. The section is concluded by extending the approach for correlated output measurements. Let us introduce the notation

$$Y = \begin{bmatrix} y(1) \\ \vdots \\ y(N) \end{bmatrix}, \quad U = \begin{bmatrix} u_2(1) \\ \vdots \\ u_2(N) \end{bmatrix}, \quad V = \begin{bmatrix} v_M(1) \\ \vdots \\ v_M(N) \end{bmatrix}. \tag{6.12}$$

From the state space representation of the closed loop, it can be derived that for the general case with nonzero initial condition, the data generating system obeys

$$Y = \mathcal{O}\bar{x}(1) + \Gamma_1 U + \Gamma_M V \tag{6.13}$$

with $\bar{x} = \text{col}(x, x_K)$, and

$$\mathcal{O} = \begin{bmatrix} \mathcal{C} \\ \mathcal{C}\mathcal{A} \\ \mathcal{C}\mathcal{A}^2 \\ \vdots \\ \mathcal{C}\mathcal{A}^{N-1} \end{bmatrix}, \quad \Gamma_1 = \begin{bmatrix} \mathcal{D}_1 & 0 & 0 & \cdots & 0 \\ \mathcal{C}\mathcal{B}_1 & \mathcal{D}_1 & 0 & \cdots & 0 \\ \mathcal{C}\mathcal{A}\mathcal{B}_1 & \mathcal{C}\mathcal{B}_1 & \mathcal{D}_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \mathcal{C}\mathcal{A}^{N-2}\mathcal{B}_1 & \mathcal{C}\mathcal{A}^{N-3}\mathcal{B}_1 & \cdots & \mathcal{C}\mathcal{B}_1 & \mathcal{D}_1 \end{bmatrix}. \quad (6.14)$$

Matrix Γ_M is defined as follows. Let $(\mathcal{A}_M, \mathcal{B}_M, \mathcal{C}_M, \mathcal{D}_M)$ be a minimal realization for $M(q)$ [3]. Matrix Γ_M holds the same structure as Γ_1, but then evaluated for this realization.

Observe that expression (6.13) is linear in the unknown initial state $\bar{x}(1)$ and unknown matrix $B_{P,2}$. In appendix B.3.1 it is shown that we can express $V_N(\theta, Z^N)$ as

$$V_N(\theta, Z^N) = |\Gamma_M^{-1}(Y - \Psi\theta)|^2 \quad (6.15)$$

with parameter vector θ the vector of all unknowns (including the initial state)

$$\theta = \begin{bmatrix} \bar{x}(1) \\ \text{vec}(B_{P,2}) \end{bmatrix} \quad (6.16)$$

and regression matrix Ψ defined as

$$\Psi = \begin{bmatrix} \mathcal{O} & \mathcal{U}_1\mathcal{O}_P & \mathcal{U}_2\mathcal{O}_P & \cdots & \mathcal{U}_{m_2}\mathcal{O}_P \end{bmatrix} \quad (6.17)$$

where $\mathcal{O}_P$ are the first n columns of $\mathcal{O}$, i.e.

$$\mathcal{O} = \begin{bmatrix} \mathcal{O}_P & \mathcal{O}_C \end{bmatrix}, \quad (6.18)$$

and

$$\mathcal{U}_i = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ u_{2,i}(1) & 0 & 0 & \cdots & 0 \\ u_{2,i}(2) & u_{2,i}(1) & 0 & \cdots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ u_{2,i}(N-1) & u_{2,i}(N-2) & \cdots & u_{2,i}(1) & 0 \end{bmatrix} \otimes I.$$

Here we have used $u_{2,i}(t)$ to indicate element number i of the input vector u_2 at time instant t. Minimization of (6.15) results in a standard linear regression problem, having solution

$$\hat{\theta} = (\Psi_\Gamma^T \Psi_\Gamma)^{-1} \Psi_\Gamma^T Y_\Gamma \quad (6.19)$$

---

[3] Observe that if the spectral factorization approach in Section 6.3 is followed, a realization for $M(q)$ will be obtained satisfying $\mathcal{A}_M = \mathcal{A}$, $\mathcal{B}_M = LR^{1/2}$, $\mathcal{C}_M = \mathcal{C}$, and $\mathcal{D}_M = R^{1/2}$, with L and R as defined in Proposition 3.5.2.

with $\Psi_\Gamma = \Gamma_M^{-1}\Psi$ and $Y_\Gamma = \Gamma_M^{-1}Y$. Observe that that in equation (6.19), regression matrix Ψ_Γ consists of values of the input signal filtered with a known filter, like in a classical FIR model.

In order to study the bias and covariance of the estimator, let us define the estimation error $\tilde{\theta} = \hat{\theta} - \theta_0$, with θ_0 representing the true parameter values. The estimate of θ in (6.19) is unbiased, i.e. $\mathbb{E}\tilde{\theta} = 0$, following directly from the whiteness of v_M. The covariance of the estimate is given by

$$\Pi = \mathbb{E}(\hat{\theta} - \mathbb{E}(\hat{\theta}))(\hat{\theta} - \mathbb{E}(\hat{\theta}))^T \tag{6.20}$$

$$= \mathbb{E}\tilde{\theta}\tilde{\theta}^T \tag{6.21}$$

$$= (\Psi_\Gamma^T\Psi_\Gamma)^{-1}\sigma_v^2. \tag{6.22}$$

With these expressions for the bias and covariance of $\hat{\theta}$, we can also compute the bias and covariance of $P_2(e^{j\omega})$. For that purpose define

$$\tilde{P}_2(e^{j\omega}) = \hat{P}_2(e^{j\omega}) - P_2(e^{j\omega})$$

$$= C_P(e^{j\omega}I - A_P)^{-1}(\hat{B}_{P,2} - B_{P,2}).$$

From the unbiasedness of $\hat{\theta}$ it immediately follows that $\mathbb{E}\tilde{P}_2(e^{j\omega}) = 0$. In order to express the covariance of $\hat{P}_2(e^{j\omega})$, partition covariance matrix Π according to the dimensions of $x(1)$ and $\text{vec}(B_{B,2})$:

$$\Pi = \left[\begin{array}{cc} \Pi_{11} & \Pi_{12} \\ \Pi_{21} & \Pi_{22} \end{array} \right]$$

Then the covariance of $\hat{P}_2(e^{j\omega})$ can be expressed as

$$\text{COV}(\hat{P}_2(e^{j\omega})) = \mathbb{E}\text{vec}(\tilde{P}_2(e^{j\omega}))\text{vec}(\tilde{P}_2(e^{j\omega}))^H \tag{6.23}$$

$$= [I \otimes C_P(e^{j\omega}I - A_P)^{-1}]\Pi_{11}[I \otimes C_P(e^{j\omega}I - A_P)^{-1}]^H \tag{6.24}$$

where it is used that $\text{vec}(\tilde{P}_2(e^{j\omega})) = [I \otimes C_P(e^{j\omega}I - A_P)^{-1}]\text{vec}(\hat{B}_{P,2} - B_{P,2})$.

Consider now the case that the output measurements are not white. In that case the data-generating system is assumed to obey

$$y(t) = T_0(q)u_2(t) + S_0(q)v(t) \tag{6.25}$$

with $v(t) = H_0(q)e(t)$ with $H_0(q)$ a known proper monic transfer function, which as usual is assumed stable and minimum phase, and $e(t)$ a white noise process with covariance $\mathbb{E}ee^T = \sigma_e^2I$. Using the same notation as in equation (6.12), we can now express

$$Y = \mathcal{O}\bar{x}(1) + \Gamma_1 U + \Gamma_M\mathcal{H}V \tag{6.26}$$

139

with $\mathcal{H}$ the lower block Toeplitz matrix with Markov parameters of $H(q)$. With this, the identification procedure is solved analogously the case with white noise output measurement disturbance, yielding LS estimate

$$\hat{\theta} = (\Psi_\Gamma^T \Psi_\Gamma)^{-1} \Psi_\Gamma^T Y_\Gamma \tag{6.27}$$

with $\Psi_\Gamma = (\Gamma_M \mathcal{H})^{-1} \Psi$ and $Y_\Gamma = (\Gamma_M \mathcal{H})^{-1} Y$.

6.4 Formulation of a recursive estimation scheme

For larger data sets, good algorithms are needed to limit the computation time. Here an efficient recursive estimation approach is given. Let us introduce the following notation:

$$Y(k) = \begin{bmatrix} y(1+kN) \\ \vdots \\ y(N+kN) \end{bmatrix}, \quad V(k) = \begin{bmatrix} v_M(1+kN) \\ \vdots \\ v_M(N+kN) \end{bmatrix}, \quad \theta(k) = \begin{bmatrix} \bar{x}(1+kN) \\ \text{vec}(B_{P,2}) \end{bmatrix} \tag{6.28}$$

Using the state space representation of (6.3) we can derive that

$$\theta(k+1) = \phi(k)\theta(k) + \Gamma_B V(k) \tag{6.29}$$
$$Y(k) = \Psi(k)\theta(k) + \Gamma_M V(k) \tag{6.30}$$

with

$$\phi(k) = \begin{bmatrix} \mathcal{A}^N & H(k) \\ 0 & I \end{bmatrix}$$

$$H(k) = \left[u^T(1+kN) \otimes \mathcal{A}^{N-1}\begin{pmatrix} I \\ 0 \end{pmatrix} \right] + \left[u^T(2+kN) \otimes \mathcal{A}^{N-2}\begin{pmatrix} I \\ 0 \end{pmatrix} \right] + \cdots + $$
$$+ \left[u^T(N+kN) \otimes I\begin{pmatrix} I \\ 0 \end{pmatrix} \right]$$

$$\Gamma_B = \begin{bmatrix} \mathcal{A}^{N-1}\mathcal{B}_2 & \mathcal{A}^{N-2}\mathcal{B}_2 & \dots & \mathcal{B}_2 \end{bmatrix}$$

$$\Psi(k) = \begin{bmatrix} \mathcal{O} & \mathcal{U}_1(k)\mathcal{O}_P & \mathcal{U}_2(k)\mathcal{O}_P & \cdots & \mathcal{U}_{m_2}(k)\mathcal{O}_P \end{bmatrix}$$

$$\mathcal{U}_i(k) = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ u_{2,i}(1+kN) & 0 & 0 & \cdots & 0 \\ u_{2,i}(2+kN) & u_{2,i}(1+kN) & 0 & \cdots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ u_{2,i}(N-1+kN) & u_{2,i}(N-2+kN) & \cdots & u_{2,i}(1+kN) & 0 \end{bmatrix} \otimes I$$

and $\mathcal{O}$, $\mathcal{O}_P$ and Γ_2 as defined before. Observe that $V(k)$ appears in both the state update and measurement equation. We can eliminate this correlation in the state

and measurement noise, and bring the system in the standard form

$$\theta(k+1) = \phi_s(k)\theta(k) + \Gamma_B Y_\Gamma(k) \tag{6.31}$$
$$Y_\Gamma(k) = \Psi_\Gamma(k)\theta(k) + V(k) \tag{6.32}$$

with $\phi_s(k) = \phi(k) - \Gamma_B\Psi_\Gamma(k)$, $Y_\Gamma(k) = \Gamma_M^{-1}Y(k)$, and $\Psi_\Gamma(k) = \Gamma_M^{-1}\Psi(k)$. We use the well-known measurement and time update equations for the time-variant Kalman filter to obtain a recursive estimator for $B_{P,2}$ [83]:

- $k = 0$: Here we write the measurement update in information form

$$\Pi(0|0) = \left(\Pi(0|-1)^{-1} + \frac{1}{\sigma_v^2}\Psi_\Gamma(0)\Psi_\Gamma^T(0) \right)^{-1} \tag{6.33}$$

$$K(0) = \Pi(0|0)\frac{1}{\sigma_v^2}\Psi_\Gamma^T(0) \tag{6.34}$$

$$\hat{\theta}(0|0) = \theta(0|-1) + K(0)\left[Y(0) - \Psi(0)\hat{\theta}(0|-1)\right] \tag{6.35}$$

so that when $\hat{\theta}(0|-1) = 0$, $\Pi(0|-1) \to \infty$, these equations reduce to

$$K(0) = \left(\Psi_\Gamma(0)\Psi_\Gamma^T(0)\right)^{-1}\Psi_\Gamma^T(0) \tag{6.36}$$
$$\hat{\theta}(0|0) = K(0)Y_\Gamma(k) \tag{6.37}$$
$$\Pi(0|0) = \sigma_v^2\left(\Psi_\Gamma(0)\Psi_\Gamma^T(0)\right)^{-1} \tag{6.38}$$

in which we recognize the estimator of the previous section.

- $k = 0$: The time update is given by

$$\hat{\theta}(1|0) = \phi_s(0)\theta(0|0) + \Gamma_B Y_\Gamma(0) \tag{6.39}$$
$$\Pi(1|0) = \phi_s(0)\Pi(0|0)\phi_s^T(0) \tag{6.40}$$

- $k > 0$: Here we write the measurement update in innovation form

$$K(k) = \Pi(k|k-1)\Psi_\Gamma^T(k)\left[\Psi_\Gamma(k)\Pi(k|k-1)\Psi_\Gamma^T(k) + \sigma_v^2 I\right]^{-1} \tag{6.41}$$
$$\hat{\theta}(k|k) = \hat{\theta}(k|k-1) + K(k)\left[Y_\Gamma(k) - \Psi_\Gamma(k)\hat{\theta}(k|k-1)\right] \tag{6.42}$$
$$\Pi(k|k) = \left[I - K(k)\Psi_\Gamma(k)\right]\Pi(k|k-1) \tag{6.43}$$

- $k > 0$ The time update is given by

$$\hat{\theta}(k+1|k) = \phi_s(k)\theta(k|k) + \Gamma_B Y_\Gamma(k) \tag{6.44}$$
$$\Pi(k+1|k) = \phi_s(k)\Pi(k|k)\phi_s^T(k) \tag{6.45}$$

With this recursion an computational efficient algorithm for identification of $B_{P,2}$ has been obtained.

6.5   Results

By lack of an experimental setup able to produce reliable data, validation of the method in this chapter using experimental data was not possible. Instead, the approach has been tested in simulation in Matlab/Simulink, and applied to a realistic model of the EAAT spindle described in Chapter 2. Using the described FE modeling approach, followed by balanced truncation [118] and discretization, a state space model of the spindle has been obtained with $n = 16$, meaning that the model captured the rigid body behavior, as well as the first two flexural modes. The AMB spindle was stabilized by four decentralized PID controllers. The noise disturbance on the position measurements was white with $\sigma_v = 0.1\mu$m. As in Chapter 3, the input consisted of a simulated cutting force signal, based on a cutting force model in the literature for a micro-end mill with two teeth [45]. A rotational speed of 30 krpm was selected. With this cutting force signal and the resulting model of the spindle, a cutting experiment with the AMB spindle (i.e. the data-generating system) has been simulated.

With this simulation data the identification procedure was tested. Using the simulated cutting force signal u_2 and the simulated displacement signal y, matrix $B_{P,2}$ has been estimated according to the procedure set out in this chapter. This procedure requires knowledge of the bearing dynamics and the controller, for which the true dynamics were used (i.e. the dynamics that were used to simulate the data-generating system).

Using the recursive estimator, a total set of 10, 000 samples were used to estimate $B_{P,2}$, in blocks of $N = 100$ samples. The result obtained with the Kalman recursion of Section 6.4 is given in Figure 6.2, showing the small resulting estimation error of $\hat{P}_2$.

6.6   Discussion

The simulations show that favorable results can be obtained with the described results. Concerning the method we make the following closing remarks:

- In the given approach, knowledge on the bearing dynamics, including the noise on the position measurements, is assumed to be known. In the context of this thesis, this information is obtained in prior identification experiments. One could argue that the identification of the bearing dynamics and the tooltip dynamics can be done simultaneously. After all, in the data of used in the identification of the tooltip dynamics, information is also present on the bearing dynamics and the noise dynamics on the position measurements. Nevertheless, the added value from combining the two identification problems is limited. The amount of data used in the identification of the bearing dynamics by and large exceeds the data used in the identification of the tooltip dynamics. Hence, the decrease of the variance of the estimated pa-
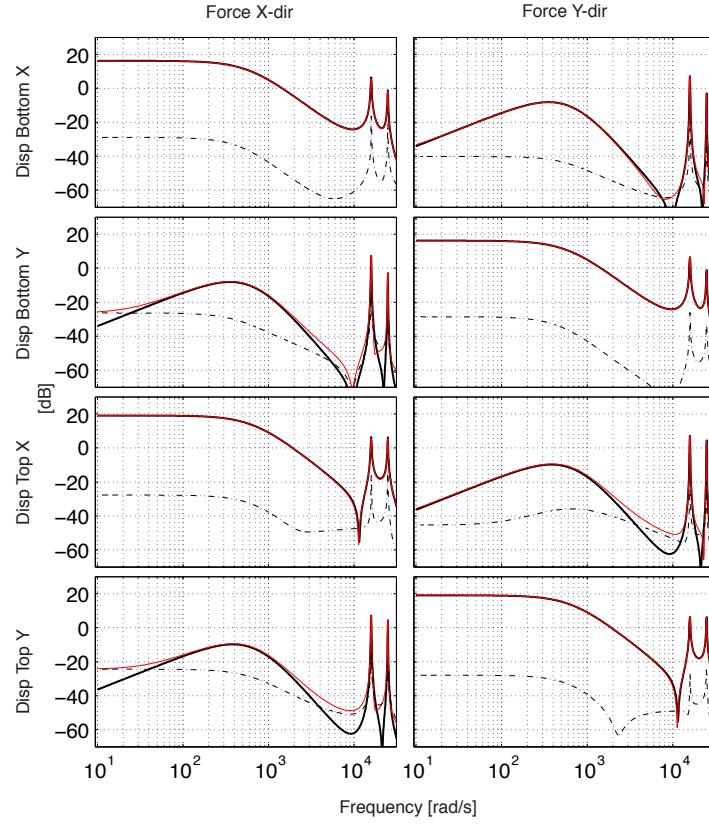
Figure 6.2: Identification of the tooltip dynamics with the recursive estimator using a total data set of 10, 000 samples. Solid black: true plant $P_2(e^{j\omega})$, solid red: estimated plant $\hat{P}_2(e^{j\omega})$, dotted: estimation error $P_2(e^{j\omega}) - \hat{P}_2(e^{j\omega})$.

rameters of the model of the bearing dynamics by using also the data used for identification of the tooltip dynamics, will be limited.

- As we are dealing with experiments in this method in which the input is determined by the milling process, and not user-defined, one may end up in the situation that the input is not sufficiently exciting. The possible remedies in this case are straightforward, and would involve changing the cutting conditions in order to get a cutting process that is more exciting (e.g. by applying varying cutting depth and width), or by reduction of the model complexity.

6.7 Summary and conclusions

In this chapter, identification of the tooltip dynamics was addressed. The main challenge in this identification problem is to apply a known excitation force to the tooltip. Having reviewed several alternatives, the route followed in this chapter is to identify the tooltip dynamics using data obtained during a milling experiment in which the cutting forces are measured. The amount of data that can be generated in this way is limited, as is the control over the spectral properties of the input. Hence, in order to reduce the complexity of the identification, usage is made of the property of the AMB system that all modes of the AMB spindle in the relevant frequency range are observable in the bearing dynamics. The bearing dynamics of the AMB spindle, including the noise dynamics, are assumed known, for example from prior identification experiments.

It is shown that under these conditions, identification of the tooltip dynamics can be expressed as a prediction error identification problem, with a particular closed-loop model parametrization that is linear in the parameters, and a fixed noise model that is non-minimum phase. Computation of a causally invertible spectral factorization of the noise spectrum, allows to formulate a stable predictor model for the resulting system. Since the prediction error is affine in the parameters results, the identification problem can be solved using a linear regression approach. An extension of the resulting LS estimator to recursive estimation was made by application of the Kalman state estimator, allowing to process also longer measurements efficiently.

A simulation study based on the properties of the EAAT AMB spindle was used to demonstrate the favorable properties of the given method.

Chapter 7

Compensation of runout disturbances in AMB signals

7.1 Introduction

In order to obtain useful estimates of the cutting forces from the signals of the AMB spindle, it is mandatory that the disturbances present in these signals due to runout are corrected for. In fact, this issue is dealt with in several of the indirect force estimation procedures that we reviewed in Section 1.3.1, see [122, 7, 30]. The procedure in these contributions is as follows. The key idea is that the runout disturbance is a function of the angular position. A measurement of this function is obtained by adding a rotary encoder to the spindle and measuring the runout disturbance for one or more rotations of the spindle, while no cutting is performed. During cutting, this measured runout disturbance is then substracted from the measurements, using the encoder information for synchronization.

This procedure has two main disadvantages. The first is that with the above method, only measurements of the runout disturbance are obtained at a discrete angular position grid (which is not necessarily equidistant). Consequently, interpolation is needed to compensate for runout at angular positions not included in the measurement grid. Besides, in this method, noise in the measurements is not taken into account explicitly.

The second disadvantage is that this procedure relies on the availability of a rotary encoder. Frequently spindles are not equipped with such a sensor. Since even very small variations of the rotational speed would result in loss of synchronization between the runout signal and its correction, this method cannot be applied when no angular position information is available.

To deal with these issues, a model-based runout compensation strategy will be presented in this chapter. In this method, the runout disturbance is modeled by

a parametrized truncated Fourier series expansion, of which the parameters are estimated prior to a cutting job. For this, several rotations of the spindle can be used to obtain parameter estimates with small variance. Using the estimated parameters, the model directly predicts the runout for a given angular position during cutting, which can be used to correct the runout disturbance.

A main issue in this chapter is how to deal with distortions in the information on the angular position. We will arrive at solutions from two different angles. Considering first the case that *perfect information* on the angular position is available, resulting in a standard linear regression scheme, we derive a solution for the case that noisy angular position measurements are available. The solution follows a strategy to eliminate the bias that would be incurred by naive application of the estimator, derived from the assumption that the angular position information is noiseless. The other route starts with the assumption that *no information* on the angular position is available. In this approach, the angular position is estimated from the observations using an Extended Kalman Filter. To this end, the prior knowledge on the almost linear increase of the angular position of the rotor as function of time is included in the model of the observations. In the case that noisy angular positions are available, these can be employed to improve the performance of the estimator. The performance of the resulting runout identification and correction schemes varies, depending on the level of disturbances in the observations, and are analyzed in this chapter.

This chapter is structured as follows. In Section 7.2 we will first define the problems of runout identification and runout compensation. Subsequently in Section 7.3, solutions are formulated for the case that exact angular position information is available, the results of which are used to analyze the situation that noisy angular position measurements are at hand (Section 7.4). For the case no information on the angular position is available we will use an Extended Kalman Filter in Section 7.5 to solve the runout identification and compensation problems. In the same section this approach is extended for the case that noisy angular positions are available. In Section 7.6 the approaches are compared by simulation and results obtained by application to real data are shown.

7.2 Problem description

In this section, the problems of runout idenfication and compensation are formulated. In order to do so, consider the measurement setup depicted in Figure 7.1. Here $y_0(t)$ is a signal that is to be measured, which is disturbed by zero mean measurement noise v, as well as a disturbance w which has the property that it is almost periodic. To make this explicit, we model w as the continuous mapping $\phi \to w(\phi)$, with ϕ the phase angle. This mapping is periodic, i.e. it has the property that $w(\phi + 2\pi) = w(\phi)$. With this, we can write $w(t) = w(\phi(t))$, where $\phi(t)$ is used to denote the phase angle at time instant t. Observe that in the application of
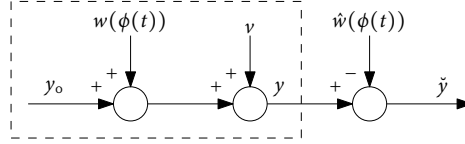
Figure 7.1: Measurement model for identification and correction of almost periodic disturbances. The quasi-periodic disturbance $w(\phi(t))$ is corrected for by estimating the parameters of a model of $w(\phi(t))$, and then substracting the model from the measurements y.

runout estimation in the AMB spindle signal, ϕ has the interpretation of the angular position of the rotor. Concerning $w(\phi(t))$ the following assumptions are made:

1. Define the increments of the phase angle as $\Delta\phi(t) = \phi(t) - \phi(t-1)$. It is assumed that $\Delta\phi(t) = \Delta\phi_0 + \delta\phi(t)$ with $\delta\phi(t)$ small compared to $\Delta\phi_0$.

2. $w(\phi)$ is a static function of ϕ. The average value of $w(\phi)$ is zero for $\phi = 0 \ldots 2\pi$.[1]; the function may be different for different values of $\Delta\phi_0$.

3. $w(\phi)$ is independent of y_0.

4. Let $W(\omega)$ be the (continuous) Fourier transform of $w(\phi)$. There exists a frequency B such that $W(\omega) = 0$ for all $\omega > B$ and $B < \frac{\pi}{\max(\Delta\phi(t))}$.

Evaluating these assumptions in the context of runout estimation in the signals of AMB spindles, we note that the first assumption implies that the rotational speed is nearly constant. As discussed in Chapter 2, runout disturbances are rotational speed dependent, meaning that $w(\phi)$ will be different for different rotational speed settings. Assumption 2 entails that the variations of the rotational speed around a given setpoint are small enough to have no significant effect on the profile of the runout disturbance. The third assumption implies that the runout disturbance is independent of the position and/or current signals of the bearings, which is in line with the model of Chapter 2. The fourth assumption means that the rotational speed is small enough so that no aliasing arises when taking measurements of $w(\phi)$ at angular positions $\phi(t)$.

Under these assumptions, we can model $w(\phi)$ by a truncated Fourier series expansion, given by

$$w(\phi) = \sum_{m=1}^{M} a_m \cos(m\phi) + b_m \sin(m\phi) \tag{7.1}$$

---

[1]The assumption of zero average is merely made for convenience. A situation where w has non-zero average can simply be dealt with by determining the average separately. This is straightforward.

which we may also write as $w(\phi) = H^T(\phi)\theta_o$ with

$$H^T(\phi) = \begin{bmatrix} \cos(\phi) & \dots & \cos(M\phi) & \sin(\phi) & \dots & \sin(M\phi) \end{bmatrix}$$
$$\theta_0^T = \begin{bmatrix} a_1 & \dots & a_M & b_1 & b_2 & \dots & b_M \end{bmatrix}.$$

With this, the obtained model is

$$y(t) = y_0(t) + H(\phi(t))\theta_0 + v(t). \tag{7.2}$$

We define the following two problems:

1. **Runout estimation:** With a set of measurement data $\{y(t)\}_{t=1\dots N}$, for which $y_0 = 0$, estimate θ_0.

2. **Runout compensation:** Using estimate $\hat{\theta}$, estimate the disturbance $w(\phi(t))$, denoted by $\hat{w}(\phi(t))$, and with that perform the runout compensation

$$\breve{y}(t) = y(t) - \hat{w}(\phi(t)). \tag{7.3}$$

In the following sections we will address these two problems. We will first consider the ideal situation where phase angle $\phi(t)$ is known exactly (Section 7.3), resulting in an ordinary linear Least Squares (LS) problem. Subsequently in Section 7.4 we treat the case that noisy measurements of $\phi(t)$ are available. This results in an errors in variables problem. In solving this problem, we will avoid the complexity of striving for statistical efficiency and use the results of Section 7.3 to formulate a Bias Compensated Least Squares (BCLS) approach. Section 7.5 deals with the case that no phase information is available. Using the prior knowledge that the phase increases almost linearly, we can formulate a state space model of the observations and then formulate a state estimator to estimate both the parameters and and the phase angle, where state estimation is performed using an Extended Kalman Filter (EKF). For the case that noisy phase angle measurements are available, these can easily be added as extra observations, allowing to reduce the variance of the state estimates.

7.3 Exact information on the phase angle

In this section, a runout estimator and compensator is given for the case that exact information is available on phase angle $\phi(t)$. For this, define output estimate $\hat{y}(t) = H(\phi(t))\hat{\theta}$, then we can define the optimal parameter estimate by

$$\hat{\theta} = \arg\min_{\theta} \sum_{t=1}^{N} (y(t) - \hat{y}(t))^2. \tag{7.4}$$

Employing the notation

$$
\mathcal{Y} = \begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(N) \end{bmatrix}, \qquad \mathcal{H} = \begin{bmatrix} H^T(\phi(1)) \\ H^T(\phi(2)) \\ \vdots \\ H^T(\phi(N)) \end{bmatrix}, \qquad \mathcal{V} = \begin{bmatrix} v(1) \\ v(2) \\ \vdots \\ v(N) \end{bmatrix} \tag{7.5}
$$

it follows that we can reformulate equation (7.4) as

$$
\hat{\theta} = \arg \min_{\theta} |\mathcal{Y} - \mathcal{H}\theta|^2 \tag{7.6}
$$

showing that we have obtained a standard linear regression problem, with solution [83]:

$$
\hat{\theta} = K\mathcal{Y}, \text{ where} \tag{7.7}
$$

$$
K = (\mathcal{H}^T\mathcal{H})^{-1}\mathcal{H}^T. \tag{7.8}
$$

We summarize the main properties of the estimator. The estimate $\hat{\theta}$ is unbiased, and if the covariance of $\mathcal{V}$, denoted by $\Pi_{\mathcal{V}}$, is available, minimum variance is obtained by the estimator

$$
K = (\mathcal{H}^T\Pi_{\mathcal{V}}^{-1}\mathcal{H})^{-1}\mathcal{H}^T\Pi_{\mathcal{V}}^{-1}. \tag{7.9}
$$

In that case the covariance of $\hat{\theta}$ is

$$
\Pi_{\hat{\theta}} = \mathbb{E}(\hat{\theta} - \theta_0)(\hat{\theta} - \theta_0)^T \tag{7.10}
$$

$$
= (\mathcal{H}^T\Pi_{\mathcal{V}}^{-1}\mathcal{H})^{-1}. \tag{7.11}
$$

To allow for online implementation of the estimator, we also give a recursive algorithm. Consider the situation we have a prior estimate of θ_0, which will be denoted by θ_{k-1}, and let Π_{k-1} be its covariance matrix. Assume θ_{k-1} is uncorrelated with $\mathcal{Y}$. Then the optimal estimate is given by

$$
\hat{\theta}_k = \theta_{k-1} + \arg \min_{d\theta} |\mathcal{Y} - \mathcal{H}(\theta_{k-1} + d\theta)|^2 \tag{7.12}
$$

$$
= \theta_{k-1} + K(\mathcal{Y} - \mathcal{H}\theta_{k-1}), \text{ where} \tag{7.13}
$$

$$
K = (\mathcal{H}^T\Pi_{\mathcal{V}}^{-1}\mathcal{H} + \Pi_{k-1}^{-1})^{-1}\mathcal{H}^T\Pi_{\mathcal{V}}^{-1}. \tag{7.14}
$$

In this case, the covariance Π_k is given by

$$
\Pi_k = (\mathcal{H}^T\Pi_{\mathcal{V}}^{-1}\mathcal{H} + \Pi_{k-1}^{-1})^{-1}. \tag{7.15}
$$

Runout compensation for $t > N$ in the case that exact phase information is available is performed by evaluating equation (7.3).

7.4 Noisy measurements of the phase angle

In the previous section, the estimator has been derived based on perfect knowledge of the phase angle. In this section we will investigate the situation when only noisy measurements of the phase angle are available.

Let $\psi(t) = \phi(t) + v_\phi(t)$ represent the measurement of the phase angle with noise $v_\phi(t)$. It will be assumed that this noise is a zero mean white Gaussian disturbance with variance σ_ϕ^2 that is uncorrelated with $v(t)$. At every time instant we now have two measurements:

$$y(t) = H(\phi(t))\theta_0 + v(t) \tag{7.16a}$$
$$\psi(t) = \phi(t) + v_\phi(t). \tag{7.16b}$$

With this, estimation of θ_0 becomes an errors-in-variables (EIV) problem, which is intricate due to the nonlinearity of $H(\phi)$. This makes formulation of a maximum likelihood estimator complicated. An alternative solution is to pursue a Bias Compensated Least Squares (BCLS) approach. The idea behind this is as follows (see [152]). In general, naive application of an LS estimator to EIV problems yields biased estimates. In the BCLS approach the estimator is modified in such a way that this bias is eliminated. We will follow this approach and use the results of the previous section.

First observe that direct application of a LS estimator of indeed results in asymptotically biased estimates. This is seen as follows. For infinite length data, linear regression of $y(t)$ on $H(\psi(t))$ satisfies [83]

$$\mathbb{E}H^T(\psi(t))H(\psi(t))\hat{\theta}_{LS} = \mathbb{E}H(\psi(t))^T y(t). \tag{7.17}$$

Note that from the definition of $H(\cdot)$, it is immediate that

$$H^T(\psi)H(\psi) = 1 \text{ for all } \psi. \tag{7.18}$$

Furtheremore,

$$\mathbb{E}H(\psi(t))^T y(t) = \mathbb{E}H(\psi(t))^T [H(\phi(t))\theta_0 + v(t)] \tag{7.19}$$
$$= \mathbb{E}H(\psi(t))^T H(\phi(t))\theta_0 + \mathbb{E}H(\psi(t))^T v(t) \tag{7.20}$$

Note since $v(t)$ and $\psi(t)$ are uncorrelated, the second term is zero. Using equation (7.16b), it follows for the first term that

$$\mathbb{E}H(\psi(t))^T H(\phi(t))\theta_0 = \mathbb{E}\left\{H(\phi(t) + v_\phi(t))^T\right\} H(\phi(t))\theta_0 \tag{7.21}$$

where it is used that $H(\phi(t))\theta_0$ is deterministic. In appendix B.4.1 it is shown that for v_ϕ Gaussian distributed

$$\mathbb{E}H(\phi(t) + v_\phi(t)) = \mathcal{Q}^{-1}H(\phi(t)), \tag{7.22}$$

with

$$\mathcal{Q} = \mathrm{diag}\big(e^{\sigma_\phi^2/2}, e^{(2\sigma_\phi)^2/2}, \ldots, e^{(M\sigma_\phi)^2/2}, e^{\sigma_\phi^2/2}, e^{(2\sigma_\phi)^2/2}, \ldots, e^{(M\sigma_\phi)^2/2}\big) \qquad (7.23)$$

Using equations (7.18) and (7.22) in equation (7.17), we obtain that for infinite length data (i.e. $N \to \infty$)

$$\hat{\theta}_{\mathrm{LS}} = \mathcal{Q}^{-1}\theta_0. \qquad (7.24)$$

Clearly, the LS estimator results in asymptotically biased estimates. Bias elimination can be achieved by defining

$$\hat{\theta}_{\mathrm{BCLS}} = \mathcal{Q}\hat{\theta}_{\mathrm{LS}}. \qquad (7.25)$$

For finite length data, this results in the following estimator. Let $\bar{\mathcal{H}}$ be defined by

$$\bar{\mathcal{H}} = \begin{bmatrix} H^T(\psi(1)) \\ H^T(\psi(2)) \\ \vdots \\ H^T(\psi(N)) \end{bmatrix} \qquad (7.26)$$

With this the BCLS estimator is defined as

$$K_{\mathrm{BCLS}} = \mathcal{Q}(\bar{\mathcal{H}}^T \Pi_{\mathcal{V}}^{-1} \bar{\mathcal{H}})^{-1} \bar{\mathcal{H}}^T \Pi_{\mathcal{V}}^{-1}. \qquad (7.27)$$

Recursive implementation of this estimator can be done similarly as the previous section, with

$$K = \mathcal{Q}(\bar{\mathcal{H}}^T \Pi_{\mathcal{V}}^{-1} \bar{\mathcal{H}} + \Pi_{k-1}^{-1})^{-1} \bar{\mathcal{H}}^T \Pi_{\mathcal{V}}^{-1} \qquad (7.28)$$

and where the covariance Π_k is now estimated as

$$\Pi_k = \mathcal{Q}(\bar{\mathcal{H}}^T \Pi_{\mathcal{V}}^{-1} \bar{\mathcal{H}} + \Pi_{k-1}^{-1})^{-1} \mathcal{Q}. \qquad (7.29)$$

Again, runout compensation is straightforward by estimating the runout disturbance according to

$$\hat{w}(\phi(t)) = H(\psi(t))\hat{\theta}_{\mathrm{BCLS}}, \text{ for } t > N \qquad (7.30)$$

which uses the measured phase angle $\psi(t)$ to compute the estimate of the quasi-periodic signal.

7.5   No information on the phase angle

In this section we will consider the situation that no information on the phase angle is available, in which case the phase angle must be estimated from the data as well. The solution for runout estimation in this section is taken from the EKF solution for frequency tracking of nonsinusoidal periodic signals, presented by Parker and Anderson [125]. First this solution is given. Subsequently, a solution is given for runout correction, following a similar EKF approach, but with a reduced model complexity. Based on these runout estimation and correction approaches, we will give extensions that allows to improve the estimate of the phase angle in case noisy phase angle measurements are available.

Recall from Section 7.2, that the observations are modeled as $y(t) = y_0(t) + w(\phi(t)) + v(t)$, with the quasi-periodic signal denoted as $w(\phi(t))$, $y_0(t) = 0$ during runout estimation, and $v(t)$ additive noise. The starting point for the EKF approach is to model the mapping $w(\phi(t))$ using a state space model, given by

$$x(t+1) = Ax(t) + n(t) \tag{7.31a}$$
$$y(t) = w(x(t)) + v(t) \tag{7.31b}$$

where

$$x(t) = \begin{bmatrix} \theta_0 & \Omega(t) & \phi(t) \end{bmatrix} \tag{7.32}$$

with θ_0 and ϕ as defined before, and where Ω is the angular frequency. Furthermore,

$$A = \left[\begin{array}{c|cc} I_{2M \times 2M} & \multicolumn{2}{c}{0} \\ \hline 0 & 1 & 0 \\ & T_s & 1 \end{array}\right], \text{ and}$$

$$w(x(t)) = \sum_{m=1}^{M} a_m \cos(m\phi(t)) + b_m \sin(m\phi(t))$$

and n is a white noise process that is uncorrelated with v, and that has zero mean and variance Q. In this model, v is modeled as a white noise process. We remark that this no limitation. When it is known a priori that the measurement noise can be represented by $v(t) = H(q)e(t)$ with $e(t)$ a white noise process and $H(q)$ a known transfer function, the state space model in (7.31) can easily be extended to include the states of $H(q)$.

As we model θ_0 to be static parameters, typically covariance matrix Q is nonzero only at element $(2M + 1, 2M + 1)$, meaning that angular frequency $\Omega(t)$ is modeled as a random walk with a given variance. From the structure of the state space model, it follows that the model dictates the angular phase to evolve as

$$\phi(t) = \phi_0 + \sum_{\tau=1}^{t} \Omega(\tau) T_s. \tag{7.33}$$

with ϕ_0 the initial value. Using the above state space model of the observations, estimation of θ_0 and the unknown $\phi(t)$ can now be treated as a state estimation problem. Due to the nonlinearity of the model, the EKF is applied [83, 125]. This filter is given by the following recursion:

- Measurement update:

$$\hat{x}(t|t) = \hat{x}(t|t-1) + K(t)[y(t) - w(\hat{x}(t|t-1))] \tag{7.34}$$

$$K(t) = \Pi(t|t-1)W(t)[W(t)\Pi(t|t-1)W^T(t) + R]^{-1} \tag{7.35}$$

$$\Pi(t|t) = [I - K(t)W(t)]\Pi(t|t-1) \tag{7.36}$$

with $K(t)$ the Kalman gain, $R = \sigma_v^2$, and $\Pi(t|t)$ the error covariance.

- Time update:

$$\hat{x}(t+1|t) = A\hat{x}(t|t) \tag{7.37}$$

$$\Pi(t+1|1) = A\Pi(t|t)A^T + Q \tag{7.38}$$

where $W(t)$ is the Jacobian of $w(x(t))$ at $\hat{x}(t|t-1)$, i.e.

$$W(t) = \frac{\partial}{\partial x}w(x(t))|_{x(t)=\hat{x}(t|t-1)} = \begin{bmatrix} \cos(\hat{\phi}(t|t-1)) \\ \vdots \\ \cos(M\hat{\phi}(t|t-1)) \\ \sin(\hat{\phi}(t|t-1)) \\ \vdots \\ \sin(M\hat{\phi}(t|t-1)) \\ 0 \\ \gamma(t|t-1) \end{bmatrix}^T \tag{7.39}$$

where

$$\gamma(t|t-1) = \sum_{m=1}^{M} -\hat{a}_m(t|t-1)m\sin(m\hat{\phi}(t|t-1)) + \hat{b}_m(t|t-1)m\cos(m\hat{\phi}(t|t-1)). \tag{7.40}$$

The procedure for runout correction is similar, the main difference is that the parameters θ are now assumed to be known. Hence, when switching from runout estimation to runout correction at time $t = N+1$, we reduce the order of the model and continue the state estimation with the model

$$x(t+1) = Ax(t) + n(t) \tag{7.41a}$$

$$y(t) = w(x(t)) + v(t) \tag{7.41b}$$

where $x(t) = \begin{bmatrix} \Omega(t) & \phi(t) \end{bmatrix}$,

$$A = \begin{bmatrix} 1 & 0 \\ T_s & 1 \end{bmatrix},$$

and $w(x(t))$ as define before, now parametrized with $\hat{\theta} = \hat{\theta}(N|N)$. The EKF equations are similar, with

$$H(t) = \begin{bmatrix} 0 \\ \gamma(t|t-1) \end{bmatrix}$$

and

$$\gamma(t|t-1) = \sum_{m=1}^{M} -\hat{a}_m m \sin(m\hat{\phi}(t|t-1)) + \hat{b}_m m \cos(m\hat{\phi}(t|t-1)). \qquad (7.42)$$

The runout correction recursion is initiated by the last state estimates of the runout identification recursion and the corresponding covariance.

In the runout correction stage, we no longer require y_0 to be zero. In the model of equation (7.41), presence of y_0 is not explicitly modeled. Instead, nonzero y_0 is considered to be included in the noise process v. Hence in switching from runout identification to runout correction, it may be necessary to adjust to modeled variance of v.

In [125], Parker and Anderson also give a theoretical analysis of the above EKF approach. This analysis shows that there is a measure of decoupling in the estimator, i.e. the Fourier coefficients are estimated as if the phase angle and angular frequency estimates are correct, and the phases and frequencies are estimated as if the amplitude estimates are correct. It is demonstrated that the EKF can be interpreted as a set of interconnected phase-locked loops, able to track the frequency variations in the observations.

Although this property is an attractive feature of the EKF approach, there are also some related pitfalls. The first is that the estimator needs to lock onto the right fundamental frequency, while it has a tendency to lock onto fractions or multiples of the true fundamental frequency. Such behavior can be avoided by providing good initial estimates of the fundamental frequency. Also, providing a smaller value of the variance related to this initial estimate can slow down and tamper the initial transient behavior, increasing the probability of a successful lock on. The second pitfall is to maintain good frequency tracking, without loosing the frequency lock. The rate of variation is mainly determined by the matrix Q. Larger values of Q result in a faster (and also noisier) response, but also increase the risk of loosing the frequency lock. Hence, practical implementation of the EKF approach requires careful selection of the initial settings and model matrices.

While the EKF approach is introduced for the case that no phase information is available, it is quite simply adapted for the situation we do have noisy phase measurements. This is an attractive extension, since adding such additional information to the EKF is expected to improve the performance. In fact, we may not only

find improved estimates of the Fourier coefficients, but also improved performance regarding frequency locking and tracking. Depending on the noise level v_ϕ, the resulting EKF can also have attractive features compared to the BCLS estimator. Although the latter may be unbiased, for larger values of noise level v_ϕ we can expect the variances to deteriorate[2]. Moreover, at the runout correction stage, no compensation is performed for the errors in ψ, and hence for larger v_ϕ the bias in the runout corrections can be large. In contrast, the EKF that is provided with the phase measurements, also uses the phase information that is present in the observations y. As a result, both in the runout identification and the runout correction stage, phase estimates are used with a smaller variance than ψ, resulting in a better performance. This comes at a price of a more complex identification and correction filter.

With measurement of the phase available the EKF can be adapted as follows. The state space model of (7.31) now satisfies

$$x(t+1) = Ax(t) + n(t) \tag{7.43a}$$

$$z(t) = \left[\begin{array}{c} y(t) \\ \psi(t) \end{array}\right] = \left[\begin{array}{c} w(x(t)) \\ \phi(t) \end{array}\right] + \left[\begin{array}{c} v(t) \\ v_\phi(t) \end{array}\right]. \tag{7.43b}$$

Replacing in the EKF recursion equations

$$\hat{x}(t|t) = \hat{x}(t|t-1) + K(t)\left[z(t) - \hat{z}(t|t-1)\right] \tag{7.44}$$

and letting $W(t)$ and R satisfy

$$W(t) = \left[\begin{array}{cc} \cos(\hat{\phi}(t|t-1)) & 0 \\ \vdots & \vdots \\ \cos(M\hat{\phi}(t|t-1)) & 0 \\ \sin(\hat{\phi}(t|t-1)) & 0 \\ \vdots & \vdots \\ \sin(M\hat{\phi}(t|t-1)) & 0 \\ 0 & 0 \\ \gamma(t|t-1) & 1 \end{array}\right]^T, \quad R = \left[\begin{array}{cc} \sigma_v^2 & 0 \\ 0 & \sigma_\phi^2 \end{array}\right] \tag{7.45}$$

we obtain an EKF that uses the observations ψ to improve the estimate of ϕ (and the other parameters). This procedure can be performed similarly for the EKF used for runout correction.

We conclude this section by noting that the EKF approach is easily extended to the multivariable case. In the case of runout identification and correction in AMB signals, the development of the angular position as a function of time is identical for

---

[2]This is easily seen from equations (7.27). For growing values of v_ϕ, the elements of $\mathcal{Q}$ corresponding to the highest order Fourier terms grow rapidly. As a result, the variance of the parameters belonging to these terms grows accordingly.

all observations. Hence, instead of estimating it from each observation separately, it is straightforward to formulate a state space model that includes all observations, allowing a more accurate estimation of the angular position. The computational effort for such multivariable approach is significantly higher though. This is particularly seen from the expression for the Kalman gain in equation (7.35), where the dimension of the matrix that is inverted equal is to the dimension of the vector of observations.

7.6 Results

In this section we will demonstrate the performance of the estimators described in the previous sections. We will first give a comparative simulation study of the different approaches in this chapter by means of Monte Carlo study (Section 7.6.1). Then in Section 7.6.2 we apply the EKF approach in a realistic simulation enviroment to assess its performance. Results obtained after application to real measurement data are discussed in Section 7.6.3.

7.6.1 Monte carlo simulations

To study the properties of the estimators discussed in the previous section, a simulation study has been performed. Observations $y(t)$ were generated according to equation (7.2), with $y_0 = 0$, $M = 5$ and

$$
\begin{aligned}
a_1 &= 1.000 & b_1 &= 0.6000 \\
a_2 &= 0.030 & b_2 &= 0.1300 \\
a_3 &= 0.110 & b_3 &= 0.0500 \\
a_4 &= 0.005 & b_4 &= -0.1200 \\
a_5 &= 0.040 & b_5 &= 0.0400.
\end{aligned}
$$

The runout disturbance was modeled to have a fundamental frequency of $\omega_0 = 1000\pi$ rad/s, with variations of up to ± 2.0 rad/s[3]. The sampling time T_s satisfies $T_s = 50\mu$s. The observations were disturbed with zero mean white noise v, with variance set such that the signal-to-noise ratio, defined as SNR $= 10\log_{10}\left(\sum_{m=1}^{5} a_m^2 + b_m^2\right)/\sigma_v^2$ was 46.5 dB, 30.5 dB, or 14.5 dB, depending on the trial. Furthermore, in each trial a measurement of the phase $\psi(t)$ was generated, according to equation (7.16). Here v_ϕ is a zero mean white Gaussian noise with variance $\sigma_\phi^2 = \left\{\left(\frac{2\pi}{1000}\right)^2, \left(\frac{2\pi}{100}\right)^2, \left(\frac{2\pi}{10}\right)^2\right\}$. For each SNR level and noise level of v_ϕ, 200 Monte Carlo simulations were run, where in each simulation a different realization of v and v_ϕ was generated, while $w(\phi(t))$ was identical for all simulations. Each simulation was run over over $N_f =$

---

[3]Note that although these variations are small in comparison to the base frequency, modeling the disturbance to have a constant frequency would quickly lead to very significicant phase errors.

4000 samples; the samples for $t = 1..N$, with $N = 2000$ were used for runout identification, and the samples for $t = N + 1..N_f$ for runout correction. On each data set, four runout identification and correction schemes discussed in this chapter were applied, i.e.

1. LS: runout identification with the LS estimator; runout correction by prediction via (7.3), both performed using the noise free $\phi(t)$. In this scheme the phase measurement is not used, so evaluation of the results for different levels of σ_ϕ is unnessary. Hence, the simulations are only run over the data sets corresponding one single value of σ_ϕ, in this case $\sigma_\phi = \frac{2\pi}{1000}$.

2. BCLS: runout identification with the BCLS estimator; runout correction by prediction via (7.30), both using the noisy phase measurements $\psi(t)$.

3. EKF: Runout identification and correction with the EKF without phase information (as with the LS scheme, this scheme is only run over the datasets corresponding to $\sigma_\phi = \frac{2\pi}{1000}$).

4. EKF+: Runout identification and correction with the EKF using the noisy phase measurements $\psi(t)$.

Both EKF solutions were initiated with $\hat{a}_1(0|-1) = \hat{b}_1(0|-1) = \left(\frac{1}{2N}\sum_{t=1}^{N} y(t)^2\right)^{1/2}$, $\hat{\Omega}(0|-1) = 3110$ rad/s, and $\hat{\phi}(0|-1) = 0$ rad. All other values of $\hat{x}(0|-1)$ were set to zero. Furthermore, $\Pi(0|-1) = \text{diag}(\rho, \rho/4, \rho/4, \rho/4, \rho/4, \rho, \rho/4, \rho/4, \rho/4, \rho/4)$, where $\rho = \hat{a}_1(0|-1)^2/4$. Covariance matrix Q had all zeros, except $Q(2M+1, 2M+1) = (0.02)^2$.

From the estimates $\hat{\theta}$ obtained over all 200 Monte Carlo simulations the bias and standard deviation are estimated. The results are listed in table 7.1 and 7.2, where only the statistics for the parameters a_m, $m = 1 \ldots M$ are given, those for b_m give a similar picture. From these tables, we observe the following:

- As expected, the LS estimator with perfect phase information has the lowest bias and variance for all values of the SNR and σ_ϕ.

- For $\sigma_\phi = 0.006$ and $\sigma_\phi = 0.063$, the performance of the BCLS estimator is close to that of the LS estimator. Here we see that in particular if the SNR is low, the variance error in the estimates is mostly determined by the measurement noise v, and the loss of performance due to disturbance v_ϕ is small. The results seem to corroborate the unbiasedness of the BCLS estimator, as in all simulations the average error is significantly smaller than the standard deviation. However, for $\sigma_\phi = 0.628$, the performance of the BCLS estimator is very poor. While coefficient a_1 is still estimated with reasonable accuracy, the variance of the higher order Fourier coefficients is increasingly worse. This effect can directly be understood from the structure of matrix $\mathcal{Q}$ in equation (7.26): for growing values of v_ϕ, the elements of $\mathcal{Q}$ corresponding to the

highest order Fourier terms grow very quickly. As a result, the variance of the parameters that belong to these terms increase rapidly as well.

- Comparing the standard deviations of the estimates obtained with the EKF to those of the LS estimator using perfect phase information, we see the EKF only performs slightly worse for smaller SNR. The logical explanation is that at small SNR the variance is mostly determined by v, and to a smaller extent by errors in the phase estimate. For larger SNR, these errors become more dominant, as can be observed from the table. The estimates obtained with the EKF also seem to have a bias. This should not be given too much emphasis, as in fact the model used in the EKF is overparametrized (both ϕ and b_1 represent a phase shift in the fundamental term of the expansion). Closer inspection shows that also the estimates of ϕ are slightly biased, compensating the bias in the Fourier coefficients.

- As expected, the EKF+ outperforms both the EKF and the BCLS estimator. Compared to the BCLS, the improvement is small when $\sigma_\phi = 0.006$. However, for $\sigma_\phi = 0.063$ the standard deviations of the estimates are reduced significantly, particularly for SNR = 46.5. The improvement for $\sigma_\phi = 0.628$ clearly shows that the BCLS is useless for this noise level: although the estimate produced by this estimator is unbiased, its variance is so large that no meaningful results are obtained. On the other hand, the EKF gives useful results. The improved locking and tracking of the EKF+ compared to the EKF has also been observed. Figure 7.2 illustrates this, where the estimates $\hat{\Omega}(t)$ obtained with both approaches are compared to the true values (for one of the Monte Carlo simulations with SNR = 30.5 dB, $\sigma_\phi = 0.006$). In this figure it is clear that in the initial stage the EKF+ converges to the true frequency faster, and also tracks the frequency variations better. It is also observed that the successful initial lock on is achieved for a much larger range of initial estimates.

Summarizing, we can conclude from these simulations that if no phase information is present, we can still obtain estimates with the EKF that have variance errors approaching the levels that would be attainable with perfect phase information. If we have phase information, then the EKF+ that includes this information yields the smallest variance and would be the estimator of choice. However, when σ_ϕ is very small, the cost for getting the best performance may not balance against the increased complexity. In such cases the BCLS might be a more attractive option.

To evaluate the performance of the runout correction for each scheme, we introduce the runout correction error defined as

$$\sigma_e = \left[\frac{1}{N_f - N} \sum_{\tau=N+1}^{N_f} \left(y(t) - \hat{w}(\phi(t))^2\right) \right]^{1/2}. \tag{7.46}$$
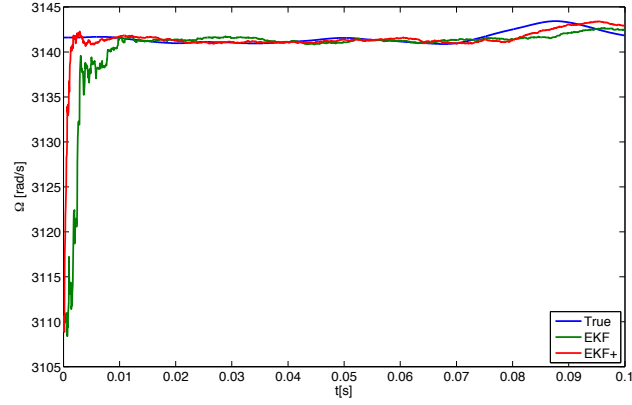
Figure 7.2: Comparison of the estimate of rotational speed $\hat{\Omega}(t)$ obtained with the EKF without phase information (green), and the EKF with phase information (SNR = 30.5, $\sigma_\phi = 0.006$) (red). True $\Omega(t)$ is depicted in blue.

The value of σ_e obtained with any of the four runout identification and correction scheme and averaged over the 200 Monte Carlo simulations, is listed in table 7.3 for each SNR level and noise level of ν_ϕ. This table leads to the same conclusion as with runout identification: with no phase information present, runout correction can be achieved with the EKF yielding error levels that approach the error levels achieved with perfect phase information (in the LS scheme). With phase information present, the best choice is the EFK+ approach that uses this information, albeit the increase for small values of σ_ϕ might not trade-off against higher complexity of the filter.

7.6.2 Application to AMB spindle setup, simulation

As a second test, a simulation experiment was performed in which a realistic simulation of a micro-milling experiment was made. For this, the simulation model of the EAAT AMB spindle described in Chapter 2 was used. Here the AMB spindle model was extended to also include runout disturbances. As the EAAT spindle has no angular encoder, the EKF approach for runout detection and correction was selected.

The simulation settings were as follows. The AMB spindle was set to rotate at 30 krpm, where the rotational speed was varied with ±7 rpm. Sample time T_s was 50μs. From $t = 0$ to $t = 0.5$ the cutting forces were nil, after that the milling was

7. COMPENSATION OF RUNOUT DISTURBANCES IN AMB SIGNALS

LS estimator using perfect phase information

σ_ϕ rad	SNR [dB]	$a_1 = 1.000$ Bias [·10⁻³]	Std.dev [·10⁻³]	$a_2 = 0.030$ Bias [·10⁻³]	Std.dev [·10⁻³]	$a_3 = 0.110$ Bias [·10⁻³]	Std.dev [·10⁻³]	$a_4 = 0.005$ Bias [·10⁻³]	Std.dev [·10⁻³]	$a_5 = 0.040$ Bias [·10⁻³]	Std.dev [·10⁻³]
	46.5	-0.014	0.183	0.001	0.185	-0.013	0.180	0.000	0.192	-0.007	0.164
	30.5	-0.019	1.161	-0.144	1.175	-0.195	1.209	-0.017	1.095	-0.044	1.131
	14.5	0.440	6.904	-0.914	6.383	0.633	7.197	-0.190	6.516	0.575	7.635
0.006	46.5	-0.009	0.255	-0.004	0.273	0.005	0.234	-0.003	0.266	-0.014	0.276
	30.5	-0.015	1.196	-0.148	1.186	-0.207	1.209	-0.032	1.105	-0.051	1.168
	14.5	0.474	6.925	-0.899	6.392	0.646	7.188	-0.198	6.500	0.584	7.649

BCLS estimator using noisy phase measurements

σ_ϕ rad	SNR [dB]	$a_1 = 1.000$ Bias [·10⁻³]	Std.dev [·10⁻³]	$a_2 = 0.030$ Bias [·10⁻³]	Std.dev [·10⁻³]	$a_3 = 0.110$ Bias [·10⁻³]	Std.dev [·10⁻³]	$a_4 = 0.005$ Bias [·10⁻³]	Std.dev [·10⁻³]	$a_5 = 0.040$ Bias [·10⁻³]	Std.dev [·10⁻³]
0.063	46.5	0.059	1.969	-0.057	1.731	0.007	1.813	-0.103	1.926	0.132	2.059
	30.5	0.044	2.042	0.084	2.396	0.012	2.065	-0.041	1.978	-0.043	2.443
	14.5	-0.268	7.829	1.051	6.820	0.120	7.543	0.722	6.918	0.540	7.265
0.628	46.5	0.375	12.68	-0.934	28.89	-3.727	83.82	-6.600	343.6	-50.60	2126
	30.5	0.577	13.58	1.849	28.71	-0.774	87.18	3.830	375.92	198.0	2185
	14.5	0.377	14.93	2.118	33.28	5.931	92.07	12.953	444.0	-164.2	2506

Table 7.1: Statistics of the LS estimator using perfect phase information and the BCLS estimator using noisy phase measurements, obtained in the Monte Carlo simulation study with the four schemes in Section 7.6.1.

EKF without using phase information

σ_ϕ rad	SNR [dB]	$a_1 = 1.000$		$a_2 = 0.030$		$a_3 = 0.110$		$a_4 = 0.005$		$a_5 = 0.040$	
		Bias [$\cdot 10^{-3}$]	Std.dev [$\cdot 10^{-3}$]	Bias [$\cdot 10^{-3}$]	Std.dev [$\cdot 10^{-3}$]	Bias [$\cdot 10^{-3}$]	Std.dev [$\cdot 10^{-3}$]	Bias [$\cdot 10^{-3}$]	Std.dev [$\cdot 10^{-3}$]	Bias [$\cdot 10^{-3}$]	Std.dev [$\cdot 10^{-3}$]
	46.5	3.344	1.585	1.474	0.732	0.813	0.431	-2.701	1.294	1.096	0.545
	30.5	3.284	2.377	1.327	1.454	0.624	1.296	-2.720	2.022	1.023	1.334
	14.5	3.206	9.355	0.407	6.752	1.232	7.240	-2.730	8.529	1.242	8.227

EKF+ using noisy phase measurements

σ_ϕ rad	SNR [dB]	$a_1 = 1.000$		$a_2 = 0.030$		$a_3 = 0.110$		$a_4 = 0.005$		$a_5 = 0.040$	
		Bias [$\cdot 10^{-3}$]	Std.dev [$\cdot 10^{-3}$]	Bias [$\cdot 10^{-3}$]	Std.dev [$\cdot 10^{-3}$]	Bias [$\cdot 10^{-3}$]	Std.dev [$\cdot 10^{-3}$]	Bias [$\cdot 10^{-3}$]	Std.dev [$\cdot 10^{-3}$]	Bias [$\cdot 10^{-3}$]	Std.dev [$\cdot 10^{-3}$]
0.006	46.5	-0.020	0.192	0.004	0.190	-0.011	0.181	-0.001	0.205	-0.006	0.168
	30.5	-0.035	1.175	-0.138	1.172	-0.191	1.207	-0.017	1.095	-0.045	1.133
	14.5	0.457	6.908	-0.906	6.376	0.590	7.189	-0.209	6.516	0.562	7.627
0.063	46.5	0.532	0.762	0.229	0.368	0.128	0.262	-0.404	0.603	0.166	0.296
	30.5	0.268	1.353	0.175	1.229	0.004	1.094	-0.169	1.226	0.081	1.154
	14.5	-0.431	7.355	0.993	6.671	-0.013	7.196	0.480	6.611	0.766	6.701
0.628	46.5	3.145	1.590	1.391	0.695	0.777	0.446	-2.517	1.256	1.034	0.538
	30.5	2.829	2.130	1.416	1.324	0.715	1.248	-2.546	1.912	1.161	1.315
	14.5	0.555	8.106	-0.085	7.246	0.419	6.935	-1.251	7.617	0.653	7.242

Table 7.2: Statistics of the EKF that uses no phase information, and the EKF that use noisy phase measurements obtained in the Monte Carlo simulation study with the four schemes in Section 7.6.1.

7. COMPENSATION OF RUNOUT DISTURBANCES IN AMB SIGNALS

		AVG runout correction error $[\cdot 10^{-3}]$			
σ_ϕ	SNR	LS	BCLS	EKF	EKF+
	46.5	5.6	8.3	5.8	5.7
0.006	30.5	35.5	36.0	35.7	35.0
	14.5	223.6	223.7	224.1	223.6
	46.5		61.1		5.8
0.063	30.5		70.5		36.0
	14.5		231.9		223.8
	46.5		1995		5.8
0.628	30.5		2057		36.0
	14.5		2313		224.4

LS: using perfect phase measurements

BCLS: using noisy phase measurements

EKF: no phase phase measurements

EKF+: using noisy phase measurements

Table 7.3: Comparison of the average runout correction error obtained in the Monte Carlo simulation study with the four schemes in Section 7.6.1. The AVG runout correction error obtained with the BCLS compared is acceptable for small σ_ϕ (and small SNR). However, for large σ_ϕ, the EKF+ clearly outperforms the BCLS approach.

simulated to start. The resulting position signal of the bottom X bearing is depicted in the top plot of Figure 7.3.

The EKF with order $M = 5$ was applied, where the initial settings were chosen in a similar fashion as in the simulation study in the previous section. Here the first 4000 samples were used for runout indentification, and from $t = 0.2$s onwards runout correction was applied using the estimated Fourier coefficients. The result of the EKF after runout correction is depicted in the middle plot of Figure 7.3. The corrected signal is compared to the signal that was obtained by simulating the AMB spindle setup without the runout disturbance, the resulting error is shown in the bottom plot of Figure 7.3.

From these plots we observe the success of the runout correction, the runout correction error is small in comparison to the response of the AMB spindle to the noise and cutting force signal. Noteworthy to point out, is that the noise in this position measurements is not white (as it is filtered by the sensitivity of the closed-loop), and neither is the response to the cutting force in the measurement data. Nevertheless, the EKF performs well in tracking the frequency variations, also after $t = 0.5$ when the milling has started. This can also be observed in Figure 7.4, in which the estimated angular frequency is compared to the true angular frequency.
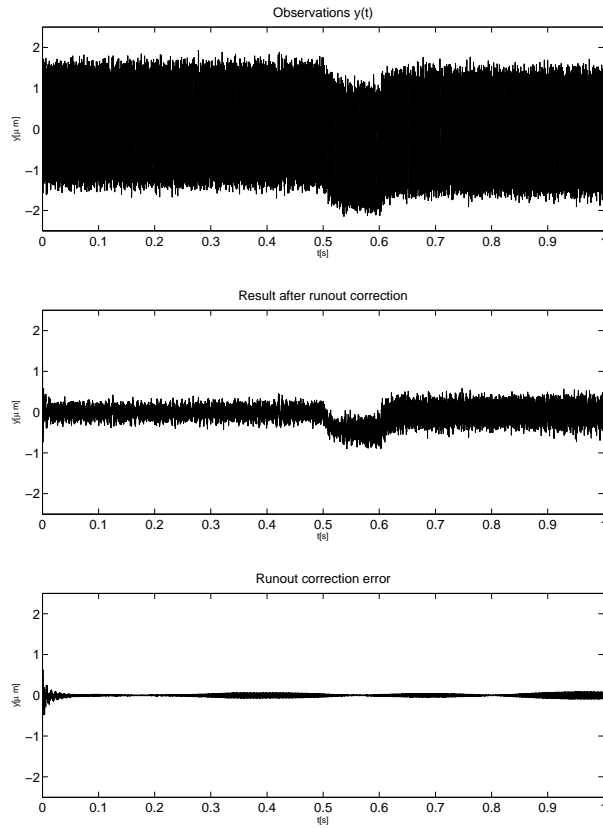
Figure 7.3: Application of the EKF for runout identification and correction on a simulated micro-milling experiment. Cutting starts after $t = 0.5s$. Runout identification is performed up to $t = 0.2s$, correction from $t = 0.2s$ onwards. The top figure shows the displacement signal, the middle the result after runout correction, and the bottom figure the runout correction error. The runout correction error is small compared to the position measurement without runout distortion.
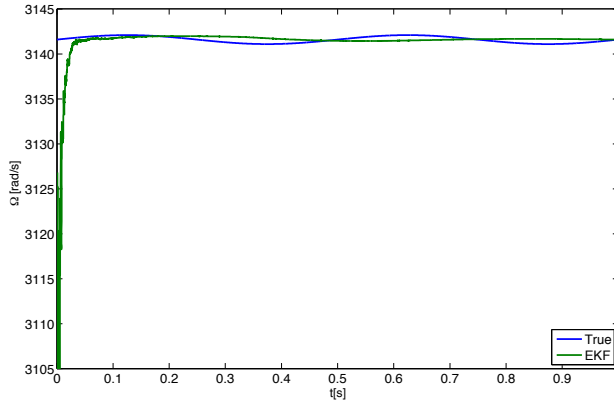
Figure 7.4: EKF applied to runout identification and correction in an simulated AMB spinde signal: comparison of the estimate of rotational speed $\hat{\Omega}(t)$ to the true $\Omega(t)$.

7.6.3 Application to AMB spindle setup, measured data

To conclude the test, the EKF was also applied to real measurement data of a micro-milling cutting experiment performed with the EAAT spindle[4]. The spindle was set to a rotational speed of 40 krpm and the signals of the bearings were recorded at a sampling frequency of 20 kHz. The top plot of Figure 7.5 depicts the displacement signal of the bottom X bearing. As can be observed from this plot, the cutting process started at $t = 0.28$s. The data from $t = 0$ to $t = 0.2$ was used to perform runout identification, after that runout correction was performed. For this the EKF filter was applied with similar settings as the EKF used in the simulated milling experiment of the previous section. The result after runout correction is depicted in the bottom plot of Figure 7.5. To show that the runout is accurately identified, a zoomed in plot of the measured position signal and the estimated runout disturbance is shown in Figure 7.6. This figure shows a high correspondence between a quasi-periodic component in the measured signal and the estimated runout with the EKF.

In order to verify the frequency tracking, the EKF is applied to both the signals from the bottom X and Y bearing independently. The estimated angular frequency from both signals is depicted in Figure 7.7 for $t = 0 \ldots 3$s. From this plot, it is concluded that the EKF is able to track the frequency variations in the AMB spindle signals very accurately (note that the angular frequency estimate obtained from the

[4]Experiment settings: up-milling, $f_z = 0.005$, $d = 1.0$mm, $a_e = 50\mu$m, $a_p = 0.15$mm.
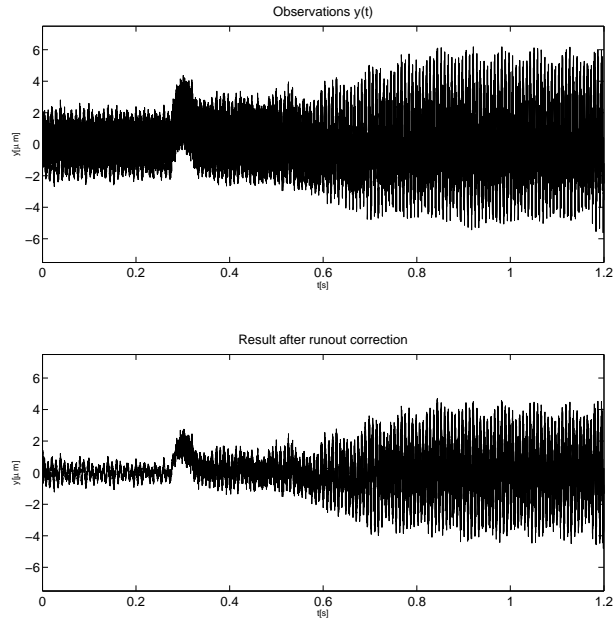
Figure 7.5: Application of the EKF for runout identification and correction on a real micro-milling experiment. Cutting starts after $t = 0.28s$. Runout identification is performed up to $t = 0.2s$, correction from $t = 0.2s$ onwards. The top figure shows the displacement signal, the bottom the result after runout correction.

position measurement in the X-direction responds clearly more to the start of the cutting process than the angular frequency estimate obtained from the measurement in the Y-direction. Still, the difference in this transient period is still very small, only a few rpm).

7.7 Summary and conclusions

Model-based correction of runout disturbances in measurements of the positions and currents of AMB spindle has been considered. Such disturbances are synchronous with the rotation of the spindle and hence almost periodic. A parametrized truncated Fourier series expansion model for the runout disturbance as a function of the angular position is used, allowing to formulate runout identification as a pa-

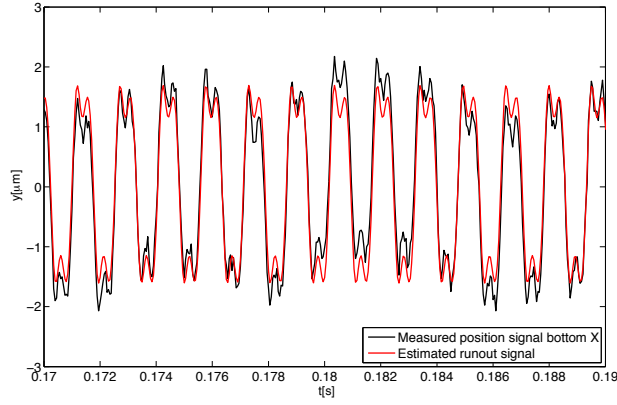7. Compensation of runout disturbances in AMB signals

Figure 7.6: Application of the EKF for runout identification and correction on a real micro-milling experiment: zoom in on the measured displacement signal y (black), and the identified runout disturbance in the signal $\hat{w}$ (red). The EKF appears to have succesfully identified the runout disturbance.

rameter estimation problem.

In correcting for the runout disturbances, the main issue is dealing with the uncertainty in the angular position measurements, or the total lack of such measurements. In this chapter, this issue is attacked from two angles. Considering first the case we have *perfect information* on the angular position, resulting in a standard linear regression scheme, we applied the Bias Compensated Least Squares (BCLS) solution for the case noisy angular position measurement are available. From this angle, the solution follows a strategy to eliminate the bias that would be incurred by naive application of the estimator derived from the assumption the angular position information is noiseless. The other route starts with the assumption that *no information* on the angular position is available. In this solution, the angular position is estimated from the observations using an Extended Kalman Filter (EKF), that for this purpose includes a model of dynamics of angular position. In the case noisy angular positions are available, these can be employed to improve the performance of the angular position estimates, which resulted in the EFK+ approach.

A Monte Carlo simulation study was performed to compare the different approaches for different levels of the SNR in the observations, as well as different levels of the noise in the angular position measurements. This study demonstrated that if no phase information is present, we can still obtain estimates with the EKF that have variance errors that approach the levels that would be attainable with perfect
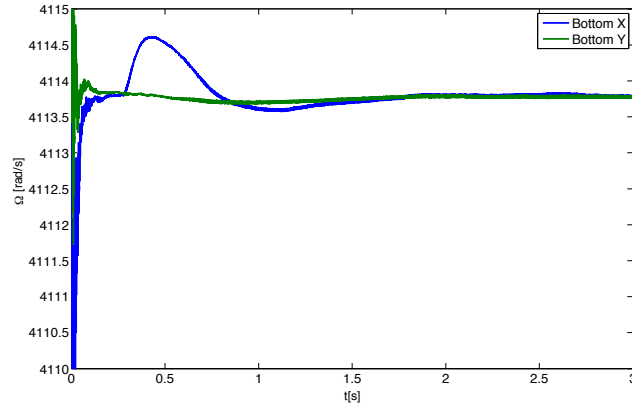
Figure 7.7: Application of the EKF for runout identification and correction on a real micro-milling experiment. Shown are the estimate of rotational speed $\hat{\Omega}(t)$ obtained from the bottom X and bottom Y displacement measurement. The true $\Omega(t)$ is unknown, but the EKF of both directions give almost the same estimate. The start of the milling process at $t = 0.28s$ causes a transient effect in the estimation of the rotational speed.

phase information. If we have angular position information, then the EKF+ that includes this information yields the least variance and would be the estimator of choice. This is understood from the additional a priori information on the dynamics of the angular position, included in this approach. However, when variance of the noise on the angular position measurements is very small, the cost for getting the best performance may not balance out against the increased complexity. In such cases the BCLS might be a more attractive option.

The EKF approach was further demonstrated by applying it to a realistic simulation of an AMB spindle performing a milling experiment, as well as to actual measurement data. In both cases, runout was accurately identified and corrected, and the frequency variations in the signal very closely tracked.
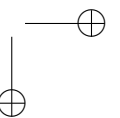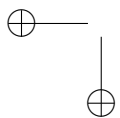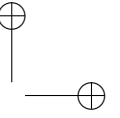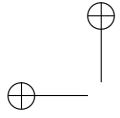
Chapter 8

Conclusions and recommendations

8.1   Conclusions

The research reported in this thesis was aimed at investigating the opportunities for model-based process monitoring and control, in order to improve the micro-milling process using the intrinsic properties of AMB spindles. In the first chapter, it was set out that the approach to this objective was threefold:

1. Provide an approach to model-based cutting force estimation in micro-milling, using the signals of the AMBs;

2. Provide a method for system identification of an high speed AMB micro-milling spindle;

3. Provide a compensation approach for the almost periodic disturbances in the bearing signals, resulting from unbalance forces and roundness errors.

The main conclusions will be given with regard to each of these three aspects, followed by the overall conclusion of the thesis.

Main conclusion A

An approach is developed for optimal model-based estimation of the micro-milling cutting forces from the signals of the AMBs, dealing with the following items:

- *the dynamics of the AMB spindle, including its instable and multivariable character, as well as the increased coupling due to gyroscopy at higher rotational speeds;*

- *the closed-loop configuration of the setup, where knowledge of the controller might not be available;*

- *information on the spectral content of the cutting force signal and the noise disturbance on the bearing signals.*

Main conclusion A follows from chapters 2 and 3, in which it was concluded that:

A.1 The AMB spindle in a micro-milling configuration can be represented using a model of the AMB spindle dynamics and a model of the cutting dynamics, resulting in a configuration with two closed-loops. With a reasonable simplification of this configuration, the cutting force estimation problem can be expressed as an input estimation problem, where the cutting forces are an unknown input to the closed-loop AMB spindle system (chapter 2).

A.2 A minimum mean square error input estimator has been developed to estimate the cutting forces from the signals of the AMBs. When controller knowledge is not available, an additional constraint can be imposed to the optimal estimator, ensuring equal performance of estimator for any controller. Smoothed estimators are derived, allowing to obtain smaller estimation errors when a delay in the estimation result is tolerable (chapter 3).

A.3 The approach was tested in simulation with a realistic model of an experimental setup, showing favorable results. With the commonly used random walk model of the unknown input, an estimator designed without delay still produces delayed estimates. Designing an estimator with the same delay results in significantly better estimates. A filter with just a few time steps extra delay results yields a reduction of error due to the measurement noise of up to 60%. Estimation results can be improved by using the a priori information on the spectral content of the cutting forces (chapter 3).

Main conclusion B

An approach to identifying a linear parametric model of a high-speed AMB micro-milling spindle from measured data sequences is given, in which:

- *the effect of nonlinear distortions on the estimate is quantified, and careful experiment design is performed to obtain small variance errors, as well as small errors from nonlinearities;*

- *accurate black-box models are estimated of the dynamics of the full high-order multivariable AMB system operating at high rotational speeds, including both its rigid body, as well as its flexible body behavior;*

- *both the bearing dynamics (the transfer function from the bearing currents to the displacement of the rotor at the bearings), as well as the tooltip dynamics (the transfer function from the force on the tip of a micro-mill to the displacement of the rotor at the bearings) are modeled.*

Main conclusion B follows from the material in chapters 4, 5 and 6, from which the main conclusions are:

B.1 Iterative linear regression algorithms are given for estimation of output error models in left or right matrix fraction descriptions from frequency response data. These algorithms are extensions of the SISO IV-based linear regression algorithm, which has the property that convergence of the algorithm implies that a stationary point of the cost function is reached. This property, in combination with the freedom in the definition of the model set and the possibility to incorporate pre, post or element-wise multivariable frequency weighting, make this an attractive approach for MIMO frequency domain identification of output error models (chapter 4).

B.2 A frequency domain approach is taken to identify the bearing dynamics of the AMB spindle in the frequency range relevant for process monitoring and control. This approach consists of two stages:

– In the first stage, accurate estimates of the multivariable FRF of the bearing dynamics are made by excitation of the closed-loop with orthogonal random phase multisine signals. A method to detect nonlinear distortions in the joint input/output (JIO) estimate of the FRF of a MIMO system is given, which is based on the variance analysis for detection of nonlinearities of Pintelon et al. in [131]. With this approach, nonlinear distortion in the FRF estimate of an experimental AMB spindle setup have been detected particularly at the lower frequencies. A strategy is formulated to reduce the nonlinear distortion in the JIO FRF estimate, and it has been shown that this resulted in a reduction of the nonlinear distortion of up to 25 dB in the experimental results.

– In the second stage, a parametric model of the AMB spindle is determined using the estimated FRF and its covariance. A Schur-weighted Output Error criterion is formulated, which has been shown to approximate the maximum likelihood cost criterion under mild assumptions. Minimization of the criterion function is achieved by application of the IV-based iterative linear regression method of chapter 4. Multi-band modeling is performed to deal with the large dynamic range. Experimental results show the high correspondence of the estimated parametric model to the estimated FRF data. The resulting estimated models have been verified to hold a structure as predicted from the physics of the AMB spindle.

(chapter 5)

B.3 The main challenge in identifying the tooltip dynamics is applying a known excitation force to the tooltip. An approach has been formulated that uses

data obtained during a milling experiment in which the cutting forces are measured. The amount of data that can be generated in this way is limited, as is the control over the spectral properties of the input. Hence, in order to reduce the complexity of the identification, usage is made of the property of the AMB system that all modes of the AMB spindle in the relevant frequency range are observable in the bearing dynamics. This results in an identification problem with a model set with a particular closed-loop parametrization that is linear in the parameters and a known, but non-minimum phase noise model, for which an approach has been formulated. A simulation study based on the properties of the EAAT AMB spindle has been used to demonstrate the favorable properties of the given method (chapter 6).

Main conclusion C

Effective model-based compensation approaches have been given for almost periodic disturbances in the bearing signals, resulting from unbalance forces and roundness errors.

This conclusion follows from chapter 7:

C.1 A Model-based correction of runout disturbances in measurements of the positions and currents of an AMB spindle has been considered, using a parametrized truncated Fourier series expansion model for the runout disturbance as a function of the angular position. The main issue in this approach is how to deal with the uncertainty in the angular position measurements, or the total lack of such measurements. This issue is attacked from two angles:

– In case *perfect information* on the angular position is available, a standard linear regression scheme is obtained. From this, a bias compensation approach has been formulated for the case noisy angular position measurement are available.

– When *no information* on the angular position is available, the angular position is estimated from the observations using an Extended Kalman Filter (EKF), that for this purpose includes a model of dynamics of angular position. In the case noisy angular positions are available, these can be employed to improve the performance of the angular position estimates, resulting in the EFK+ approach.

C.2 The properties of both estimators have been investigated using a Monte Carlo simulation study, demonstrating that in the case of noisy angular position measurements, the EFK+ estimation performs best. The EKF approach has been further demonstrated by applying it both to data obtained through simulation of an AMB spindle performing a milling experiment, as well as to measurement data from actual experiments. In both cases, runout has been

accurately identified and corrected, and the frequency variations in the signal are very closely tracked.

Overall conclusion

The objective of the research in this thesis is to investigate the possibilities for performing process monitoring and control of micro-milling using AMBs. In Chapter 1, it was argued that it is an essential step to investigate the feasibility of obtaining cutting forces from the bearing signals. As reflected in conclusions A, B, and C, this thesis provides new methods to accomplish cutting force estimation using AMB spindles, where the following model-based steps are applied:

A. A model-based approach to cutting force estimation from the bearing signals is given;

B. An approach is given to obtain highly accurate models of the AMB spindle dynamics from measured data, needed for such an estimator;

C. A model-based approach is given to compensate for almost periodic disturbances present in the measurements of the bearing signals, which are the input for the force estimator.

With this methodology, cutting force estimation of micro-milling forces from AMB signals can be performed, opening up several other process monitoring and control applications. Concluding, the research goal of this thesis has been achieved.

8.2 Recommendations for future research

Based on the conclusions, some improvements and recommendations for further research can be formulated.

- The methods in this thesis have been validated individually, some only in simulation by lack of an experimental setup able to produce reliable data. It is recommended that end-to-end testing of the approach is performed to validate it.

- In Chapter 1, it was discussed that several opportunities exist to improve the micro-milling process by developing process monitoring and control using the active character of AMB spindles. In this thesis a first step has been taken. It is recommended to perform more research to develop model-based technogies that exploit the mentioned opportunities, including tool wear estimation, breaking detection, force control, and chatter control. In contrast to the cutting force estimation problem, for these applications models of the cutting dynamics will be needed. Research is required to estabish what models will be appropriate in these cases.

8. Conclusions and recommendations

- Adding to the previous recommendation, Chapter 1 also identified several opportunites to improve the cutting process with AMB spindles by improving the contour control. Particularly interesting is the opportunity to perform bending compensation. The objective of reducing the bending of a flexible micro-milling tool can be formulated as to design a controller such, that the transfer function from cutting force on the tooltip (u_2) to displacement of the tooltip (y_2) is made small (in some appropriate measure). Such a control objective can be very well formulated in an H_∞ control design framework. The effect of such a control strategy is that the controller compensates the high compliance at the tooltip by controlling the orientation and position of the rotor. The attractiveness of such technique is that the effective stiffness, as experienced by the process, is increased. In conventional manufacturing, some results are available on this method, showing favorable results for the cutting process (see [90, 51]).

- The input estimation approach in chapter 3 has some limitations. In this approach perfect model knowledge is assumed, as well as linearity of the plant. When these assumptions cannot be justified, the approach is to be extended to incorporate model uncertainty and/or nonlinearity in order to obtain smaller errors. This requires further research.

- More work is needed to establish the relation between the physical properties of an AMB spindle and the limits of performance of an input estimator used to estimate the cutting forces. This knowledge will be very valuable in the design process of an AMB spindle, when it is specifically targeted at a micro-milling configuration.

- Concerning the IV-based iterative linear regression approach to frequency domain identification of MIMO system, the following recommendations are made. It has been shown that convergence of the iterations implies optimality, however, more work is needed to establish the convergence properties of the algorithm. Furthermore, the algorithm suffers from poor numerical conditioning for high-order models, particularly if the data extends over a large dynamic range and a large frequency range. For the iterative linear regression method by Sanathanan and Koerner, improved numerical conditioning has been achieved by computing an alternative basis for the model (with orthogonal polynomials, see e.g. [137]). It is suggested to extend this work to the IV-based iterative linear regression approach.

- Regarding the identification approach of the AMB spindle in Chapter 5, the following problems are still open:

  – Parametric modeling of the plant dynamics has been performed, but only non-parametric noise modeling was done. In order to improve the accuracy of the cutting force estimator in Chapter 3, which incorporates

both a plant and noise model, it is suggested to extend the identification method to also perform parametric noise modeling.

– Linear models are obtained for a fixed rotational speed of the spindle. This is a limitation, as it implies that for every rotational speed setting the identification procedure needs to be repeated. It is therefore recommended to enhance the method to identification of the AMB spindle for a range of spindle speeds. This leads to a closed-loop MIMO LPV identification problem.

– An approach has been given to identify the level of nonlinear disturbance in the multivariable frequency response function of the AMB spindle. For particular spindle designs, this might lead to the conclusion that this level is too high to justify approximation of the dynamics with a linear model. For such cases, extensions are needed to identify the nonlinearities as part of the model.

• The approach to identify the tooltip dynamics assumes that direct cutting force measurements are available. In practice, these measurements are obtained with devices that exhibit dynamical behavior: dynamometers have strong resonances. The effect of these dynamic properties is that only a filtered version of the cutting force signal is available. Further research is required to establish the consequences of this phenomenon.

• The runout compensation approaches are described for scalar signals. In the Extended Kalman filter approach, the angular position of the rotor is estimated from the measurement. In an AMB spindle application however, the angular position is the same for all measured signals. Hence, instead of estimating it from each measurement separately, a natural extension is to perform runout compensation for vector-valued signals, allowing for a more accurate estimation of the angular position.
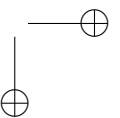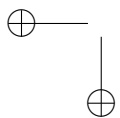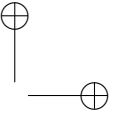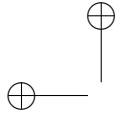
Appendix A

Modeling of an AMB spindle

This appendix shows the in-depth analysis of a number of modeling steps in chapter 2. It starts with first principles modeling of the dynamics of the rotor and the electromagnetic actuator. Both have been extensively studied in the literature (e.g. see [149, 148, 91, 94, 110] and the references therein). Based on these works, we give a summary of the modeling procedure that was adopted for this thesis. In section A.1 modeling of the flexible rotor will be discussed, followed by modeling of a simple configuration of the electromagnetic actuator in section A.2. Section A.3 discusses how discrete-time models from the continuous-time models can be obtained. This is done for the two measurement configurations discussed in chapter 2, i.e. the direct measurement configuration and the integrated configuration.

A.1 Flexible rotor model

In this context we consider the rotor and the cutting tool that is attached to the rotor as one rotating body. Given the extended frequency region of interest, we can not allow treating this as a rigid body and need to model this as a flexible system. Flexible systems are in principle infinite dimensional systems. Practical solutions to model this are generally found by approximating the dynamics with a limited number of dimensions, leading to a finite element approach. Here we will also pursue such approach by modeling the rotor and tool as a chain of flexible beam elements. These beam elements have a given length and diameter, for which the equations of motion are derived in section A.1.1. Using the beam elements as building blocks, the procedure to construct the equations of motion for the full rotor will be discussed in section A.1.2. We conclude by discussing how distributed damping can be added to the model in section A.1.3.

A.1.1 The beam element

The beam element is modeled as a flexible body with a constant circular cross-section and length L (see figure A.1). It has two nodes, with four degrees of freedom each (displacement in x and y direction and rotations along these axes, denoted by ϕ and θ respectively). Hence
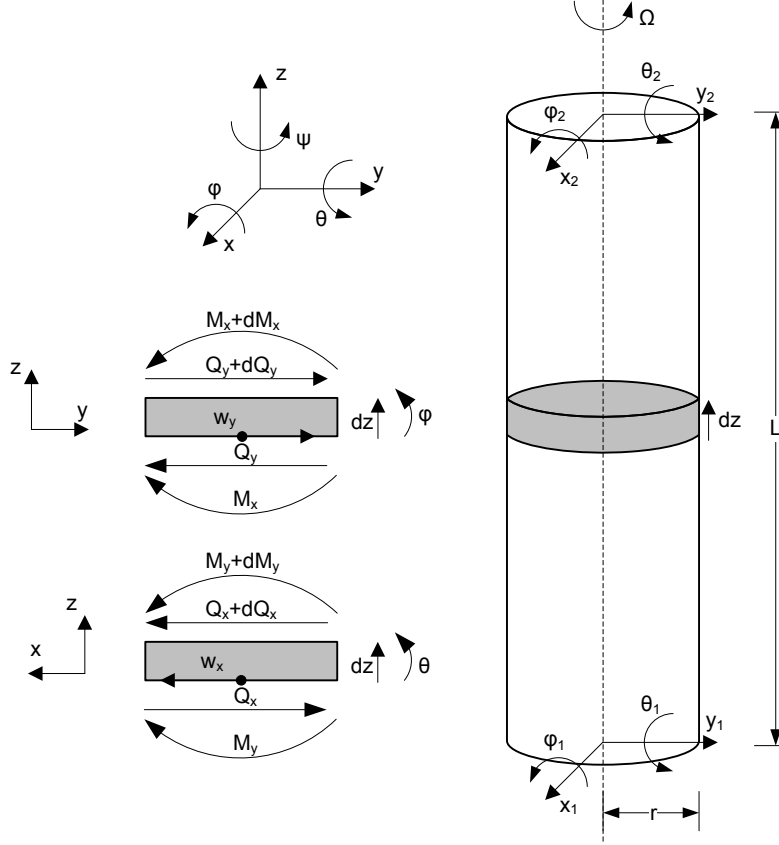
Figure A.1: The beam element with definitions of directions and coordinates

we obtain the vector of generalized coordinates

$$\mathbf{w}_b = \begin{bmatrix} x_1 & y_1 & \phi_1 & \theta_1 & x_2 & y_2 & \phi_2 & \theta_2 \end{bmatrix}^T$$

The corresponding vector of external generalized forces is

$$\mathbf{F}_{\text{ext}} = \begin{bmatrix} F_{x1} & F_{y1} & M_{\phi1} & M_{\theta1} & F_{x2} & F_{y2} & M_{\phi2} & M_{\theta2} \end{bmatrix}$$

The rotational speed of the beam element around its z-axis is assumed to be constant. We will denote the rotation angle with ψ and the rotational speed by $\Omega = \dot{\phi}$. The angles ϕ, θ and ψ are defined as Euler angles, meaning that the transformation from world coordinate system to body coordinate system is obtained by first performing a rotation of ϕ radians around the x-axis, then θ radians around the y axis and finally ϕ radians around the z axis. The material

properties are described by mass density ρ, modulus of elasticy E, shear modulus G, which is related to E by $G = E/[2(1 + \nu)]$ where ν is Poisson's ratio.

We will use an energy method to derive the equation of motion for such beam element. This implies that we take an infinitesimal cross section, and calculate the potential and kinetic energies in this section. Then a shape interpolation function based on the nodal coordinates $\mathbf{w}_b$ is determined and this is used to integrate the energies along the beam. Subsequently, the theorem of Lagrange is applied to find the equations of motion of the beam in terms of the above coordinates and forces.

We will start with deriving the differential equations for a small cross section. Applying the definitions in figure A.1, from Newton's laws, it follows that

$$dQ_x = \mu \frac{d^2}{dt^2} w_x dz \qquad\qquad dM_y + Q_x dz = \breve{\mu} \frac{d^2}{dt^2} \theta dz$$

$$dQ_y = \mu \frac{d^2}{dt^2} w_y dz \qquad\qquad dM_x - Q_y dz = \breve{\mu} \frac{d^2}{dt^2} \phi dz$$

with μ the mass per unit length and $\breve{\mu}$ the rotatory mass per unit length. Now defining $u_x = -\sin\theta x \approx -\theta x$ and $u_y = \sin\phi x \approx \phi x$, we can calculate the axial strains

$$\epsilon_x = \frac{\partial u_x}{\partial z} = -\theta' x \qquad\qquad \epsilon_y = \frac{\partial u_y}{\partial z} = \phi' y.$$

Assuming $\sigma = E\epsilon$ and noting that

$$M_y = -\int \sigma_x x dA$$

we obtain

$$M_y = \int E\theta' x^2 dA = EI_x \theta'.$$

Similarly, we find that

$$M_x = \int \sigma_y y dA = \int E\phi' y^2 dA = EI_y \phi'.$$

For the average shear strain we can write

$$\gamma_x = \frac{\partial u_x}{\partial x} + \frac{\partial w_x}{\partial z} = -\theta + w'_x \qquad\qquad \gamma_y = \frac{\partial u_y}{\partial y} + \frac{\partial w_y}{\partial z} = \phi + w'_y.$$

With shear relation $\tau = G\gamma$ we obtain that

$$Q_x = -\int \tau_x dA = GA(\theta - w'_x) \qquad\qquad Q_y = -\int \tau_y dA = GA(-\phi - w'_y).$$

However, as was shown by Timoshenko [55], this relation is not quite accurate, which can be corrected by introducing the *effective shear area* A^s which for a circular cross section is $0.89A$. Since the cross section is circular, we can write $I = I_x = I_y$, and obtain the following set of equations

$$-Q'_x = \mu \ddot{w}_x \qquad\qquad Q_x - M'_y = \breve{\mu} \ddot{\theta}$$

$$-Q'_y = \mu \ddot{w}_y \qquad\qquad -Q_y - M'_x = \breve{\mu} \ddot{\phi}$$

$$Q_x = GA^s(\theta - w'_x) \qquad\qquad M_y = EI\theta'$$

$$Q_y = GA^s(-\phi - w'_y) \qquad\qquad M_x = EI\phi'.$$

A. Modeling of an AMB spindle

To proceed, we need to find a relationship between the nodal coordinates $\mathbf{w}_b$ and the coordinates w_x, w_y, ϕ and θ along the beam element. Instead of solving the differential equations directly, in general a polynomial shape function of the beam is chosen that is uniquely determined by $\mathbf{w}_b$. If we ignore the shear strain and the rotory intertia in the directions θ and ϕ, there would exist a relationship between w_x and θ (i.e. $\theta = w_x'$) and likewise w_y and ϕ (i.e. $\phi = -w_y'$)[1]. In this case we obtain a fourth order bending problem in both x and y direction, while $\mathbf{w}_b$ has four degrees of freedom in both x and y direction, hence the interpolation problem can be solved. However, when the shear strain and the rotory intertia cannot be ignored, a difficulty we encounter is that w_x and θ, and likewise w_y and ϕ are in fact kinematically independent of one another. This results in a fourth order bending and a second order shear interpolation in both directions, so that too few boundary conditions are available to uniquely solve the interpolation problem. A number of solutions have been formulated[110]:

Second order polynomials In this approach the variables are indeed treated completely separately and independent polynomials are chosen for w_x, w_y, ϕ and θ. The result of solving the equations is that bending deformation is represented as a linear function, and shear as a second order. This result is quite non-satisfactory, particularly concerning the representation of the bending deformation.

Splitting assumption Higher order polyniomals are obtained in this approach by increasing the number of boundary conditions. It is assumed that the bending deformation can be 'split off' from the shear deformation by writing $w_x = w_{x1} + w_{x2}$ with $w_{x1}' = \theta$ and $w_{x2}' = \gamma_x$ and similarly for the other direction $w_y = w_{y1} + w_{y2}$ with $w_{y1}' = -\phi$ and $w_{y2}' = \gamma_y$. With this assumption, the following set of equations is obtained

$$-Q_x' = \mu(\ddot{w}_{x1} + \ddot{w}_{x2}) \qquad\qquad Q_x - M_y' = \breve{\mu}\ddot{w}_{x1}'$$
$$-Q_y' = \mu(\ddot{w}_{y1} + \ddot{w}_{y2}) \qquad\qquad -Q_y - M_x' = -\breve{\mu}\ddot{w}_{y1}'$$
$$Q_x = -GA^s w_{x2}' \qquad\qquad M_y = EIw_{x1}''$$
$$Q_y = -GA^s w_{y2}' \qquad\qquad M_x = EIw_{y1}''$$

Eliminating the forces and moments from these equations, we get

$$GA^s w_{x2}'' - \mu(\ddot{w}_{x1} + \ddot{w}_{x2}) = 0$$
$$GA^s w_{y2}'' - \mu(\ddot{w}_{y1} + \ddot{w}_{y2}) = 0$$
$$EIw_{x1}'''' + \breve{\mu}\ddot{w}_{x1}'' + \mu(\ddot{w}_{x1} + \ddot{w}_{x2}) = 0$$
$$EIw_{y1}'''' - \breve{\mu}\ddot{w}_{y1}'' - \mu(\ddot{w}_{y1} + \ddot{w}_{y2}) = 0$$

These equations show two fourth-order equations and two second order equations with a coupled mass term. As shown by Marguerre and Wölfel [110], by separately parameterizing w_{x1} and w_{x2}, the first with a cubic polynomial, the second with a second order polynomial (and similarly for the y direction), equations of motion can be obtained. The corresponding boundary vectors for displacement and force will have twelve elements each, instead of eight. This approach is useful particularly when dealing with simple chain structure that have no interaction with external structures. Hence most useful for our application will be the third option:

---

[1] This is would result in a Bernoulli beam analysis.

Mixed assumption Like with the Bernoulli beam, in this case it will be assumed that there exists a relationship with w_x and θ and w_y, and w_y and ϕ. However, this relationship is derived from the beam equations. If for the purpose of deriving this relationship, we ignore the rotory inertia[2], i.e. we assume $\ddot{\psi} = \ddot{\theta} = 0$, we obtain that

$$w_x' = \theta - \frac{EI}{GA^s}\theta'' \qquad\qquad w_y' = -\phi - \frac{EI}{GA^s}\phi''$$

With this assumption, we can proceed to derive the equations of motion for the the beam element.

If we define coordinate vector $\mathbf{w}(z,t) = \begin{bmatrix} w_x & w_y & \phi & \theta \end{bmatrix}^T (z,t)$ and boundary coordinate vector $\mathbf{w}_b(t) = \begin{bmatrix} \mathbf{w}(0,t) & \mathbf{w}(L,t) \end{bmatrix}$, then as shown by Marguerre and Wölfel [110], we may write

$$w_x(z,t) = H_x(z)\mathbf{w}_b(t) \qquad\qquad \phi(z,t) = H_\phi(z)\mathbf{w}_b(t) \qquad (A.1)$$
$$w_y(z,t) = H_y(z)\mathbf{w}_b(t) \qquad\qquad \theta(z,t) = H_\theta(z)\mathbf{w}_b(t) \qquad (A.2)$$

with

$$H_x(z) = \begin{bmatrix} \frac{1}{1+\kappa}\left(2(z/L)^3 - 3(z/L)^2 - \kappa(z/L) + (1+\kappa)\right) \\ 0 \\ 0 \\ \frac{L}{1+\kappa}\left((z/L)^3 - (2+\kappa/2)(z/L)^2 + (1+\kappa/2)(z/L)\right) \\ \frac{1}{1+\kappa}\left(-2(z/L)^3 + 3(z/L)^2 + \kappa(z/L)\right) \\ 0 \\ 0 \\ \frac{L}{1+\kappa}\left((z/L)^3 - (1-\kappa/2)(z/L)^2 + (\kappa/2)(z/L)\right) \end{bmatrix}^T$$

and

$$H_\theta(z) = \begin{bmatrix} \frac{6/L}{1+\kappa}\left((z/L)^2 - (z/L)\right) \\ 0 \\ 0 \\ \frac{1}{1+\kappa}\left(3(z/L)^2 - (4+\kappa)(z/L) + (1+\kappa)\right) \\ \frac{6/L}{1+\kappa}\left(-(z/L)^2 + (z/L)\right) \\ 0 \\ 0 \\ \frac{1}{1+\kappa}\left(3(z/L)^2 - (2-\kappa)(z/L)\right) \end{bmatrix}^T$$

with $\kappa = 12EI/(L^2 GA^s)$. Similar expressions for H_y, H_θ can be derived.

For the strain energy (i.e. the potential energy) we can write:

$$dV = \frac{1}{2}\left(\; * \; \right)^T \begin{bmatrix} EI & & & \\ & EI & & \\ & & GA^s & \\ & & & GA^s \end{bmatrix} \begin{pmatrix} \theta' \\ \phi' \\ (\theta - w_x') \\ (-\phi - w_y') \end{pmatrix} dz$$

---

[2] This approximation explains why approach is called the mixed assumption, since for the derivation of the dynamics, an approximation is made based on the static behavior.

which we can also write as

$$dV = \frac{1}{2}\mathbf{w}_b^T T^T \begin{bmatrix} EI & & & \\ & EI & & \\ & & GA^s & \\ & & & GA^s \end{bmatrix} T\mathbf{w}_b dz, \text{ where } T = \begin{bmatrix} H'_\theta \\ H'_\phi \\ -H'_x + H_\theta \\ -H'_y - H_\phi \end{bmatrix}$$

or, for brevity, $dV = \frac{1}{2}\mathbf{w}_b^T U\mathbf{w}_b$ with

$$U = T^T \begin{bmatrix} EI & & & \\ & EI & & \\ & & GA^s & \\ & & & GA^s \end{bmatrix} T.$$

To proceed with deriving the kinetic energy, let us define $\mathbf{x} = \begin{bmatrix} x & y & z & p & q & r \end{bmatrix}$ the vector of coordinates and angular displacements in the body reference frame (i.e. the reference frame attached to the disc). The kinetic energy of the disc with width dz we can express as

$$dT = \frac{1}{2}\dot{\mathbf{x}}^T \begin{bmatrix} \mathbf{M} & 0 \\ 0 & \mathbf{I} \end{bmatrix} \dot{\mathbf{x}} dz$$

For a circular disc with radius r, mass density ρ and with its normal aligned with the z-axis, we can write

$$\mathbf{M} = \text{diag}\begin{pmatrix} \mu & \mu & \mu \end{pmatrix} \qquad\qquad \mathbf{I} = \text{diag}\begin{pmatrix} \breve{\mu} & \breve{\mu} & 2\breve{\mu} \end{pmatrix}$$

with $\mu = \rho\pi r^2$ and $\breve{\mu} = \frac{1}{4}\rho\pi r^4$, where we have used that for this geometry the area moments of inertia $I_x = I_y = \frac{1}{4}r^4$ and $I_z = 2I_y$.

Let R be the matrix that maps the Euler angles $\begin{bmatrix} \phi & \theta & \psi \end{bmatrix}^T$ in the frame attached to the rotor. Then writing the vector of world coordinates as $\mathbf{w}_a = \begin{bmatrix} w_x & w_y & w_z & \phi & \theta & \psi \end{bmatrix}$ we can derive the coordinate transformation

$$\mathbf{x} = \begin{bmatrix} I & 0 \\ 0 & R \end{bmatrix} \mathbf{w}_a$$

This transformation we can use to express the kinetic energy in terms of world coordinates $\mathbf{w_a}$

$$dT = \frac{1}{2}\dot{\mathbf{w}}_a^T \begin{bmatrix} \mathbf{M} & 0 \\ 0 & R^T\mathbf{I}R \end{bmatrix} \dot{\mathbf{w}}_a dz$$

As can easily be verified, R is given by

$$R = \begin{bmatrix} \cos\theta\cos\psi & -\sin\psi & 0 \\ \cos\theta\sin\psi & \cos\psi & 0 \\ \sin\theta & 0 & 1 \end{bmatrix}$$

With this we obtain

$$
R^T \mathbf{I} R = \begin{bmatrix} \breve{\mu}(1 + \sin^2 \theta) & 0 & 2\mu \sin \theta \\ 0 & \breve{\mu} & 0 \\ 2\mu \sin \theta & 0 & 2\mu \end{bmatrix}
$$

which for small θ reduces to

$$
R^T I R = \begin{bmatrix} \breve{\mu} & 0 & 2\mu\theta \\ 0 & \breve{\mu} & 0 \\ 2\mu\theta & 0 & 2\breve{\mu} \end{bmatrix}
$$

Now observe that

$$
\dot{\mathbf{w}}_a^T \begin{bmatrix} \mathbf{M} & 0 \\ 0 & R^T \mathbf{I} R \end{bmatrix} \dot{\mathbf{w}}_a = \dot{\mathbf{w}}_a^T \begin{bmatrix} \mathbf{M} & 0 \\ 0 & \mathbf{I} \end{bmatrix} \dot{\mathbf{w}}_a + 4\breve{\mu}\phi\dot{\theta}\dot{\psi}.
$$

Noting that we do not consider translation in the z-direction (hence $\dot{w}_z = 0$) and the rotational speed is constant (hence $\dot{\psi} = \Omega$), we can derive that

$$
\mathbf{w}_a(z, t) = H(z)\mathbf{w}_b(t) + \mathrm{col}(0, 0, w_z, 0, 0, \psi)
$$

with $H(z) = \mathrm{col}(H_x, H_y, 0, H_\phi, H_\theta, 0)$, and consequently

$$
\dot{\mathbf{w}}_a(z, t) = H(z)\dot{\mathbf{w}}_b(t) + \mathrm{col}(0, 0, 0, 0, 0, \Omega).
$$

Using this, we obtain

$$
\dot{\mathbf{w}}_a^T \begin{bmatrix} \mathbf{M} & 0 \\ 0 & R^T \mathbf{I} R \end{bmatrix} \dot{\mathbf{w}}_a = \dot{\mathbf{w}}_b^T H^T \begin{bmatrix} \mathbf{M} & 0 \\ 0 & \mathbf{I} \end{bmatrix} H\dot{\mathbf{w}}_b + 4\breve{\mu}\Omega\mathbf{w}_b^T H_\phi^T H_\theta \dot{\mathbf{w}}_b + 2\breve{\mu}\Omega^2
$$

Integrating the potential and kinetic energy over the length of the beam, we can construct the Lagrangian and obtain

$$
L = T - V = \dot{\mathbf{w}}_b^T \left\{ \frac{1}{2} \int_0^L H^T \begin{bmatrix} \mathbf{M} & 0 \\ 0 & \mathbf{I} \end{bmatrix} H dz \right\} \dot{\mathbf{w}}_b(t) + \mathbf{w}_b^T \left\{ 2\breve{\mu}\Omega \int_0^L H_\phi^T H_\theta dz \right\} \dot{\mathbf{w}}_b^T(t) +
$$

$$
- \mathbf{w}_b(t)^T \left\{ \frac{1}{2} \int_0^L U dz \right\} \mathbf{w}_b(t) + 2\breve{\mu}\Omega^2
$$

Now applying Lagrange's theorem:

$$
\frac{d}{dt} \left(\frac{\partial T}{\partial \dot{\mathbf{w}}_b} \right) - \frac{\partial T}{\partial \mathbf{w}_b} = M\ddot{\mathbf{w}}_b + G\dot{\mathbf{w}}_b + K\mathbf{w}_b = \mathbf{F}_{ext}
$$

where from the symmetry properties of the integrands it follows that

$$
M = \int_0^L H^T \begin{bmatrix} \mathbf{M} & 0 \\ 0 & \mathbf{I} \end{bmatrix} H dz
$$

$$
K = \int_0^L U dz
$$

and furthermore

$$
G = 2\breve{\mu}\Omega \int_0^L H_\phi^T H_\theta - H_\theta^T H_\phi dz.
$$

A.1.2 Connection of beam elements

A finite element model of the rotor can be obtained by modeling it as a finite series of connected beam elements. Assuming rigid connection between the elements, the connection of two elements can be modeled as follows. Consider beam element A of which its top node is rigidly connnected to the bottom node of beam element B. Let $\mathbf{w}_i = \begin{bmatrix} x_i & y_i & \phi_i & \theta_i \end{bmatrix}^T$, with $i = 1, 2$ denote the generalized coordinates of the top and bottom coordinate of a beam element respectively . Then the rigid connection of beam element A and B can be expressed mathematically as $\mathbf{w}_2^A = \mathbf{w}_1^B$. We use the following notation, indication the partioning of the element matrices with respect to the top and bottom nodal coordinates:

$$\mathbf{M}^A = \begin{bmatrix} M_{11}^A & M_{12}^A \\ M_{21}^A & M_{22}^A \end{bmatrix}, \qquad \mathbf{G}^A = \begin{bmatrix} G_{11}^A & G_{12}^A \\ G_{21}^A & G_{22}^A \end{bmatrix}, \qquad \mathbf{K}^A = \begin{bmatrix} K_{11}^A & K_{12}^A \\ K_{21}^A & K_{22}^A \end{bmatrix}, \qquad (A.3)$$

$$\mathbf{M}^B = \begin{bmatrix} M_{11}^B & M_{12}^B \\ M_{21}^B & M_{22}^B \end{bmatrix}, \qquad \mathbf{G}^B = \begin{bmatrix} G_{11}^B & G_{12}^B \\ G_{21}^B & G_{22}^B \end{bmatrix}, \qquad \mathbf{K}^B = \begin{bmatrix} K_{11}^B & K_{12}^B \\ K_{21}^B & K_{22}^B \end{bmatrix}. \qquad (A.4)$$

Defining the extended coordinate vector

$$\mathbf{w} = \begin{bmatrix} \mathbf{w}_1^A \\ \mathbf{w}_2^A = \mathbf{w}_1^B \\ \mathbf{w}_2^B \end{bmatrix}$$

it can easily be verified that the connected system can be described as

$$\mathbf{M}\ddot{\mathbf{w}} + \mathbf{G}\dot{\mathbf{w}} + \mathbf{K}\mathbf{w} = \mathbf{F}_{\text{ext}}$$

with

$$\mathbf{M} = \begin{bmatrix} M_{11}^A & M_{12}^A & 0 \\ M_{21}^A & M_{22}^A + M_{11}^B & M_{12}^B \\ 0 & M_{21}^B & M_{22}^B \end{bmatrix}, \qquad \mathbf{G} = \begin{bmatrix} G_{11}^A & G_{12}^A & 0 \\ G_{21}^A & G_{22}^A + G_{11}^B & G_{12}^B \\ 0 & G_{21}^B & G_{22}^B \end{bmatrix},$$

$$\mathbf{K} = \begin{bmatrix} K_{11}^A & K_{12}^A & 0 \\ K_{21}^A & K_{22}^A + K_{11}^B & K_{12}^B \\ 0 & K_{21}^B & K_{22}^B \end{bmatrix}$$

and $\mathbf{F}_{\text{ext}}$ the vector of external forces and moments corresponding to each of the degrees of freedom. The above procedure can be extended for longer connections of beam elements.

If the joint between elements is not rigid (as might be the case in the coupling between milling tool and rotor), radially stiff joints or flexible joints with stiffness and damping can be included as described by Krämer[91].

A.1.3 Adding damping

Viscous damping can be added to the model by adding a matrix $\mathbf{D}$ to the equations of motion, yielding

$$\mathbf{M}\ddot{\mathbf{w}} + (\mathbf{D} + \mathbf{G})\dot{\mathbf{w}} + \mathbf{K}\mathbf{w} = \mathbf{F}_{\text{ext}}.$$

It is particularly difficult to model distributed (c.q. material) damping from first principles. Hence, in the literature, frequently approximations are made, where Rayleigh damping and
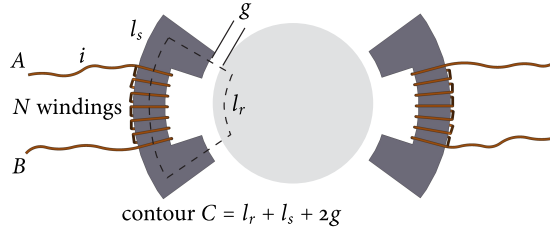
Figure A.2: Configuration of the electromagnetic actuator

modal damping are most common (see e.g.[91]). Rayleigh involves setting the damping matrix as a linear combination of the mass and stiffness matrix, i.e. $\mathbf{D} = \alpha\mathbf{M} + \beta\mathbf{K}$, where α and β are chosen such that damping of the two most important modes is appropriately modeled. Alternatively, modal damping can be applied, which involves computation of a modal decomposition of the system (for $\Omega = 0$). Then for each mode a damping value is selected. With this, matrix D can be computed.

A.2 The electromagnetic actuator

In this section, a model will be derived for the electromagnetic actuator. For this, we will consider the most common configuration of a magnetic bearing, *the differential driving mode*, which consists of two electromagnetic actuators that are positioned at opposite sides of the rotor. Furthermore, in order to keep the derivation simple, we will consider a design of actuator with independent flux paths of both actuators.

In section A.2.1 a model for a single actuator will be derived. This model will be used in section A.2.2 to derive the equations for a magenetic bearing in differential driving mode.

A.2.1 The electromagnetic actuator

Consider the actuator depicted in figure A.2, showing a rotor, with on opposite sides an electromagnetic coil. The purpose of this section is to derive the electrical and electromechanical properties of one coil.

Laws of electromagnetism

In order to analyze the actuator, we will start by briefly reviewing Maxwell's equations, where we will use the integral form [33]:

- Maxwell's first law

$$\oint_C \mathbf{H} \cdot d\mathbf{l} = \int_S \mathbf{J} \cdot d\mathbf{A} + \frac{d}{dt} \int_S \mathbf{D} \cdot d\mathbf{A}$$

A. Modeling of an AMB spindle

- Maxwell's second law or Faraday's law of induction

$$\oint_C \mathbf{E} \cdot d\mathbf{l} = -\frac{d}{dt} \int_S \mathbf{B} \cdot d\mathbf{A}$$

- Gauss' law for magnetism

$$\oint_S \mathbf{D} \cdot d\mathbf{A} = \int_V \rho dV$$

- Gauss' law

$$\oint_S \mathbf{B} \cdot d\mathbf{A} = 0.$$

Here $\mathbf{E}$ is the electric field $[V/m]$, $\mathbf{H}$ is the magnetic field $[A/m]$, $\mathbf{D}$ is the electric displacement field (also known as electric flux density) $[C/m^2]$, $\mathbf{B}$ is the magnetic flux density $[T]$, $\mathbf{J}$ is the free current density $[A/m^2]$, and ρ is the free electric charge density $[C/m^3]$. The magnetic flux Φ through surface S is defined as

$$\Phi = \int_S \mathbf{B} \cdot d\mathbf{A}$$

In the analysis that follows we will use two common and reasonable simplifications:

- Linear materials: the relation between magnetic field and magnetic flux density is assumed to be known and linear, i.e. $\mathbf{B} = \mu\mathbf{H} = \mu_0\mu_r\mathbf{H}$, with μ the (absolute) permeability, μ_0 the permability of vaccuum and μ_r the relative permeability.
- Quasi-static field approximation: relative to the propagation speed of electromagnetic fields (the speed of light), the distances and translation speeds in the systems are small, hence changes of electric displacement field can be neglected as source for the magnetic field. With this, Maxwell's first equation reduces to

$$\oint_C \mathbf{H} \cdot d\mathbf{l} = \int_S \mathbf{J} \cdot d\mathbf{A}. \tag{A.5}$$

Magnetic circuit

Considering a current carrying wire wound N times around a ferromagnetic core, it follows from equation (A.5):

$$\oint_C \mathbf{H} \cdot d\mathbf{l} = \int_S \mathbf{J} \cdot d\mathbf{A} = Ni$$

Contour C is built up in four segments (stator part, rotor part and two air gaps). Assuming the resulting magnetic field density is constant and always tangential to the coil core, we may simplify this to $H_s l_s + H_r l_r + 2H_g g = Ni$. Assuming the permeability of the core material is sufficiently high, the field will concentrate in the magnetic circuits, and Φ may be considered constant. This implies no spread of field in air gap and no leakage flux induced in air around coil. From this it follows that $B_r = \frac{\Phi}{A_r}$, $B_s = \frac{\Phi}{A_s}$, $B_g = \frac{\Phi}{A_g}$. With $B = \mu H$ and assuming that the relative permeability of the air gap is 1 we obtain

$$Ni = \Phi \left(\frac{l_r}{\mu_0 \mu_{r_r} A_r} + \frac{l_s}{\mu_0 \mu_{r_s} A_s} + 2\frac{g}{\mu_0 A_g} \right)$$

and thus

$$\Phi = \frac{N}{\left(\frac{l_r}{\mu_0 \mu_{r_r} A_r} + \frac{l_s}{\mu_0 \mu_{r_s} A_s} + 2\frac{g}{\mu_0 A_g} \right)} i. \tag{A.6}$$

Electrical properties

To apply Faraday's law of induction, first note that electrical current I is defined as flowing from A to B. Hence the contour integral of electric field can be expanded as

$$\oint_C \mathbf{E} \cdot d\mathbf{l} = \int_A^B \mathbf{E} \cdot d\mathbf{l} + \int_B^A \mathbf{E} \cdot d\mathbf{l}$$

$$= \int_A^B \rho \mathbf{J} \cdot d\mathbf{l} - u$$

$$= \int_A^B \rho \frac{i}{A_w} \cdot d\mathbf{l} - u$$

$$= Ri - u$$

with $R = \frac{\rho l_w}{A_w}$, and ρ the electrical resistivity, A_w the cross section of the wire and l_w the length of the wire. Noting that

$$-\frac{d}{dt} \int_S \mathbf{B} \cdot d\mathbf{A} = -\frac{d}{dt}\psi$$

with ψ the flux coupled with contour C. Combining the above results we obtain

$$-\frac{d}{dt}\psi = Ri - u.$$

Note that coupled flux is given by $\psi = N\Phi$ and using equation A.6 we obtain $\psi = Li$ with

$$L = \frac{N^2}{\left(\frac{l_r}{\mu_0 \mu_{r_r} A_r} + \frac{l_s}{\mu_0 \mu_s A_s} + 2\frac{g}{\mu_0 A_g} \right)}$$

Since both i and g are both functions of time, also is L. Hence we can write

$$\frac{d}{dt}\psi = \frac{\partial L}{\partial g}\frac{dg}{dt}i + L\frac{di}{dt}$$

$$= -L^2 \frac{2}{N^2 \mu_0 A_g}\frac{dg}{dt}i + L\frac{d}{dt}i$$

Note that if $\mu_{r_r} \gg 1$ and $\mu_{r_s} \gg 1$ that $L^2 \frac{2}{N^2 \mu_0 A_g} \approx \frac{L}{g}$. It can then easily be verified that even if g varies with high frequencies, the first term is several orders in magnitude smaller than the first. Hence, we can safely ignore the voltage induced by rotor movement and write

$$u = Ri + L\frac{d}{dt}i$$

Mechanical properties

There is a number of ways to derive the electromechanical force that is exerted on the rotor. Here we will apply the principle of virtual work. Assume an infinitesimal displacement ∂g as a result of the force (which stays constant along this small displacement). Then by the principle of energy conservation, the work done by this force must be equal to the change of the electromagnetic energy in both air gaps. Mathematically we can express this as $\partial W = F \partial g$. Noting that the energy density of a magnetic field is given by

$$w = \frac{1}{2\mu_0} \mathbf{B} \cdot \mathbf{B}$$

we find that the electromagnetic energy is given by

$$W = \frac{1}{2\mu_0} \int_V \mathbf{B} \cdot \mathbf{B} \, dV = \frac{1}{2} \frac{2g A_g B^2}{\mu_0} = \frac{g \Phi^2}{\mu_0 A_g}$$

Hence we obtain

$$F = \frac{\partial W}{\partial g} = \frac{\Phi^2}{\mu_0 A_g}$$

$$= \frac{(Ni)^2}{\mu_0 A_g \left(\frac{l_r}{\mu_0 \mu_{r_r} A_r} + \frac{l_s}{\mu_0 \mu_s A_s} + 2\frac{g}{\mu_0 A_g} \right)^2}.$$

If we ignore the magnetoresistance of the stator and rotor and hence assume that $\mu_{r_r} \gg 1$ and $\mu_{r_s} \gg 1$, we can simplify this equation to

$$F = \frac{N^2 \mu_0 A_g}{4} \frac{i^2}{g^2} = \lambda \frac{i^2}{g^2}$$

with $\lambda = \frac{1}{4} N^2 \mu_0 A_g$.

A.2.2 Active magnetic bearings in differential driving mode

As mentioned in the introduction of this section, we consider electromagnetic actuators in differential driving mode. In this configuration two magnetic coils are setup opposite to each other, together enabling the control of the position of an object in one dimension. Using the result of the previous section we can write

$$F_{\text{left}} = \lambda_{\text{left}} \left(\frac{i_{\text{left}}}{g_0 + x} \right)^2 \qquad\qquad F_{\text{right}} = \lambda_{\text{right}} \left(\frac{i_{\text{right}}}{g_0 - x} \right)^2$$

where i_{left} and i_{right} are the currents through the left and right coil respectively, g_0 is the nominal gap length beteen the object and the coil, and x is the displacement of the object from its center position. In the differential driving mode configuration, the current of two coils is chosen such, that both have an equal bias value, augmented with a control current with opposite sign:

$$i_{\text{left}} = i_{\text{bias}} - i \qquad\qquad i_{\text{right}} = i_{\text{bias}} + i$$

If we assume that the electromagnetic coils have the same properties, i.e. $\lambda = \lambda_{\text{left}} = \lambda_{\text{right}}$, we can then the express the net force that is exerted on the object as

$$F = F_{\text{right}} - F_{\text{left}}$$

$$= \lambda \left\{ \left(\frac{i_{\text{bias}} + i}{g_0 - x} \right)^2 - \left(\frac{i_{\text{bias}} - i}{g_0 + x} \right)^2 \right\}.$$

Note that due to this particular configuration, the resulting magnetic force becomes a linear function of the control current i when the rotor is at its nominal position (i.e. $x = 0$):

$$F(i)|_{x=0} = \frac{4\lambda i_{\text{bias}}}{g_0^2} \cdot i.$$

Since the rotor will generally be operating closely around this nominal position, it makes sense to approximate expression A.7a with a linear model. A first order Taylor expansion around $i = 0$ and $x = 0$ yields

$$F \approx \frac{\partial F}{\partial x} x + \frac{\partial F}{\partial i} i$$

yielding the well-known and frequently-used relation for the AMB

$$F^{\text{amb}} = k_x x + k_i i$$

with

$$k_i = \frac{4\lambda i_{\text{bias}}}{g_0^2} = \frac{N^2 \mu_0 A_g i_{\text{bias}}}{g_0^2}$$

$$k_x = \frac{4\lambda i_{\text{bias}}^2}{g_0^3} = \frac{N^2 \mu_0 A_g i_{\text{bias}}^2}{g_0^3}.$$

A.3   Discrete-time modeling

In this part, we analyze how a discrete-time model of the AMB spindle system can be derived for both configurations disucssed in the section 2.7.

A.3.1   Direct measurement of the bearing signals

The configuration in which the bearing signals are obtained by direct measurement is illustrated in figure A.3. As indicated in this figure, the output $y(t)$ is low-pass filtered by filter $L(s)$ to avoid aliasing, yielding signal $y_L(t)$, which is then sampled at intervals T_s yielding the discrete time sequence $y_L(nT_s)$. Similarly the control currents are measured, yielding the discrete time sequence $u_{1,L}(nT_s)$. It is assumed that both anti-aliasing filters $L(s)$ are identical. Let sample frequency ω_s be defined as $\omega_s = \frac{2\pi}{T_s}$. The discrete-time representation of the cutting force signal $u_{2,L}(nT_s)$ is defined as the sequence that is obtained by filtering the continuous-time cutting force signal $u_{2,0}(t)$ by $L(s)$ and then sampling the result at frequency ω_s.
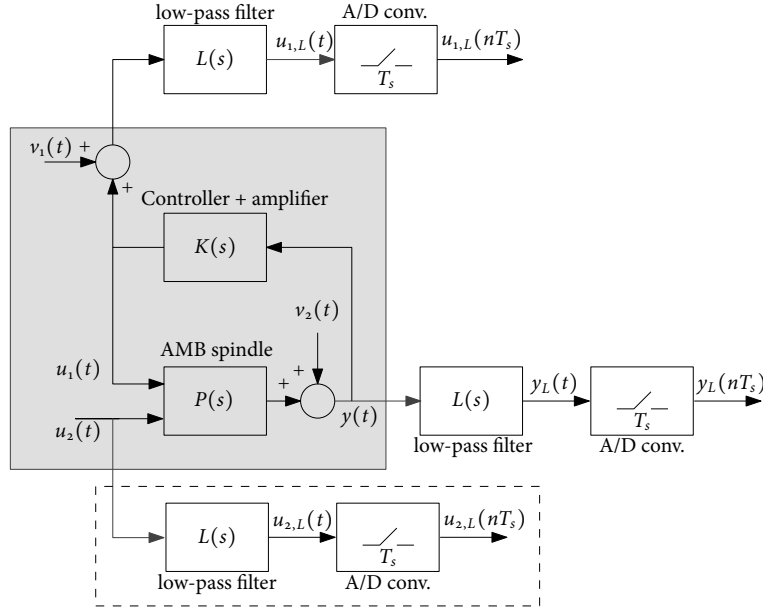
A. MODELING OF AN AMB SPINDLE

Figure A.3: Block diagram illustrating the configuration of an AMB spindle where the signals of the bearings are available through a digital control environment

In order to obtain an equivalent discrete-time model of the AMB spindle, we will derive the relation between the Discrete-Time Fourier Transform (DTFT) of $u_{1,L}(nTs)$, $u_{2,L}(nTs)$ and $y_{1,L}(nTs)$, here denoted as $U_{1,L}(e^{j\omega T_s})$, $U_{2,L}(e^{j\omega T_s})$ and $Y_{1,L}(e^{j\omega T_s})$ respectively. For this we will consider the noiseless case. First note that we can express

$$U_{1,L}(e^{j\omega T_s}) = \frac{1}{T_s} \sum_{k=-\infty}^{\infty} U_{1,L}(j\omega - k\omega_s) \tag{A.7a}$$

$$U_{2,L}(e^{j\omega T_s}) = \frac{1}{T_s} \sum_{k=-\infty}^{\infty} U_{2,L}(j\omega - k\omega_s) \tag{A.7b}$$

$$Y_{1,L}(e^{j\omega T_s}) = \frac{1}{T_s} \sum_{k=-\infty}^{\infty} Y_{1,L}(j\omega - k\omega_s). \tag{A.7c}$$

where we use the notation $X(j\omega)$ to denote the Continuous-Time Fourier Transform (CTFT) of a signal $x(t)$. For the analysis, we make the following assumptions concerning the low-pass filter $L(s)$

1. Each channel is identically low-pass filtered, i.e. $L(s)$ has the following structure:

$$L(s) = I_{4\times4} \cdot l(s) \tag{A.8}$$

2. The low-pass filter $l(s)$ satisfies $l(j\omega) = 0$, for $\omega > \frac{\omega_s}{2}$
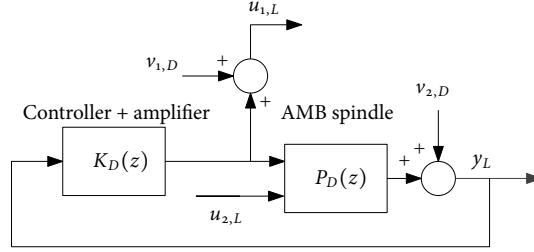
Figure A.4: Block diagram of equivalent discrete-time model of the direct measurement configuration

From these assumptions it immediately follows that for $|\omega| < \frac{\omega_s}{2}$,

$$U_{1,L}(e^{j\omega T_s}) = \frac{1}{T_s} L(j\omega) U_1(j\omega) \tag{A.9a}$$

$$U_{2,L}(e^{j\omega T_s}) = \frac{1}{T_s} L(j\omega) U_2(j\omega) \tag{A.9b}$$

$$Y_{1,L}(e^{j\omega T_s}) = \frac{1}{T_s} L(j\omega) \left[P_1(j\omega) U_1(j\omega) + P_2(j\omega) U_2(j\omega) \right] \tag{A.9c}$$

which when combined results in

$$Y_{1,L}(e^{j\omega T_s}) = P_1(j\omega) U_{1,L}(e^{j\omega T_s}) + P_2(j\omega) U_{2,L}(e^{j\omega T_s}). \tag{A.10}$$

Hence an equivalent discrete-time representation $P_D(z)$ of $P(s)$ satisfies

$$P_D(e^{j\omega T_s}) = P(j\omega), \quad |\omega| < \omega_s/2, \tag{A.11}$$

An approximation of $P_D(z)$ can be obtained using e.g. Tustin transformation of $P(s)$. With equivalent discrete-time representation of the noises in the setup, the equivalent discrete-time model for this configuration is depicted in figure A.4. Observe that this representation also contains a discrete-time equivalent representation of the combined controller and amplifier, denoted as $K_D(z)$. In this configuration we will not make any further assumptions on $K_D(z)$ other than that the closed-loop is stable.

A.3.2 Integration with a digital control environment

When the force estimator can be integrated with the digital controller of the AMB spindle, the configuration that is obtained is as depicted in figure A.5. The output $y(t)$ is low-pass filtered by filter $L(s)$ yielding signal $y_L(t)$, which is then sampled at intervals T_s yielding the discrete time sequence $y_L(nT_s)$. This discrete signal is the input for the digital controller $K(z)$, which yields the discrete control sequence $u_{1,D}(nT_s)$. This signal is reconstructed through zero-order hold, and the reconstructed signal forms the input of the current amplifier that generates the currents through the coils of the bearings. The discrete-time equivalent force signal $u_{2,L}(nTs)$ is defined as in the previous section, and filter $L(s)$ satisfies the
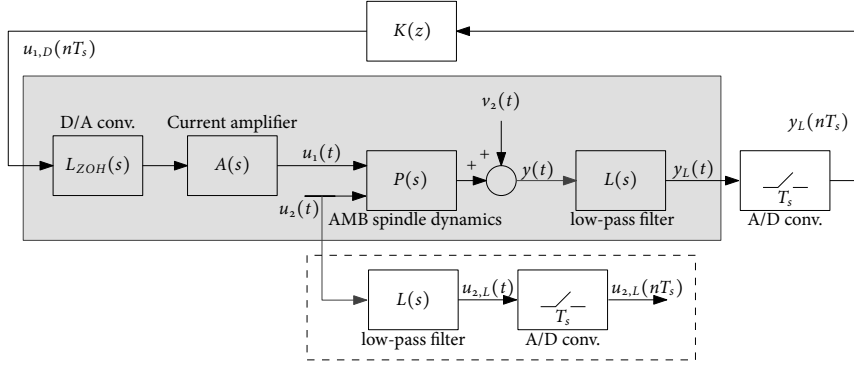
A. Modeling of an AMB spindle

Figure A.5: Block diagram illustrating the configuration of a AMB spindle where the signals of the bearings are obtained through direct measurement
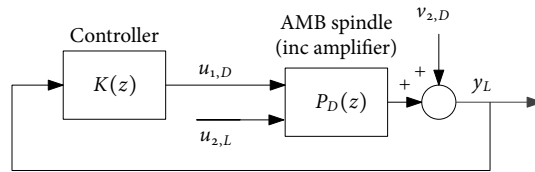

Figure A.6: Block diagram of equivalent discrete-time model of the integrated configuration

same assumptions. Following the same reasoning as the previous section, we derive that for $|\omega| < \omega_s/2$

$$Y_L(e^{j\omega T_s}) = L(j\omega)P_1(j\omega)A(j\omega)L_{ZOH}(j\omega)U_{1,d}(e^{j\omega T_s}) + P_2(j\omega)U_{2,L}(e^{j\omega T_s}), \qquad \text{(A.12)}$$

implying that for $|\omega| < \omega_s/2$ the equivalent discrete-time model $P_D(z)$ satisfies

$$P_{D,1}(e^{j\omega T_s}) = L(j\omega)P_1(j\omega)A(j\omega)L_{ZOH}(j\omega) \qquad \text{(A.13a)}$$

$$P_{D,2}(e^{j\omega T_s}) = P_2(j\omega). \qquad \text{(A.13b)}$$

This shows that $P_{D,1}(z)$ can be obtained by computing the ZOH equivalent of the cascade of the current amplifier, AMB spindle dynamics and the anti-aliasing filter, while $P_{D,2}(z)$ can be approximated by e.g. a Tustin transformation. With the discrete-time equivalent representation of the measurement noise $v_{2,D}$, the resulting discrete-time model is depicted in figure A.6.

Appendix B

Proofs

B.1 Proofs of chapter 3

B.1.1 Proof of proposition 3.5.3

In this section we prove proposition 3.5.3. First note that with some algebraic manipulations, $Psi(z)$ can be brought in the standard form of equation (3.5.1)

$$\Psi(z) = \left[\begin{array}{cc} C_c(zI - A_c)^{-1} & I \end{array}\right] \left[\begin{array}{cc} B_{c,2}QB_{c,2}^T & B_{c,2}S \\ S^T B_{c,2}^T & R_{v_2} \end{array}\right] \left[\begin{array}{c} (z^{-1}I - A_c^T)^{-1}C_c^T \\ I \end{array}\right] \tag{B.1}$$

where

$$S = \left[\begin{array}{cc} 0 & R_{v_2} \end{array}\right]^T \tag{B.2}$$

We first verify that the system satisfies the conditions of proposition 3.5.2.

Lemma B.1.1 *With* $\{A_c, B_{c,2}, C_c, Q, R_{v_2}, S\}$ *as defined in equations* (3.85), (3.86), (3.87), *and* (B.2), *define*

$$A^s = A_c - B_{c,2}SR_{v_2}^{-1}C_c \text{ and } Q^s = Q - SR_{v_2}^{-1}S^T. \tag{B.3}$$

Then we have the following facts:

1. $\left[\begin{array}{cc} B_{c,2}QB_{c,2}^T & B_{c,2}S \\ S^T B_{c,2}^T & R_{v_2} \end{array}\right] \geq 0.$
2. (C_c, A_c) *is observable.*
3. $(A^s, B_{c,2}(Q^s)^{1/2})$ *is unit-circle controllable.*

Proof We provide proof of each of the facts:

1. Observe that

$$\left[\begin{array}{cc} B_{c,2}QB_{c,2}^T & B_{c,2}S \\ S^T B_{c,2}^T & R_{v_2} \end{array}\right] = \left[\begin{array}{c} B_{c,2} \\ D_{c,2} \end{array}\right] Q \left[\begin{array}{cc} B_{c,2}^T & D_{c,2}^T \end{array}\right] \tag{B.4}$$

Positive semi definiteness follows from $Q > 0$.

2. We establish observability of (C_c, A_c) using the BPH test, i.e. (C_c, A_c) is controllable if (and only if) rank $Z(\lambda) = n + n_u + n_{v_2} \, \forall \lambda \in \mathbb{C}$, with

$$Z(\lambda) = \begin{bmatrix} \lambda I - A_u & 0 & 0 \\ B_{P,2}C_u & \lambda I - A & 0 \\ 0 & 0 & \lambda I - A_{v_2} \\ 0 & C & C_{v_2} \end{bmatrix}. \tag{B.5}$$

Assume (C_c, A_c) is not observable. Then there exists a vector $z = \text{col}(x, \xi, \zeta) \neq 0$ and a $\lambda \in \mathbb{C}$ such that $Z(\lambda)z = 0$. We examine the system for the cases that $\lambda \notin \sigma(A_u)$ and $\lambda \in \sigma(A_u)$:

- Assume $\lambda \in \{\sigma(A) \cup \sigma(A_{v_2})\}$ and $\lambda \notin \sigma(A_u)$. Then $Z(\lambda)$ has full rank based on observability of (C_P, A_P) and the minimality of the realization $(A_{v_2}, B_{v_2}, C_{v_2})$. Consequently $\forall \lambda \in \mathbb{C}$, $Z(\lambda)z = 0$ implies $z = 0$. Hence we arrive at a contradiction.

- Assume $\lambda \in \sigma(A_u)$ (and possibly also $\lambda \in \sigma(A) \cup \sigma(A_{v_2})$). Then by minimality of the realization (A_u, B_u, C_u), $(\lambda I - A_u)\xi = 0$ for $\xi \neq 0$ implies $C_u\xi \neq 0$. Let $\xi' = C_u\xi$. Then $Z(\lambda)z = 0$ implies

$$\begin{bmatrix} B_{P,2} & \lambda I - A_P & 0 \\ 0 & 0 & \lambda I - A_{v_2} \\ 0 & C_P & C_{v_2} \end{bmatrix} \begin{pmatrix} \xi' \\ x \\ \zeta \end{pmatrix} = 0 \tag{B.6}$$

or equivalently

$$\begin{bmatrix} B_{P,2}^1 & \lambda I - A_P^{11} & -A_P^{12} & 0 \\ 0 & 0 & \lambda I - A_P^{22} & 0 \\ 0 & 0 & 0 & \lambda I - A_{v_2} \\ 0 & C_P^1 & C_P^2 & C_{v_2} \end{bmatrix} \begin{pmatrix} \xi^{1\prime} \\ \xi^{2\prime} \\ x \\ \zeta \end{pmatrix} = 0 \tag{B.7}$$

From observability of (C_P, A_P), equation (3.80) and the minimality of the realization $(A_{v_2}, B_{v_2}, C_{v_2})$ we conclude that the matrix in the above equality has full column rank, from which it follows $\text{col}(\xi', x, \zeta) = 0$. Again, we arrive at a contradiction.

In both cases we arrive at a contradiction, which proves the observability of (C_c, A_c). Observe that (3.80) is sufficient for (C_c, A_c) to be observable, although it is not necessary. From the above, it can be derived easily that a necessary requirement is that the triplet $(A_P^{11}, B_{P,2}^1, C_P^1)$ has no invariant zeros that coincide with the eigenvalues of A_u. The more conservative condition (3.80) ensures observability for any input spectrum Φ_u (see also the remark under assumption 3.4.8).

3. Unit-circle controllability of $(A^s, B_{c,2}(Q^s)^{1/2})$ is equivalent to

$$\text{rank} \begin{bmatrix} \lambda I - A^s & B_{c,2}(Q^s)^{1/2} \end{bmatrix} = n_u + n + n_{v_2} \quad \text{for all } |\lambda| = 1 \tag{B.8}$$

with $n_u = \dim(A_u)$ and $n_{v_2} = \dim(A_{v_2})$. It is easily verified that

$$\begin{bmatrix} \lambda I - A^s & B_{c,2}(Q^s)^{1/2} \end{bmatrix} = \begin{bmatrix} \lambda I - A_u & 0 & 0 & B_u R_u^{1/2} \\ B_{P,2} & \lambda I - A_P & 0 & 0 \\ 0 & B_{v_2}C_P & \lambda I - A_{v_2} + B_{v_2}C_{v_2} & 0 \end{bmatrix}.$$

From minimality of the realizations (A_u, B_u, C_u), $(A_{v_2}, B_{v_2}, C_{v_2})$ and equation (3.78), identity (B.8) is immediate. ■

Using proposition 3.5.2 and lemma B.1.1, we can now obtain an explicit expression for $M(z)$, i.e.

$$M^{-1}(z) = I - C_c(zI - \bar{A}_c)^{-1}L, \text{ where} \tag{B.9}$$

$$L = (A_c \Pi C_c^T + B_{c,2}S)R^{-1} \tag{B.10}$$

$$R = R_{v_2} + C_c \Pi C_c^T \tag{B.11}$$

and Π is the unique positive semi-definite solution to the discrete algebraic Ricatti equation (DARE)

$$\Pi = A_c \Pi A_c^T + B_{c,2}QB_{c,2}^T - LRL^T \tag{B.12}$$

such that $\bar{A}_c = A_c - LC_c$ is stable.

B.1.2 Proof of proposition 3.5.4

To complete the expression for $F_{0,N}(z)$ we need to evaluate the causality operator in (3.92). Observe that $\mathcal{F}(z)$ can also be brought in the form

$$\mathcal{F}(z) = \begin{bmatrix} C_u(zI - A_u)^{-1} & I \end{bmatrix} \begin{bmatrix} B_u \begin{bmatrix} R_u & 0 \end{bmatrix} B_{c,2}^T & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} (z^{-1}I - \bar{A}_c)^{-1}C_c^T \\ I \end{bmatrix} \tag{B.13}$$

Observe that due to the stability of A_u the first factor of (B.13) is strictly causal, while the stability of $\bar{A}_c$ implies that the third factor is strictly anti-causal. Hence, by transforming expression (B.13) in a form in which the right-upper block of the center matrix is zero, we can separate the causal and anti-causal part of $\mathcal{F}(z)$. To achieve this, we use the following lemma:

Lemma B.1.2 *Any z-cross spectrum of the form*

$$S_{12}(z) = \begin{bmatrix} C_1(zI - A_1)^{-1} & I \end{bmatrix} \begin{bmatrix} K & L \\ M & N \end{bmatrix} \begin{bmatrix} (z^{-1}I - A_2^T)^{-1}C_2^T \end{bmatrix} \tag{B.14}$$

is invariant under transformation

$$\begin{bmatrix} K & L \\ M & N \end{bmatrix} \rightarrow \begin{bmatrix} K - Q + A_1 Q A_2^T & L + A_2 Q C_2^T \\ M + C_1 Q A_2^T & N + C_1 Q C_2^T \end{bmatrix} \tag{B.15}$$

for any matrix Q with appropriate dimensions.

Proof A proof for the case $A_1 = A_2$ and $C_1 = C_2$ in given in [83]. The more general formulation in this lemma is proven similarly. ■

B. Proofs

Now, note that we can express for the left-upper block of the center matrix of $\mathcal{F}(z)$

$$B_u \begin{bmatrix} R_u & 0 \end{bmatrix} B_{c,2}^T = \begin{bmatrix} I & 0 & 0 \end{bmatrix} B_{c,2} Q B_{c,2}^T.$$

From equation (B.12) we infer that

$$B_{c,2} Q B_{c,2}^T = \Pi - A_c \Pi \bar{A}_c^T + B_{c,2} S L^T.$$

Furthermore, partition Π according to the partitioning of A_c as

$$\Pi = \begin{bmatrix} \Pi_{11} & \Pi_{12} & \Pi_{13} \\ \Pi_{21} & \Pi_{22} & \Pi_{23} \\ \Pi_{31} & \Pi_{32} & \Pi_{33} \end{bmatrix}$$

and let $Z = \begin{bmatrix} \Pi_{11} & \Pi_{12} & \Pi_{13} \end{bmatrix}$. With this, we obtain

$$\begin{aligned} B_u \begin{bmatrix} R_u & 0 \end{bmatrix} B_{c,2}^T &= \begin{bmatrix} I & 0 & 0 \end{bmatrix} (\Pi - A_c \Pi \bar{A}_c^T + B_{c,2} S L^T) \\ &= Z - A_u Z \bar{A}_c^T \end{aligned}$$

where it used that $\begin{bmatrix} I & 0 & 0 \end{bmatrix} B_{c,2} S = 0$. Applying lemma B.1.2 with $Q = Z$ yields

$$\begin{aligned} \mathcal{F}(z) &= \begin{bmatrix} C_u(zI - A_u)^{-1} & I \end{bmatrix} \begin{bmatrix} Z - A_u Z \bar{A}_c^T & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} (z^{-1}I - \bar{A}_c^T)^{-1} C_c^T \\ I \end{bmatrix} \\ &= \begin{bmatrix} C_u(zI - A_u)^{-1} & I \end{bmatrix} \begin{bmatrix} 0 & A_u Z C_c^T \\ C_u Z \bar{A}_c^T & C_u Z C_c^T \end{bmatrix} \begin{bmatrix} (z^{-1}I - \bar{A}_c^T)^{-1} C_c^T \\ I \end{bmatrix} \end{aligned}$$

or equivalently

$$\mathcal{F}(z) = \underbrace{C_u(zI - A_u)^{-1} A_u Z C_c^T}_{:=\mathcal{F}_1(z)} + \underbrace{C_u Z C_c^T + C_u Z \bar{A}_c^T (z^{-1}I - \bar{A}_c^T)^{-1} C_c^T}_{:=\mathcal{F}_2(z)} \tag{B.16}$$

where we observe that $\mathcal{F}_1(z)$ is strictly causal and $\mathcal{F}_2(z)$ is anti-causal. From this, we can directly identify the causal part of $z^{-N}\mathcal{F}(z)$:

$$\left\{ z^{-N}\mathcal{F}(z) \right\}_+ = z^{-N}\mathcal{F}_1(z) + \left\{ z^{-N}\mathcal{F}_2(z) \right\}_+. \tag{B.17}$$

To evaluate the second term, observe that we can write

$$\mathcal{F}_2(z) = C_u Z (I + \bar{A}_c^T (z^{-1}I - \bar{A}_c^T)^{-1}) C_c^T \tag{B.18}$$

$$= C_u Z \sum_{m=1}^{\infty} z^{m-1} (\bar{A}_c^T)^{m-1} C_c^T \tag{B.19}$$

from which it follows that

$$\left\{ z^{-N}\mathcal{F}_2(z) \right\}_+ = C_u Z \sum_{m=1}^{N+1} z^{m-N-1} (\bar{A}_c^T)^{m-1} C_c^T \tag{B.20}$$

196

Using (B.17) we can express the optimal filter solution (3.92) as

$$z^{-N}F_{0,N} = \left[\begin{array}{cc} 0 & \underbrace{z^{-N}\mathcal{F}_1(z)R^{-1}M^{-1}(z)S(z)^{-1}}_{:=F_{0,N;2a}(z)} + \underbrace{\{z^{-N}\mathcal{F}_2(z)\}_+ R^{-1}M^{-1}(z)S(z)^{-1}}_{:=F_{0,N;2b}(z)} \end{array} \right].$$

(B.21)

We will simplify both terms $F_{0,N;2a}(z)$ and $F_{0,N;2b}(z)$. First note that we can express

$$M^{-1}(z)S^{-1}(z) = (I - C_c(zI - \bar{A}_c)^{-1}L)(I - P_1(z)K(z)) \tag{B.22}$$

$$= (I - C_c(zI - \bar{A}_c)^{-1}L)(I - C_P(zI - A_P)^{-1}B_{P,1}K(z)) \tag{B.23}$$

$$= (I - C_c(zI - \bar{A}_c)^{-1}L)(I - C_c(zI - A_c)^{-1}B_{c,1}K(z)) \tag{B.24}$$

$$= I - C_c(zI - \bar{A}_c)^{-1}(L + B_{c,1}K(z) \tag{B.25}$$

To simplify $F_{0,N;2a}(z)$, partition $L = \text{col}(L_1, L_2, L_3)$ according the partitioning of A_c. From equation (B.30) it follows that $L_1 = A_u Z C_c^T R^{-1}$. With this it can easily be verified that $\mathcal{F}_1(z)R^{-1} = \left[\begin{array}{ccc} C_u & 0 & 0 \end{array} \right](zI - A_c)^{-1}L$, so that

$$F_{0,N;2a}(z) = z^{-N}\left[\begin{array}{ccc} C_u & 0 & 0 \end{array} \right](zI - A_c)^{-1}L(I - C_c(zI - \bar{A}_c)^{-1}(L + B_{c,1}K(z))$$

$$= z^{-N}\left[\begin{array}{ccc} C_u & 0 & 0 \end{array} \right](zI - \bar{A}_c)^{-1}(L + B_{c,1}K(z))$$

Using equation (B.25), we can obtain for $F_{0,N;2b}(z)$:

$$F_{0,N;2b}(z) = \left[\begin{array}{ccc} C_u & 0 & 0 \end{array} \right]\Pi \sum_{m=1}^{N+1} z^{m-N-1}(\bar{A}_c^T)^{m-1}C_c^T R^{-1}\cdot$$

$$\cdot (I - C_c(zI - \bar{A}_c)^{-1}(L + B_{P,1}K(z)).$$

B.1.3 Proof of proposition 3.5.5

In this appendix, the proof of proposition 3.5.5 is given. First, observe that we can express for $\bar{\Psi}$

$$\bar{\Psi}(z) = P_{c,2}(z)QP_{c,2}^H(z) \tag{B.26}$$

$$= \left[\begin{array}{cc} C_c(zI - A_c)^{-1} & I \end{array} \right]\left[\begin{array}{cc} B_{c,2}QB_{c,2}^T & B_{c,2}S \\ S^T B_{c,2}^T & R_{v_2} \end{array} \right]\left[\begin{array}{c} (z^{-1}I - A_c^T)^{-1}C_c^T \\ I \end{array} \right] \tag{B.27}$$

which has the form of the spectrum defined in proposition 3.5.1. As in the proof of proposition 3.5.3, we first verify that this realization ensures that $\bar{\Psi}(z) > 0$, for $|z| = 1$, i.e. if the conditions of proposition 3.5.2 are satisfied.

Lemma B.1.3 *With $\{A_c, B_{c,2}, C_c, Q, R_{v_2}, S\}$ as defined in equations (3.85), (3.96), and (3.97), define*

$$A^s = A_c - B_{c,2}SR_{v_2}^{-1}C_c \text{ and } Q^s = Q - SR_{v_2}^{-1}S^T. \tag{B.28}$$

Then we have the following facts:

197

1. $\begin{bmatrix} B_{c,2} Q B_{c,2}^T & B_{c,2} S \\ S^T B_{c,2}^T & R_{v_2} \end{bmatrix} \geq 0.$

2. (C_c, A_c) is observable.

3. $(A^s, B_{c,2}(Q^s)^{1/2})$ is unit-circle controllable.

Proof The proof of this lemma goes along the same lines as the proof of lemma B.1.1. Hence we only give the main arguments.

1. Follows from $Q > 0$.

2. Follows from minimality of the realizations (A_u, B_u, C_u), $(A_{v_i}, B_{v_i}, C_{v_i})$, $i = 1, 2$, and identities (3.80) and (3.81).

3. Follows from minimality of the realizations (A_u, B_u, C_u), $(A_{v_i}, B_{v_i}, C_{v_i})$, $i = 1, 2$ and identity (3.78).

Having established that $\bar{\Psi}(z) > 0$, for $|z| = 1$, we again use proposition 3.5.2 to obtain an explicit expression for $\bar{M}^{-1}(z)$, i.e.

$$\bar{M}^{-1}(z) = I - C_c(zI - \bar{A}_c)^{-1} L, \quad \text{where} \tag{B.29}$$

$$L = (A_c \Pi C_c^T + B_{c,2} S) \bar{R}^{-1} \tag{B.30}$$

$$\bar{R} = R_{v_2} + C_c \Pi C_c^T \tag{B.31}$$

and Π is the unique positive semi-definite solution to the discrete algebraic Ricatti equation (DARE)

$$\Pi = A_c \Pi A_c^T + B_{c,2} Q B_{c,2}^T - L \bar{R} L^T \tag{B.32}$$

such that $\bar{A}_c = A_c - LC_c$ is stable.

B.1.4 Proof of proposition 3.5.6

Recall from proposition 3.4.7 that the solution of the N-causal input estimation problem without controller knowledge is given by[1]

$$z^{-N} F_{0,N} = \{z^{-N} \underbrace{\Phi_u(z) P_2^H(z) \bar{M}^{-H}(z)}_{\mathcal{F}(z)}\}_+ \bar{R}^{-1} \bar{M}^{-1}(z) \begin{bmatrix} -P_1(z) & I \end{bmatrix} \tag{B.33}$$

where $\bar{M}(z)$ and $\bar{R} > 0$ are found from the spectral factorization of

$$\bar{\Psi}(z) = P_2(z) \Phi_u(z) P_2^H(z) + P_1(z) \Phi_{v_1}(z) P_1^H(z) + \Phi_{v_2}(z). \tag{B.34}$$

Following a similar procedure as in the proof of proposition (3.5.4), it can be verified we can also express (B.33) as

$$z^{-N} F_{0,N} = (z^{-N} \mathcal{F}_1(z) + \{z^{-N} \mathcal{F}_2(z)\}_+) \bar{R}^{-1} \begin{bmatrix} -\bar{M}^{-1}(z) P_1(z) & \bar{M}^{-1}(z) \end{bmatrix} \tag{B.35}$$

---

[1]Again, we use the normalized factorization $\bar{\Psi} = \bar{M} \bar{R} \bar{M}^H$ with $\bar{M}(\infty) = I$

where $\mathcal{F}_1(z)$ and $\mathcal{F}_2(z)$ are as defined in equation (B.16), but then evaluated for the extended realization defined here in this section. We will perform some simplifications. First note that

$$\bar{M}^{-1}(z)P_1(z) = (I - C_c(zI - \bar{A}_c)^{-1}L)C_p(zI - A_P)^{-1}B_{P,1} \tag{B.36}$$

$$= (I - C_c(zI - \bar{A}_c)^{-1}L)C_c(zI - A_c)^{-1}B_{c,1} \tag{B.37}$$

$$= C_c(zI - \bar{A}_c)^{-1}B_{c,1} \tag{B.38}$$

so that

$$\begin{bmatrix} -\bar{M}^{-1}(z)P_1(z) & \bar{M}^{-1}(z) \end{bmatrix} = \begin{bmatrix} 0 & I \end{bmatrix} - C_c(zI - \bar{A}_c)^{-1}\begin{bmatrix} B_{c,1} & L \end{bmatrix} \tag{B.39}$$

With some algebraic manipulations, it follows that

$$z^{-N}\mathcal{F}_1(z)\bar{R}^{-1}\begin{bmatrix} -\bar{M}^{-1}(z)P_1 & \bar{M}^{-1}(z) \end{bmatrix} =$$
$$z^{-N}\begin{bmatrix} C_u & 0 & 0 & 0 \end{bmatrix}(zI - \bar{A}_c)^{-1}\begin{bmatrix} B_{c,1} & L \end{bmatrix} \tag{B.40}$$

and

$$\{z^{-N}\mathcal{F}_2(z)\}_+ R^{-1}\begin{bmatrix} -\bar{M}^{-1}(z)P_1 & \bar{M}^{-1}(z) \end{bmatrix} = \begin{bmatrix} C_u & 0 & 0 & 0 \end{bmatrix}\Pi\cdot$$
$$\cdot \sum_{m=1}^{N+1} z^{m-N-1}(\bar{A}_c^T)^{m-1}C_c^T\bar{R}^{-1}\left(\begin{bmatrix} 0 & I \end{bmatrix} - C_c(zI - \bar{A}_c)^{-1}\begin{bmatrix} B_{c,1} & L \end{bmatrix}\right) \tag{B.41}$$

B.2 Proofs of chapter 4

In this appendix we give the proof of propositions 4.4.2 and 4.4.4. We will first give the following lemma.

Lemma B.2.1 *Suppose $A(\theta)$ is a complex square matrix, depending on the complex vector θ, which takes values in an open subset $V \subseteq \mathbb{C}^l$. Furthermore, suppose that $A(\theta)$ is invertible and analytical for all $\theta \in V$. Then for $\theta \in V$*

$$\frac{d}{d\theta}\text{vec}[A^{-1}(\theta)] = [-A^{-T}(\theta) \otimes A^{-1}(\theta)]\frac{d}{d\theta}\text{vec}[A(\theta)]. \tag{B.42}$$

Proof Let θ_i be the ith element of θ. Using the product rule for matrix differentation, we derive that

$$\frac{\partial}{\partial\theta_i}(A(\theta) \cdot A^{-1}(\theta)) = \frac{\partial A(\theta)}{\partial\theta_i}A^{-1}(\theta) + A(\theta)\frac{\partial A^{-1}(\theta)}{\partial\theta_i} = 0$$

from which immediately follows that

$$\frac{\partial A^{-1}(\theta)}{\partial\theta_i} = -A^{-1}(\theta)\frac{\partial A(\theta)}{\partial\theta_i}A^{-1}(\theta).$$

Using the identity $\text{vec}(ABC) = (C^T \otimes A)\text{vec}(B)$, we infer

$$\frac{\partial}{\partial\theta_i}\text{vec}[A^{-1}(\theta)] = -[A^{-T}(\theta) \otimes A^{-1}(\theta)]\text{vec}[\frac{\partial A(\theta)}{\partial\theta_i}],$$

B. Proofs

implying

$$\frac{d}{d\theta}\text{vec}[A^{-1}(\theta)] = -[A^{-T}(\theta) \otimes A^{-1}(\theta)]\frac{d}{d\theta}\text{vec}[A(\theta)].$$

This proofs the claim.

B.2.1 Proof of proposition 4.4.2

We will first introduce the notation $\theta^T = \begin{bmatrix} \theta_A^T & \theta_B^T \end{bmatrix}$ where θ_A only contains the parameters used to define $A(\xi, \theta)$ and θ_B those to define $B(\xi, \theta)$. For brevity, we will drop the dependency of P, B and A on $\xi(\omega_k)$ from here on. Observe that with this, we can write

$$M_k(\theta) = \begin{bmatrix} \frac{\partial}{\partial\theta_A}\text{vec}[P(\theta)] & \frac{\partial}{\partial\theta_B}\text{vec}[P(\theta)] \end{bmatrix}$$

Using the identities (4.8a) and (4.8b), we rewrite this to

$$M_k(\theta) = \begin{bmatrix} [B^T(\theta) \otimes I]\frac{\partial}{\partial\theta_A}\text{vec}[A^{-1}(\theta)] & [I \otimes A^{-1}(\theta)]\frac{\partial}{\partial\theta_B}\text{vec}[B(\theta)] \end{bmatrix}. \qquad \text{(B.43)}$$

We will derive expressions for $\frac{\partial}{\partial\theta_A}\text{vec}[A^{-1}(\theta)]$ and $\frac{\partial}{\partial\theta_B}\text{vec}[B(\theta)]$. Application of lemma B.2.1 yields

$$\frac{\partial}{\partial\theta_A}\text{vec}[A^{-1}(\theta)] = \begin{bmatrix} -A^{-T}(\theta) \otimes A^{-1}(\theta) \end{bmatrix}\frac{\partial}{\partial\theta_A}\text{vec}[A(\theta)]$$

$$= \begin{bmatrix} -A^{-T}(\theta) \otimes A^{-1}(\theta) \end{bmatrix} \cdot \begin{bmatrix} I_{pp \times pp}\xi(\omega_k)^{n_a-1} & \dots & I_{pp \times pp}\xi(\omega_k)^0 \end{bmatrix}$$

$$= \begin{bmatrix} -A^{-T}(\theta) \otimes A^{-1}(\theta) \end{bmatrix} \cdot \left(\begin{bmatrix} I_{p \times p}\xi(\omega_k)^{n_a-1} & \dots & I_{p \times p}\xi(\omega_k)^0 \end{bmatrix} \otimes I_{p \times p} \right).$$

Furthermore, we derive that

$$\frac{\partial}{\partial\theta_B}\text{vec}[B(\theta)] = \begin{bmatrix} I_{mp \times mp}\xi(\omega_k)^{n_b} & \dots & I_{mp \times mp}\xi(\omega_k)^0 \end{bmatrix}$$

$$= \begin{bmatrix} I_{m \times m}\xi(\omega_k)^{n_b} & \dots & I_{m \times m}\xi(\omega_k)^0 \end{bmatrix} \otimes I_{p \times p}.$$

With this, we express the first element of $M_k(\theta)$ in (B.43) as

$$\begin{bmatrix} -B^T(\theta) \otimes I \end{bmatrix}\begin{bmatrix} -A^{-T}(\theta) \otimes A^{-1}(\theta) \end{bmatrix} \cdot$$

$$\cdot \left(\begin{bmatrix} I_{p \times p}(\xi(\omega_k)^{n_a-1} & \dots & I_{p \times p}\xi(\omega_k)^0 \end{bmatrix} \otimes I_{p \times p} \right)$$

$$= \begin{bmatrix} -P^T(\theta) \otimes A^{-1}(\theta) \end{bmatrix} \cdot \left(\begin{bmatrix} I_{p \times p}\xi(\omega_k)^{n_a-1} & \dots & I_{p \times p}\xi(\omega_k)^0 \end{bmatrix} \otimes I_{p \times p} \right)$$

$$= \begin{bmatrix} -P^T(\theta)\xi(\omega_k)^{n_a-1} & \dots & -P^T(\theta)\xi(\omega_k)^0 \end{bmatrix} \otimes A^{-1}(\theta),$$

and the second element of $M_k(\theta)$ in (B.43) as

$$\begin{bmatrix} I \otimes A^{-1}(\theta) \end{bmatrix} \cdot \left(\begin{bmatrix} I_{m \times m}\xi(\omega_k)^{n_b} & \dots & I_{m \times m}\xi(\omega_k)^0 \end{bmatrix} \otimes I_{p \times p} \right)$$

$$= \begin{bmatrix} I_{m \times m}\xi(\omega_k)^{n_b} & \dots & I_{m \times m}\xi(\omega_k)^0 \end{bmatrix} \otimes A^{-1}(\theta).$$

Combining these results, and using the identity $\begin{bmatrix} A_1 \otimes B & A_2 \otimes B \end{bmatrix} = \begin{bmatrix} A_1 & A_2 \end{bmatrix} \otimes B$, we infer

$$M_k(\theta) = \Phi_k(\theta)^T \otimes A^{-1}(\theta), \qquad \text{(B.44)}$$

which proves the claim.

B.2.2 Proof of proposition 4.4.4

To facilitate the proof, we will use the following notation. Let parameter matrix Θ be partitioned as $\Theta^T = \begin{bmatrix} \Theta_A^T & \Theta_B^T \end{bmatrix}$ where Θ_A is comprised the parameters used to define $A(\xi, \theta)$, and Θ_B is comprised of those to define $B(\xi, \theta)$. We will denote the ith column of Θ as θ_i, the ith column of Θ_A as $\theta_{A,i}$ and i^{th} column of Θ_B as $\theta_{B,i}$. Observe that with this notation we can express

$$\frac{\partial}{\partial \theta} \mathrm{vec}[P(\theta)] = \begin{bmatrix} \frac{\partial}{\partial \theta_{A,1}} \mathrm{vec}[P(\theta)] & \frac{\partial}{\partial \theta_{B,1}} \mathrm{vec}[P(\theta)] \\[2mm] \cdots & \frac{\partial}{\partial \theta_{A,m}} \mathrm{vec}[P(\theta)] & \frac{\partial}{\partial \theta_{B,m}} \mathrm{vec}[P(\theta)] \end{bmatrix}. \quad \text{(B.45)}$$

Using the identities of (4.8a), we can write

$$\frac{\partial}{\partial \theta_{A,i}} \mathrm{vec}[P(\theta)] = [I \otimes B(\theta)] \frac{\partial}{\partial \theta_{A,i}} \mathrm{vec}[A^{-1}(\theta)] \quad \text{(B.46)}$$

$$\frac{\partial}{\partial \theta_{B,i}} \mathrm{vec}[P(\theta)] = [A^{-T}(\theta) \otimes I] \frac{\partial}{\partial \theta_{A,i}} \mathrm{vec}[B(\theta)] \quad \text{(B.47)}$$

for $i = 1 \ldots m$. Following lemma B.2.1, we note that

$$\frac{\partial}{\partial \theta_{A,i}} \mathrm{vec}[A^{-1}(\theta)] = [-A^{-T}(\theta) \otimes A^{-1}(\theta)] \frac{\partial}{\partial \theta_{A,i}} \mathrm{vec}[A(\theta)]$$

$$= [-A^{-T}(\theta) \otimes A^{-1}(\theta)] \begin{bmatrix} 0_{(i-1)\cdot m \times (n_a \cdot m)} \\ \Lambda_A(\omega_k) \\ 0_{(m-i)\cdot m \times (n_a \cdot m)} \end{bmatrix}$$

with

$$\Lambda_A(\omega_k) = \begin{bmatrix} I_{m \times m} \xi(\omega_k)^{n_a - 1} & \cdots & I_{m \times m} \xi(\omega_k)^0 \end{bmatrix}.$$

Substituting this in equation (B.46) yields

$$\frac{\partial}{\partial \theta_{A,i}} \mathrm{vec}[P(\theta)] = [I \otimes B(\theta)][-A^{-T}(\theta) \otimes A^{-1}(\theta)] \begin{bmatrix} 0_{(i-1)\cdot m \times (n_a \cdot m)} \\ \Lambda_A(\omega_k) \\ 0_{(m-i)\cdot m \times (n_a \cdot m)} \end{bmatrix}$$

$$= [-A^{-T}(\theta) \otimes I][I \otimes P(\theta)] \begin{bmatrix} 0_{(i-1)\cdot m \times (n_a \cdot m)} \\ \Lambda_A(\omega_k) \\ 0_{(m-i)\cdot m \times (n_a \cdot m)} \end{bmatrix}$$

$$= [A^{-T}(\theta) \otimes I] \begin{bmatrix} 0_{(i-1)\cdot p \times (n_a \cdot m)} \\ -P(\theta)\Lambda_A(\omega_k) \\ 0_{(m-i)\cdot p \times (n_a \cdot m)} \end{bmatrix}. \quad \text{(B.48)}$$

Similarly, we can write

$$\frac{\partial}{\partial \theta_{B,i}} \mathrm{vec}[B(\theta)] = \begin{bmatrix} 0_{(i-1)\cdot p \times ((n_b+1)\cdot m)} \\ \Lambda_B(\omega_k) \\ 0_{(m-i)\cdot p \times ((n_b+1)\cdot m)} \end{bmatrix}$$

B. Proofs

with

$$\Lambda_B(\omega_k) = \left[\begin{array}{ccc} I_{p\times p}\xi(\omega_k)^{n_b} & \dots & I_{p\times p}\xi(\omega_k)^0 \end{array}\right],$$

which when substituted in equation (B.47) yields

$$\frac{\partial}{\partial\theta_{B,i}}\text{vec}[P(\theta)] = [A^{-T}(\theta)\otimes I]\left[\begin{array}{c} 0_{(i-1)\cdot p\times((n_b+1)\cdot m)} \\ \Lambda_B(\omega_k) \\ 0_{(m-i)\cdot p\times((n_b+1)\cdot m)} \end{array}\right]. \tag{B.49}$$

Combining equations (B.48) and (B.49) gives

$$\frac{\partial}{\partial\theta_i}\text{vec}[P(\theta)] = [A^{-T}(\theta)\otimes I]\left[\begin{array}{cc} 0_{(i-1)\cdot p\times(n_a\cdot m)} & 0_{(i-1)\cdot p\times((n_b+1)\cdot m)} \\ -P(\theta)\Lambda_A(\omega_k) & \Lambda_B(\omega_k) \\ 0_{(m-i)\cdot p\times(n_a\cdot m)} & 0_{(m-i)\cdot p\times((n_b+1)\cdot m)} \end{array}\right].$$

From this, it is easily seen that we can write

$$\frac{\partial}{\partial\theta}\text{vec}[P(\theta)] = [A^{-T}(\theta)\otimes I][I\otimes\Phi_k(\theta)]$$

$$= [A^{-T}(\theta)\otimes\Phi_k(\theta)] \tag{B.50}$$

with $\Phi_k(\theta) = \left[\begin{array}{cc} -P(\theta)\Lambda_A(j\omega_k) & \Lambda_B(j\omega_k) \end{array}\right]$, which completes the proof.

B.3 Proofs of chapter 6

B.3.1 Proof of equation (6.15)

First note that we can express

$$T(q,\theta)u(t) = \mathcal{C}\mathcal{A}^{t-1}\bar{x}(1) + \sum_{k=1}^{t-1}\mathcal{C}\mathcal{A}^k\left(\begin{array}{c} I \\ 0 \end{array}\right)B_{P,2}u_2(t-k)$$

$$= \mathcal{C}\mathcal{A}^{t-1}\bar{x}(1) + \sum_{k=1}^{t-1}\left[u^T(t-k)\otimes\mathcal{C}\mathcal{A}^k\left(\begin{array}{c} I \\ 0 \end{array}\right)\right]\text{vec}(B_{P,2})$$

$$= \mathcal{C}\mathcal{A}^{t-1}\bar{x}(1)+$$

$$+ \sum_{k=1}^{t-1}\left[\begin{array}{ccc} u_{2,1}(t-k)\mathcal{C}\mathcal{A}^k\left(\begin{array}{c} I \\ 0 \end{array}\right) & \dots & u_{2,m_2}(t-k)\mathcal{C}\mathcal{A}^k\left(\begin{array}{c} I \\ 0 \end{array}\right) \end{array}\right]\text{vec}(B_{P,2})$$

$$= \mathcal{C}\mathcal{A}^{t-1}\bar{x}(1)+$$

$$+ \left[\left\{\left[\begin{array}{ccc} u_{2,1}(t-1) & \dots & u_{2,1}(1) \end{array}\right]\otimes I\right\}\left[\begin{array}{c} \mathcal{C} \\ \mathcal{C}\mathcal{A} \\ \vdots \\ \mathcal{C}\mathcal{A}^{t-2} \end{array}\right]\left(\begin{array}{c} I \\ 0 \end{array}\right)\right.$$

$$\dots$$

$$\left\{\left[\begin{array}{ccc} u_{2,m_2}(t-1) & \dots & u_{2,m_2}(1) \end{array}\right]\otimes I\right\}\left[\begin{array}{c} \mathcal{C} \\ \mathcal{C}\mathcal{A} \\ \vdots \\ \mathcal{C}\mathcal{A}^{t-2} \end{array}\right]\left(\begin{array}{c} I \\ 0 \end{array}\right)\right]\text{vec}(B_{P,2})$$

Here we have used $u_{2,i}(t)$ to indicate element number i of the input vector u_2 at time instant t. Now let $E(\theta)$ represent the vector of all errors, i.e.

$$E(\theta) = \begin{bmatrix} \epsilon(1) \\ \vdots \\ \epsilon(N) \end{bmatrix} \tag{B.51}$$

With this, it can immediately be derived that

$$E(\theta) = \Gamma_M^{-1}(Y - \Psi\theta) \tag{B.52}$$

with

$$\theta = \begin{bmatrix} \bar{x}(1) \\ \text{vec}(B_{P,2}) \end{bmatrix} \tag{B.53}$$

and

$$\Psi = \begin{bmatrix} \mathcal{O} & \mathcal{U}_1\mathcal{O}_P & \mathcal{U}_2\mathcal{O}_P & \cdots & \mathcal{U}_{m_2}\mathcal{O}_P \end{bmatrix} \tag{B.54}$$

where $\mathcal{O}_P$ are the first n columns of $\mathcal{O}$, i.e.

$$\mathcal{O} = \begin{bmatrix} \mathcal{O}_P & \mathcal{O}_C \end{bmatrix}, \tag{B.55}$$

and

$$\mathcal{U}_i = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ u_{2,i}(1) & 0 & 0 & \cdots & 0 \\ u_{2,i}(2) & u_{2,i}(1) & 0 & \cdots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ u_{2,i}(N-1) & u_{2,i}(N-2) & \cdots & u_{2,i}(1) & 0 \end{bmatrix} \otimes I.$$

Noting that the squared norm of the vector $E(\theta)$ is identical to $V_N(\theta, Z^N)$, we have obtained the desired result.

B.4 Proofs of chapter 7

B.4.1 Proof of equation (7.22)

To proof equation (7.22), observe that for v_ϕ Gaussian distributed

$$\mathbb{E}\cos(m(\phi - v_\phi)) = e^{-(m\sigma_\phi)^2/2}\cos(m\phi) \tag{B.56}$$

$$\mathbb{E}\sin(m(\phi - v_\phi)) = e^{-(m\sigma_\phi)^2/2}\sin(m\phi) \tag{B.57}$$

We will prove the first of these identities, the second can be proven similarly. Note that

$$\mathbb{E}\cos(m(\phi - v_\phi)) = \cos(m\phi)\mathbb{E}\cos(mv_\phi) - \sin(m\phi)\mathbb{E}\sin(mv_\phi) \tag{B.58}$$

$$= \cos(m\phi)\mathbb{E}\cos(mv_\phi) \tag{B.59}$$

B. Proofs

where it has been used that $\sin(mv_\phi)$ has zero mean (this follows directly from the fact that $\sin(\cdot)$ is an odd function and v_ϕ is Gaussian distributed). Now,

$$\mathbb{E}\cos(mv_\phi) = \int_{-\infty}^{\infty}\cos(m\xi)p(\xi)d\xi \qquad (B.60)$$

with

$$p(\xi) = \frac{1}{\sqrt{2\pi}\,\sigma_\phi}e^{-x^2/(2\sigma_\phi^2)} \qquad (B.61)$$

the Gaussian probability density function for a process with zero mean and standard deviation σ_ϕ. Evaluating the integral in (B.60), while using the identity

$$\int_{-\infty}^{\infty}e^{-\xi^2}\cos(a\xi)d\xi = \sqrt{\pi}e^{-a^2/4} \qquad (B.62)$$

yields

$$\mathbb{E}\cos(mv_\phi) = e^{-(m\sigma_\phi)^2/2}. \qquad (B.63)$$

From equations (B.56) and (B.57) it immediately follows that

$$\mathbb{E}H(\phi(t) + v_\phi(t)) = \mathcal{Q}^{-1}H(\phi(t)), \qquad (B.64)$$

with

$$\mathcal{Q} = \text{diag}\big(e^{\sigma_\phi^2/2}, e^{(2\sigma_\phi)^2/2}, \ldots, e^{(M\sigma_\phi)^2/2}, e^{\sigma_\phi^2/2}, e^{(2\sigma_\phi)^2/2}, \ldots, e^{(M\sigma_\phi)^2/2}\big) \qquad (B.65)$$

which is the claimed result.

Bibliography

[1]   Masmicro programme website. `http://www.masmicro.net/`. (Accessed November 11th 2009).

[2]   Microned programme website. `http://www.microned.nl/`. (Accessed Februari 22th 2011).

[3]   E. Abele, M. Kreis, and M. Roth. Electromagnetic actuator for in process non-contact identification of spindle-tool frequency response functions. In *Proceedings of the CIRP 2nd International Conference on High Performance Cutting, Vancouver, BC, June 12 - 13*, 2006.

[4]   M. Aenis, E. Knopf, and R. Nordmann. Active magnetic bearings for the identification and fault diagnosis in turbomachinery. *Mechatronics*, 12:1011–1021, 2002.

[5]   B. Aeschlimann. Control aspects of high precision active magnetic bearings. *PhD Thesis, École Polytechnique Fédérale de Lausanne*, 2001.

[6]   H.-J. Ahn, S.-W. Lee, S.-H. Lee, and D.-C. Han. Frequency domain control-relevant identification of MIMO AMB rigid rotor. *Automatica*, 39:299–307, 2003.

[7]   A. Albrecht, S. S. Park, Y. Altintas, and G. Pritschow. High frequency bandwidth cutting force measurement in milling using capacitance displacement sensors. *International Journal of Machine Tools and Manufacture*, 45:993–1008, 2005.

[8]   L. Alting, F. Kimura, H. N. Hansen, and G. Bissacco. Micro engineering. *CIRP Annals*, 52(2):635–657, 2003.

[9]   Y. Altintas. *Manufacturing Automation, Metal Cutting Mechanics, Machine Tool Vibrations, and CNC Design.* Cambridge University Press, 2000.

[10]   Y. Altintas and S. S. Park. Dynamic compensation of spindle-integrated force sensors. *CIRP Annals*, 53(1):305–308, 2004.

[11]   Y. Altintas and M. Weck. Chatter stability of metal cutting and grinding. *CIRP Annals*, 53(2):619–642, 2004.

BIBLIOGRAPHY

[12] Y. M. Amraoui and N. A. J. Lieven. Noncontacting excitation and measurement of light structures. *Journal of Vibration and Acoustics*, 125:114–119, 2003.

[13] D. Anand, J. Kirk, and M. Anjanappa. Magnetic bearing spindles for enhancing tool path accuracy. *Materials and Manufacturing Processes*, 1(2):245–268, 1986.

[14] B. D. O. Anderson and J. Moore. *Optimal Filtering*. Prentice-Hall, Englewood Cliffs, New Jersey, USA, 1979.

[15] A. Aramcharoen and P. T. Mativenga. Size effect and tool geometry in micromilling of tool steel. *Precision Engineering*, 33:402 – 407, 2009.

[16] S. Auchet, P. Chevrier, M. Lacour, and P. Lipinski. A new method of cutting force measurement based on command voltages of active electro-magnetic bearings. *International Journal of Machine Tools and Manufacture*, 44:1441–1449, 2004.

[17] Y.-B. Bang, K.-M. Lee, and S. Oh. 5-axis micro milling machine for machining micro parts. *International Journal of Advanced Manufacturing Technology*, 25:888–894, 2005.

[18] W. Y. Bao and I. N. Tansel. Modeling micro-end-milling operations. Part I: analytical cutting force model. *International Journal of Machine Tools and Manufacture*, 40:2155–2173, 2000.

[19] W. Y. Bao and I. N. Tansel. Modeling micro-end-milling operations. Part II: tool runout. *International Journal of Machine Tools and Manufacture*, 40:2175–2192, 2000.

[20] W. Y. Bao and I. N. Tansel. Modeling micro-end-milling operations. Part III: influence of tool wear. *International Journal of Machine Tools and Manufacture*, 40:2193–2211, 2000.

[21] A. Baschin, P. Kahnis, and D. Biermann. Dynamic analysis of the micromilling process – influence of tool vibrations on the quality of microstructures — dynamikanalyse des mikrofräsprozesses - einfluss von werkzeugschwingungen auf die qualität von mikrostrukturen. *Materialwissenschaft und Werkstofftechnik*, 39(9):616–621, 2008.

[22] D. S. Bayard. High-order wide-band frequency domain identification using composite curve fitting. *American Control Conference 1992*, pages 3181 – 3185.

[23] D. S. Bayard. High-order multivariable transfer function curve fitting: Algorithms, sparse matrix methods and experimental results. *Automatica*, 30(9):1439–1444, 1994.

[24] S. P. Bhattacharyya. Observer design for linear systems with unknown inputs. *IEEE Transactions on Automatic Control*, 23(3):483– 484, 1978.

[25] G. Bissacco, H. N. Hansen, and J. Slunsky. Modelling the cutting edge radius size effect for force prediction in micro milling. *CIRP Annals*, 57:113–116, 2008.

[26] H. Bleuler, C. Gahler, R. Herzog, R. Larsonneur, T. Mizuno, R. Siegwart, and S.-J. Woo. Application of digital signal processors for industrial magnetic bearings. *IEEE Transactions on Control Systems Technology*, 2(4):280–289, 1994.

[27] E. Brinksmeier, R. Gläbe, O. Riemer, and S. Twardy. Potentials of precision machining processes for the manufacture of micro forming molds. *Microsystem Technologies*, 14:1983–1987, 2008.

[28] G. Byrne, D. Dornfeld, I. Inasaki, G. Ketteler, W. König, and R. Teti. Tool condition monitoring (TCM) – the status of research and industrial application. *CIRP Annals*, 44(2):541–567, 1995.

[29] P. Castellini, G. M. Revel, L. Scalise, and R. M. De Andrade. Experimental and numerical investigation on structural effects of laser pulses for modal parameter measurement. *Optics and Lasers in Engineering*, 32(6):565–581, 1999.

[30] J. Chae and S. Park. High frequency bandwidth measurements of micro cutting forces. *International Journal of Machine Tools and Manufacture*, 47:1433–1441, 2007.

[31] J. Chae, S. S. Park, and T. Freiheit. Investigation of micro-cutting operations. *International Journal of Machine Tools and Manufacture*, 46:313–332, 2006.

[32] M. Chen and C. R. Knospe. Control approaches to the suppression of machining chatter using active magnetic bearings. *IEEE Transactions on Control Systems Technology*, 15(2):220 – 232, 2007.

[33] D. K. Cheng. *Field and Wave Electromagnetics*. Addison-Wesley, Reading, MA, USA, 1989.

[34] P. Chevrier, A. Tidu, B. Bolle, P. Cezard, and J. P. Tinnes. Investigation of surface integrity in high speed end milling of a low alloyed steel. *International Journal of Machine Tools and Manufacture*, 43:1135–1142, 2003.

[35] S. Cho, S. Asfour, A. Onar, and N. Kaundinya. Tool breakage detection using support vector machine learning in a milling process. *International Journal of Machine Tools and Manufacture*, 45:241–249, 2005.

[36] Y. Choi, R. Narayanaswami, and A. Chandra. Tool wear monitoring in ramp cuts in end milling using the wavelet transform. *International Journal of Advanced Manufacturing Technology*, 23:419–428, 2004.

[37] M. Corless and J. Tu. State and input estimation for a class of uncertain systems. *Automatica*, 34(6):757–764, 1998.

[38] M. Darouach and M. Zasadzinski. Unbiased minimum variance estimation for systems with unknown exogenous inputs. *Automatica*, 33(4):717–719, 1997.

[39] M. Darouach, M. Zasadzinski, and M. Boutayeb. Extension of minimum variance estimation for systems with unknown inputs. *Automatica*, 39:867–876, 2003.

[40] M. Darouach, M. Zasadzinski, and S. J. Xu. Full-order observers for linear systems with unknown inputs. *IEEE Transactions on Automatic Control*, 39(3):606 – 609, 1994.

[41] R. De Callafon, D. De Roover, and P. M. J. Van den Hof. Multivariable least squares frequency domain identification using polynomial matrix fraction descriptions. *Proceedings of the 35th IEEE Conference on Decision and Control, Kobe, Japan*, 2:2030–2035, 1996.

[42] T. P. Dobrowiecki and J. Schoukens. Measuring a linear approximation to weakly nonlinear MIMO systems. *Automatica*, 43(10):1737–1751, 2007.

Bɪʙʟɪᴏɢʀᴀᴘʜʏ

[43]   T. P. Dobrowiecki, J. Schoukens, and P. Guillaume. Optimized excitation signals for MIMO frequency response function measurements. *IEEE Transactions on Instrumentation and Measurement*, 55(6):2072–2079, 2006.

[44]   D. Dornfeld, S. Min, and Y. Takeuchi. Recent advances in mechanical micromachining. *CIRP Annals*, 55(2):745–768, 2006.

[45]   T. A. Dow, E. L. Miller, and K. Garrard. Tool force and deflection compensation for small milling tools. *Precision Engineering*, 28:31–45, 2004.

[46]   S. P. Ebenezer, A. Papandreou-Suppappola, and S. B. Suppappola. Matching pursuit classification for time-varying acoustic emissions. *Proceedings of the 35th Asilomar Conference on Signals, Systems and Computers, California USA*, 1:715 – 719, 2001.

[47]   K. F. Ehmann. Ultra high-speed micro-spindle for micro/meso-scale machine tools, research description. `http://www.mech.northwestern.edu/MFG/AML/research/cur_spindle.html`. (Accessed Februari 22th 2011).

[48]   K. F. Ehmann, D. Bourell, M. L. Culpepper, T. J. Hodgson, T. R. Kurfess, M. Madou, K. Rajurkar, and R. E. DeVor. International assessment of research and development in micromanufacturing. Technical report, WTEC, 2005.

[49]   R. P. H. Faassen, N. van de Wouw, J. A. J. Oosterling, and H. Nijmeijer. Prediction of regenerative chatter by modelling and analysis of high-speed milling. *International Journal of Machine Tools and Manufacture*, 43:1437–1446, 2003.

[50]   S. Filiz, C. M. Conley, M. B. Wasserman, and O. B. Ozdoganlar. An experimental investigation of micro-machinability of copper 101 using tungsten carbide micro-endmills. *International Journal of Machine Tools and Manufacture*, 47:1088–1100, 2007.

[51]   R. L. Fittro, C. R. Knospe, and L. S. Stephens. μ-synthesis applied to the compliance minimization of an active magnetic bearing HSM spindle's thrust axis. *Machining Science and Technology*, 7:19–51, 2003.

[52]   B. Friedland. Treatment of bias in recursive filtering. *IEEE Transactions on Automatic Control*, 14(4):359– 367, 1969.

[53]   C. Gähler, M. Mohler, and R. Herzog. Multivariable identification of active magnetic bearing systems. *JSME International Journal Series C: Mechanical Systems Machine Elements and Manufacturing*, 40(4):584–592, 1997.

[54]   S. V. Gaikwad and D. E. Rivera. Multivariable frequency-response curve fitting with application to control-relevant parameter estimation. *Automatica*, 33(6):1169–1174, 1997.

[55]   J. M. Gere and S. P. Timoshenko. *Mechanics of Materials*. Chapman & Hall, London, 1995.

[56]   A. S. Ghersin, R. S. Smith, and R. S. Sánchez Peña. Classical, robust and LPV control of a magnetic-bearing experiment. *Identification and Control: The Gap Between Theory and Practice*, pages 277–325, 2007.

[57]   D. D. Gill, B. Jokiel Jr., J. C. Ziegert, S. W. T. Payne, and J. P. Pathak. Next generation spindles for micromilling. *Sandia Report SAND2004-6445*, 2004.

[58] S. Gillijns and B. De Moor. Unbiased minimum-variance input and state estimation for linear discrete-time systems. *Automatica*, 43:111–116, 2007.

[59] S. Gillijns and B. De Moor. Unbiased minimum-variance input and state estimation for linear discrete-time systems with direct feedthrough. *Automatica*, 43:934–937, 2007.

[60] M. J. Grimble. Polynomial matrix solutions of the H$_\infty$ filtering problem and the relationship to Ricatti equation state-space results. *IEEE Transactions on Signal Processing*, 41:67–81, 1993.

[61] Y. Guan and M. Saif. A novel approach to the design of unknown input observers. *IEEE Transactions on Automatic Control*, 36(5):632 – 635, 1991.

[62] P. Guillaume. Frequency response measurements of multivariable systems using nonlinear averaging techniques. *IEEE Transactions on Instrumentation and Measurement*, 47(3):796–800, 1998.

[63] P. Guillaume, R. Pintelon, and J. Schoukens. Nonparametric frequency response function estimators based on nonlinear averaging techniques. *IEEE Transactions on Instrumentation and Measurement*, 41(6):739–746, 1992.

[64] P. Guillaume, R. Pintelon, and J. Schoukens. Accurate estimation of multivariable frequency response functions. *Proceedings of the 13th IFAC World congress, San Francisco, USA*, pages 423–428, 1996.

[65] B. Hassibi, A. T. Erdogan, and T. Kailath. MIMO linear equalization with an H$^\infty$ criterion. *IEEE Transactions on Signal Processing*, 54(2):499– 511, 2006.

[66] B. Hassibi, A. H. Sayed, and T. Kailath. *Indefinite-Quadratic Estimation and Control, A Unified Approach to H^2 and H$^\infty$ Theories*. SIAM Studies in Applied and Numerical Mathematics, Philadelphia, PA (USA), 1999.

[67] A. Hassui, A. E. Diniz, J. F. G. Oliveira, J. Felipe, and J. J. F. Gomes. Experimental evaluation on grinding wheel wear through vibration and acoustic emission. *Wear*, 217:7–14, 1998.

[68] G. H. Hostetter and J. S. Meditch. Generalization of observers to systems with unmeasurable, unknown inputs. *Automatica*, 9:721–724, 1973.

[69] M. Hou and P. C. Müller. Design of observers for linear systems with unknown inputs. *IEEE Transactions on Automatic Control*, 37(6):871–875, 1992.

[70] M. Hou and R. J. Patton. Input observability and input reconstruction. *Automatica*, 34(6):789–794, 1998.

[71] M. Hou and R. J. Patton. Optimal filtering for systems with unknown inputs. *IEEE Transactions on Automatic Control*, 43(3):446–449, 1998.

[72] C.-S. Hsieh. Robust two-stage Kalman filters for systems with unknown inputs. *IEEE Transactions on Automatic Control*, 45(12):2374 – 2378, 2000.

BIBLIOGRAPHY

[73] D. Huo, K. Cheng, and F. Wardle. Design of a five-axis ultra-precision micro-milling machine — UltraMill. Part 1: holistic design approach, design considerations and specifications. *International Journal of Advanced Manufacturing Technology*, 47:879–890, 2009.

[74] D. Huo, K. Cheng, and F. Wardle. Design of a five-axis ultra-precision micro-milling machine — UltraMill. Part 2: integrated dynamic modelling, design optimisation and analysis. *International Journal of Advanced Manufacturing Technology*, 47:867–877, 2009.

[75] N. Ikawa, S. Shimada, and H. Tanaka. Minimum thickness of cut in micromachining. *Nanotechnology*, 3:6–9, 1992.

[76] K. Jemielniak and P. J. Arrazola. Application of AE and cutting force signals in tool condition monitoring in micro-milling. *CIRP Journal of Manufacturing Science and Technology*, 1:97–102, 2008.

[77] K. Jemielniak, S. Bombiński, and P. X. Aristimuno. Tool condition monitoring in micromilling based on hierarchical integration of signal measures. *CIRP Annals*, 57:121–124, 2008.

[78] C. D. Johnson. On observers for systems with unknown and inaccessible inputs. *International Journal of Control*, 21(5):825–831, 1975.

[79] C. D. Johnson. New results on the inverse-system/deconvolution problem for linear dynamical systems. *Circuits Systems Signal Process*, 19(4):365–383, 2000.

[80] M. B. G. Jun, R. E. DeVor, and S. G. Kapoor. Investigation of the dynamics of microend milling – Part II: Model validation and interpretation. *Journal of Manufacturing Science and Engineering*, 128:901–912, 2006.

[81] M. B. G. Jun, X. Liu, R. E. DeVor, and S. G. Kapoor. Investigation of the dynamics of microend milling – Part I: Model development. *Journal of Manufacturing Science and Engineering*, 128(4):893–900, 2006.

[82] T. Kailath. *Linear systems*. Prentice-Hall, Englewood Cliffs, New Jersey, USA, 1980.

[83] T. Kailath, A. H. Sayed, and B. Hassibi. *Linear Estimation*. Prentice Hall, 2000.

[84] E.-G. Kang, S.-J. Park, and S.-J. Lee. Development of in situ system to monitor the machining process using a piezo load cell. *International Journal of Advanced Manufacturing Technology*, 25(7-8):647–651, 2005.

[85] B. Karpuschewski. *Sensoren zur prozeßüberwachung beim Spanen*. Fortschritt-Berichte VDI, Reihe 2, Nr. 581, 2001.

[86] M. G. Kendall and A. Stuart. *The Advanced Theory of Statistics*. Charles Griffin & Company, London, 1963.

[87] C.-J. Kim, J. R. Mayor, and J. Ni. A static model of chip formation in microscale milling. *Journal of Manufacturing Science and Engineering*, 126:710–718, 2004.

[88] J.-D. Kim and D. S. Kim. Theoretical-analysis of micro-cutting characteristics in ultra-precision machining. *Journal of Materials Processing Technology*, 49:387–398, 1995.

[89] M. H. Kimman, H. H. Langen, J. Van Eijk, and R. H. Munnig Schmidt. Design and realization of a miniature spindle test setup with active magnetic bearings. *IEEE/ASME International Conference on Advanced Intelligent Mechatronics, AIM*, 2007.

[90] C. R. Knospe. Active magnetic bearings for machining applications. *Control Engineering Practice*, 15:307–313, 2007.

[91] E. Krämer. *Dynamics of Rotors and Foundations*. Springer-Verlag, Berlin, 1993.

[92] P. Kudva, N. Viswanadham, and A. Ramakrishna. Observers for linear systems with unknown inputs. *IEEE Transactions on Automatic Control*, 25:113– 115, 1980.

[93] J.-H. Kyung and C.-W. Lee. Controller design for a magnetically suspended milling spindle based on chatter stability analysis. *JSME International Journal Series C: Mechanical Systems Machine Elements and Manufacturing*, 46(2):416–422, 2003.

[94] M. Lalanne and G. Ferraris. *Rotordynamics, Prediction in Engineering*. John Wiley & Sons, Chichester, England, 1990.

[95] H. H. Langen. Microfactory research topics in the Netherlands. *Proceedings of the 5th International Workshop on Microfactories, Besancon, France, Oct. 25-27*, 2006.

[96] H. U. Lee, D.-W. Cho, and K. F. Ehmann. A mechanistic model of cutting forces in micro-end-milling with cutting-condition-independent cutting force coefficients. *Journal of Manufacturing Science and Engineering*, 130:031102/1–031102/9, 2008.

[97] K. Lee and D. A. Dornfeld. Micro-burr formation and minimization through process control. *Precision Engineering*, 29:246–252, 2005.

[98] P. Li. *Micromilling of Hardened Tool Steels*. PhD thesis, Delft University of Technology, 2009.

[99] P. Li, R. S. Blom, J. A. J. Oosterling, and M. Achtsnick. Design requirements of an advanced micromilling setup. *Proceedings of the 4th International Conference on Manufacturing Research, Liverpool UK, September 5-7*, 2006.

[100] X. Li. On-line detection of the breakage of small diameter drills using current signature wavelet transform. *International Journal of Machine Tools and Manufacture*, 39:157–164, 1999.

[101] X. Li. A brief review: acoustic emission method for tool wear monitoring during turning. *International Journal of Machine Tools and Manufacture*, 42:157–165, 2002.

[102] X. Li, H.-X. Li, X.-P. Guan, and R. Du. Fuzzy estimation of feed-cutting force from current measurement-a case study on intelligent tool wear condition monitoring. *IEEE Transactions on Systems, Man, and Cybernetics Part C: Applications and Reviews*, 34(4):506–512, 2004.

[103] K. Liu, T. Waumans, J. Peirs, and D. Reynaerts. Precision manufacturing of key components for an ultra miniature gas turbine unit for power generation. *Microsystem Technologies*, 15:1417–1425, 2009.

[104] X. Liu, R. E. DeVor, S. G. Kapoor, and K. F. Ehmann. The mechanics of machining at the microscale: Assessment of the current state of the science. *Journal of Manufacturing Science and Engineering*, 126:666–678, 2004.

BIBLIOGRAPHY

[105]   L. Ljung. *System Identification, Theory for the User*. Prentice Hall, Upper Saddle River, NJ, USA, 1999.

[106]   L. N. López de Lacalle and A. Lamikiz, editors. *Machine Tools for High Performance Machining*. Springer, London, 2009.

[107]   F. Lösch. *Identification and Automated Controller Design for Active Magnetic Bearing Systems*. PhD thesis, ETH Zürich, 2002.

[108]   K. Mahata, R. Pintelon, and J. Schoukens. On parameter estimation using nonparametric noise models. *IEEE Trans. Autom. Control*, 51(10):1602–1612, 2006.

[109]   M. Malekian, S. S. Park, and M. B. G. Jun. Tool wear monitoring of micro-milling operations. *Journal of Materials Processing Technology*, 209:4903–4914, 2009.

[110]   K. Marguerre and K. Wölfel. *Mechanics of Vibration*. Sijthoff & Noordhoff, Alphen aan den Rijn, The Netherlands, 1979.

[111]   T. Masuzawa. State of the art of micromachining. *CIRP Annals*, 49(2):473–488, 2000.

[112]   T. Masuzawa and H. Tönshoff. Three-dimensional micromachining by machine tools. *CIRP Annals*, 46(2):621–628, 1997.

[113]   J. S. Meditch and G. H. Hostetter. Observers for systems with unknown and inaccessible inputs. *1973 IEEE Conference on Decision and Control, including the 12th Symposium on Adaptive Processes*, 12:120 – 124, 1973.

[114]   A. J. Mian, N. Driver, and P. T. Mativenga. Micromachining of coarse-grained multiphase material. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 223(4):377–385, 2009.

[115]   R. Mohd-Mokhtar, L. Wang, L. Qin, and T. Barry. Continuous time system identification of magnetic bearing systems using frequency response data. *Proceedings of the 5th Asian Control Conference*, 3:2066– 2072, 2004.

[116]   M. K. Müller. *On-line Process Monitoring in High Speed Milling with an Active Magnetic Bearing Spindle*. PhD thesis, ETH Zürich, 2002.

[117]   K. M. Nagpal and P. P. Khargonekar. Filtering and smoothing in an H_∞ setting. *IEEE Transactions on Automatic Control*, 36(2):152–166, 1991.

[118]   G. Obinata and B. D. O. Anderson. *Model reduction for control system design*. Springer, London, 2001.

[119]   Y. Okazaki, N. Mishima, and K. Ashida. Microfactory – concept, history, and developments. *Journal of Manufacturing Science and Engineering*, 126(4):837–844, 2004.

[120]   A. Oppenheim and R. Schafer. *Digital Signal Processing*. Prentice-Hall, Englewood Cliffs, NJ, USA, 1975.

[121]   S. S. Park. Identification of spindle integrated force sensor's transfer function for modular end mills. *Journal of Manufacturing Science and Engineering*, 128:146–153, 2006.

[122]   S. S. Park and Y. Altintas. Dynamic compensation of spindle integrated force sensors with Kalman filter. *Journal of Dynamic Systems, Measurement, and Control*, 126:443–452, 2004.

[123] S. S. Park, Y. Altintas, and M. Movahhedy. Receptance coupling for end mills. *International Journal of Machine Tools and Manufacture*, 43:889–896, 2003.

[124] Y. Park and J. L. Stein. Closed-loop, state and input observer for systems with unknown inputs. *International Journal of Control*, 48(3):1121–1136, 1988.

[125] P. J. Parker and B. D. O. Anderson. Frequency tracking of nonsinusoidal periodic signals in noise. *Signal Processing*, 20:127–152, 1990.

[126] R. Pintelon, Y. Rolain, A. Bultheel, and M. Van Barel. Frequency domain identification of multivariable systems using vector orthogonal polynomials. *Proceedings of the 16th International Symposium on Mathematical Theory of Networks and Systems, Leuven (Belgium)*, 2004.

[127] R. Pintelon and J. Schoukens. Measurement of frequency response functions using periodic excitations, corrupted by correlated input/output errors. *IEEE Transactions on Instrumentation and Measurement*, 50(6):1753–1760, 2001.

[128] R. Pintelon and J. Schoukens. *System identification: a frequency domain approach.* IEEE press, Piscataway, NJ, USA, 2001.

[129] R. Pintelon, J. Schoukens, G. Vandersteen, and K. Barbe. Estimation of nonparametric noise and frf models for multivariable systems-Part I: Theory. *Mechanical Systems and Signal Processing*, 24(3):573–595, 2010.

[130] R. Pintelon, J. Schoukens, G. Vandersteen, and K. Barbe. Estimation of nonparametric noise and frf models for multivariable systems-Part II: Extensions, applications. *Mechanical Systems and Signal Processing*, 24(3):596–616, 2010.

[131] R. Pintelon, G. Vandersteen, L. De Locht, Y. Rolain, and J. Schoukens. Experimental characterization of operational amplifiers: a system identification approach-part i: theory and simulations. *IEEE Trans. Instrum. Meas.*, 53(3):854– 862, 2004.

[132] P. W. Prickett and C. Johns. An overview of approaches to end milling tool monitoring. *International Journal of Machine Tools and Manufacture*, 39:105–122, 1999.

[133] M. B. Priestley. *Spectral analysis and time series.* Academic Press, San Diego, CA, USA, 1989.

[134] M. Rahman, A. S. Kumar, and J. R. S. Prakash. Micro milling of pure copper. *Journal of Materials Processing Technology*, 116:39–43, 2001.

[135] S. S. Rangwala and D. A. Dornfeld. Learning and optimization of machining operations using computing abilities of neural networks. *IEEE Transactions on Systems, Man and Cybernetics*, 19(2):299–314, 1989.

[136] Y. Ren and C. F. Beards. Identification of joint properties of a structure using FRF data. *Journal of Sound and Vibration*, 186(4):567–587, 1995.

[137] Y. Rolain, R. Pintelon, K. Q. Xu, and H. Vold. Best conditioned parametric identification of transfer function models in the frequency domain. *IEEE Transactions on Automatic Control*, 40(11):1954–1960, 1995.

BIBLIOGRAPHY

[138] M. A. Salgado, L. N. López De Lacalle, A. Lamikiz, J. Muñoa, and J. A. Sánchez. Evaluation of the stiffness chain on the deflection of end-mills under cutting forces. *International Journal of Machine Tools and Manufacture*, 45:727–739, 2005.

[139] S. G. Sanathanan and J. Koerner. Transfer function synthesis as a ratio of two complex polynomials. *IEEE Transactions on Automatatic Control*, pages 56–58, 1963.

[140] A. Sayed and T. Kailath. A survey of spectral factorization methods. *Numerical Linear Algebra with Applications*, 8:467–496, 2001.

[141] T. Schaller, L. Bohn, J. Mayer, and K. Schubert. Microstructure grooves with a width of less than 50 μm cut with ground hard metal micro end mills. *Precision Engineering*, 23:229–235, 1999.

[142] J. Schmidt, D. Spath, J. Elsner, V. Hüntrup, and H. Tritschler. Requirements of an industrially applicable microcutting process for steel micro-structures. *Microsystem Technologies*, 8:402–408, 2002.

[143] J. Schmidt and H. Tritschler. Micro cutting of steel. *Microsystem Technologies*, 10:167–174, 2004.

[144] T. L. Schmitz, M. A. Davies, and M. D. Kennedy. Tool point frequency response prediction for high-speed machining by RCSA. *Journal of Manufacturing Science and Engineering*, 123:700–707, 2001.

[145] J. Schoukens, R. Pintelon, T. P. Dobrowiecki, and Y. Rolain. Identification of linear systems with nonlinear distortions. *Automatica*, 41:491–504, 2005.

[146] J. Schoukens, R. Pintelon, and Y. Rolain. Box-Jenkins alike identification using non-parametric noise models. *Automatica*, 40:2083–2089, 2004.

[147] J. Schoukens, R. Pintelon, G. Vandersteen, and P. Guillaume. Frequency-domain system identification using non-parametric noise models estimated from a small number of data sets. *Automatica*, 33(6):1073–1086, 1997.

[148] G. Schweitzer, H. Bleuler, and A. Traxler. *Active Magnetic Bearings, Basics, Properties and Applications of Active Magnetic Bearings*. Hochschulverlag AG an der ETH Zürich, 1994.

[149] G. Schweitzer and E. H. Maslen, editors. *Magnetic Bearings, Theory, Design, and Application to Rotating Machinery*. Springer-Verlag, Berlin Heidelberg, 2009.

[150] M. M. Seron, J. H. Braslavsky, and G. C. Goodwin. *Fundamental Limitations in Filtering and Control*. Springer-Verlag, London, Great Britain, 1997.

[151] H. Shinno, H. Hashizume, and H. Yoshioka. Sensor-less monitoring of cutting force during ultraprecision machining. *CIRP Annals*, 52(1):303–306, 2003.

[152] T. Söderström. Errors-in-variables methods in system identification. *Automatica*, 43:939–958, 2007.

[153] M. Szafarczyk, editor. *Automatic Supervision in Manufacturing*. Springer, London, 1994.

[154] I. N. Tansel, T. T. Arkan, W. Y. Bao, N. Mahendrakar, B. Shisler, D. Smith, and M. Mc-Cool. Tool wear estimation in micro-machining. Part I: tool usage-cutting force relationship. *International Journal of Machine Tools and Manufacture*, 40:599–608, 2000.

[155] I. N. Tansel, T. T. Arkan, W. Y. Bao, N. Mahendrakar, B. Shisler, D. Smith, and M. Mc-Cool. Tool wear estimation in micro-machining. Part II: neural-network-based periodic inspector for nonmetals. *International Journal of Machine Tools and Manufacture*, 40:609–620, 2000.

[156] I. N. Tansel, A. Nedbouyan, M. Trujillo, and B. Tansel. Micro-end-milling – II. extending tool life with a smart workpiece holder (SWH). *International Journal of Machine Tools and Manufacture*, 38:1437–1448, 1998.

[157] I. N. Tansel, O. Rodriguez, M. Trujillo, E. Paz, and W. Li. Micro-end-milling – I. wear and breakage. *International Journal of Machine Tools and Manufacture*, 38:1419–1436, 1998.

[158] I. N. Tansel, M. Trujillo, A. Nedbouyan, C. Velez, W. Y. Bao, T. T. Arkan, and B. Tansel. Micro-end-milling – III. wear estimation and tool breakage detection using acoustic emission signals. *International Journal of Machine Tools and Manufacture*, 38:1449–1466, 1998.

[159] H. K. Tönshoff, T. Friemuth, and J. C. Becker. Process monitoring in grinding. *CIRP Annals*, 51(2):551–571, 2002.

[160] H. K. Tönshoff, M. Jung, S. Männel, and W. Rietz. Using acoustic emission signals for monitoring of production processes. *Ultrasonics*, 37:681–686, 2000.

[161] H. K. Tönshoff, X. Li, and C. Lapp. Application of fast Haar transform and concurrent learning to tool-breakage detection in milling. *IEEE/ASME Transactions on Mechatronics*, 8(3):414–418, 2003.

[162] R. Tuokko. Views to micro and desktop factories – from miniaturization of products to miniaturization of production equipment and factories. *Proceedings of the 2nd Philips Precision Technology Conference, Hilvarenbeek, The Netherlands*, pages 7–14, 2006.

[163] L. Uriarte, A. Herrero, M. Zatarain, G. Santiso, L. N. Lopéz de Lacalle, A. Lamikiz, and J. Albizuri. Error budget and stiffness chain assessment in a micromilling machine equipped with tools less than 0.3 mm in diameter. *Precision Engineering*, 31:1–12, 2007.

[164] P. M. J. Van den Hof. Closed-loop issues in system identification. *Annual reviews in control*, 22:173–186, 1998.

[165] P. M. J. Van den Hof and S. G. Douma. An IV-based iterative linear regression algorithm with optimal output error properties. *DCSC Technical report nr. 09-018, Delft University of Technology*, 2008.

[166] N. van Dijk. *Active chatter control in high-speed milling processes.* PhD thesis, Eindhoven University of Technology, 2011.

[167] K. Vanhoenacker, J. Schoukens, J. Swevers, and D. Vaes. Summary and comparing overview of techniques for the detection of non-linear distortions. *Proceedings of the ISMA 2002*, 2:1241–1256, 2002.

BIBLIOGRAPHY

[168] M. J. Vasile, C. R. Friedrich, B. Kikkeri, and R. McElhannon. Micrometer-scale machining: Tool fabrication and initial results. *Precision Engineering*, 19:180–186, 1996.

[169] P. Verboven. *Frequency-Domain System Identification for Modal Analysis*. PhD thesis, Vrije Universiteit Brussel, 2002.

[170] P. Verboven, P. Guillaume, and B. Cauberghe. Multivariable frequency–response curve fitting with application to modal parameter estimation. *Automatica*, 41:1773–1782, 2005.

[171] M. P. Vogler, R. E. DeVor, and S. G. Kapoor. Microstructure-level force prediction model for micro-milling of multi-phase materials. *Journal of Manufacturing Science and Engineering*, 125:202–209, 2003.

[172] M. P. Vogler, R. E. DeVor, and S. G. Kapoor. On the modeling and analysis of machining performance in micro-endmilling, Part I: Surface generation. *Journal of Manufacturing Science and Engineering*, 126(4):685–694, 2004.

[173] L. Wang and R. X. Gao, editors. *Condition Monitoring and Control for Intelligent Manufacturing*. Springer, London, 2006.

[174] S.-H. Wang, E. J. Davison, and P. Dorato. Observing the states of systems with unmeasurable disturbances. *IEEE Transactions on Automatic Control*, 20:716 – 717, 1975.

[175] P. E. Wellstead. Nonparametric methods of system-identification. *Automatica*, 17(1):55–69, 1981.

[176] E. Wernholt and S. Gunnarsson. Analysis of methods for multivariable frequency response function estimation in closed loop. *Proceedings of the 46th IEEE Conference on Decision and Control, New Orleans, December 12-14*, pages 4881–4888, 2007.

[177] E. Wernholt and S. Gunnarsson. Estimation of nonlinear effects in frequency domain identification of industrial robots. *IEEE Transactions on Instrumentation and Measurement*, 57(4):856–863, 2008.

[178] E. Wernholt and S. Moberg. Experimental comparison of methods for multivariable frequency response function estimation. *Proceedings of the 17th IFAC World congress, Seoul, Korea, July 6–11, 2008*, pages 15359–15366, 2008.

[179] H. Weule, V. Hüntrup, and H. Tritschler. Micro-cutting of steel to meet new requirements in miniaturization. *CIRP Annals*, 50(1):61–64, 2001.

[180] N. Wiener and E. Hopf. On a class of singular integral equations. *Proc. Prussian Acad. Math. — Phys. Ser.*, page 696, 1931.

[181] F. Yang and R. W. Wilde. Observers for linear systems with unknown inputs. *IEEE Transactions on Automatic Control*, 33(7):677 – 681, 1988.

[182] I. Yesilyurt. End mill breakage detection using mean frequency analysis of scalogram. *International Journal of Machine Tools and Manufacture*, 46:450–458, 2006.

[183] P. Young. Some observations on instrumental variable methods of time-series analysis. *International Journal of Control*, 23(5):593–612, 1976.

[184] P. Young, H. Garnier, and M. Gilson. *Identification of Continuous-time Models from Sampled Data*, chapter Refined instrumental variable identification of continuous-time hybrid Box-Jenkins models, pages 91–131. Springer, London, UK, 2008.

[185] M. T. Zaman, A. Senthil Kumar, M. Rahman, and S. Sreeram. A three-dimensional analytical cutting force model for micro end milling operation. *International Journal of Machine Tools and Manufacture*, 46:353–366, 2006.

[186] K. P. Zhu, G. S. Hong, and Y. S. Wong. A comparative study of feature selection for hidden markov model-based micro-milling tool wear monitoring. *Machining Science and Technology*, 12(3):348–369, 2008.

[187] K. P. Zhu, Y. S. Wong, and G. S. Hong. Multi-category micro-milling tool wear monitoring with continuous hidden markov models. *Mechanical Systems and Signal Processing*, 23(2):547–560, 2009.
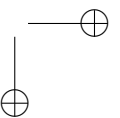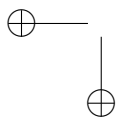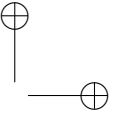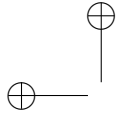
Summary

**Model-based Process Monitoring and Control of Micro-milling
using Active Magnetic Bearings**

Rogier S. Blom

The process of micro-milling is a promising technology for the fabrication of micro-parts with arbitrary 3D features in a wide range of materials. However, as a result of the reduced dimensions, the susceptibility of the process for machine tool errors and vibrations is higher, having adverse effects on accuracy and surface quality of the resulting workpieces. Furthermore, the production time and the efficiency of the process suffer from low material removal rates and excessive tool wear and breakage. To improve the micro-milling process, online process monitoring and control becomes of high importance. Signs of problems are almost unnoticeable without the use of special equipment. Techniques are needed to detect and possibly even predict anomalies in the process and to online monitor the condition of the cutting process.

Spindles with Active Magnetic Bearings are particularly interesting for the micro-milling process, not only for the achievable spindle speeds, but also because of the opportunities they offer to develop online process monitoring and control techniques. These include force monitoring, tool condition and breakage monitoring, and chatter control. However, literature thus far lacks results implementing these techniques for the micro-milling process.

The aim of this thesis is to investigate the opportunities for model-based process monitoring and control to improve the micro-milling process using the intrinsic properties of AMB spindles. This objective is narrowed down to the goal of estimating the cutting forces from the bearing signals. The approach towards this goal consists of three steps.

First an approach to model-based cutting force estimation in micro-milling us-

ing the signals of the AMBs is developed. The cutting force estimation problem is expressed as an input estimation problem, where the cutting forces are an unknown input to the closed-loop AMB spindle system. To solve this problem, a method is given for model-based optimal estimation of unknown inputs to multivariable closed-loop systems, based on Wiener filter theory. For cases in which controller knowledge is not available, an approach is formulated in which equal performance of the estimator is ensured for any controller. Smoothed estimators are derived, resulting in smaller estimation errors when a delay in the estimation result is tolerable.

Second, a method is presented for system identification of a high speed AMB micro-milling spindle in the frequency range relevant for force estimation. This problem is separated into two subproblems. The first is the identification of the dynamics from the current input to the displacement of the rotor shaft at the bearings, the *bearing dynamics*. The second problem is the identification of the *tooltip dynamics*, which are the dynamics between the force on the tooltip and the displacement of the rotor shaft at the bearings.

System identification of the bearing dynamics is approached by first making a non-parametric estimate of the multivariable frequency response function (FRF). An experiment design is given targeted at yielding small bias and variance of the FRF, as well as small error due to nonlinear distortions. Using the FRF estimate, a multivariable parametric model is estimated. Here, the main emphasis is on identification of a parametric model of the plant dynamics, leading to the choice of minimization of an output error (OE) criterion. An IV-based algorithm is given for estimation of multi-input multi-outout (MIMO) Output Error models in matrix fraction description from frequency domain data. This algorithm has the property that convergence of the iterations implies that an optimal solutions has been found.

The main challenge in identifying the tooltip dynamics is to apply a known excitation force to the tooltip. The route followed in this thesis is to identify the tooltip dynamics using data obtained during a milling experiment in which the cutting forces are measured. The amount of data that can be generated in this way is limited, as is the control over the spectral properties of the input. Hence, in order to reduce the complexity of the identification, usage is made of assumed observability and controllability properties of the system. This results in a particular closed-loop parameterization of the model-set and a known, but non-minimum phase noise model. For this particular identification problem, solutions are formulated.

The third and last step to the goal of this thesis pertains to model-based correction of runout disturbances in measurements of the positions and currents of AMB spindle. Such disturbances are synchronous with the rotation of the spindle and hence almost periodic. A parametrized truncated Fourier series expansion model for the runout disturbance as a function of the angular position is used, allowing to formulate runout identification as a parameter estimation problem. In correcting for the runout disturbances, the main issue is how to deal with the uncertainty in the angular position measurements, or the total lack of such measurements. So-

lutions are given that compensate for the errors introduced by this uncertainty, or estimate the angular position from the available data using an Extended Kalman filter approach.
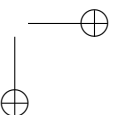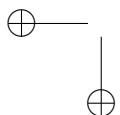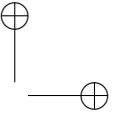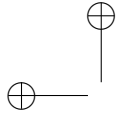
Samenvatting

**Model-gebaseerde procesmonitoring en -regeling van microfrezen
met actieve magneetlagering**

Rogier S. Blom

Het proces van micro-frezen is een veelbelovende technologie voor het fabriceren van micro-componenten met willekeurige 3D aspecten in een breed scala van materialen. Als gevolg van de gereduceerde dimensies is het proces gevoeliger voor fouten en trillingen in de bewerkingsmachine, met alle nadelige consequenties voor de nauwkeurigheid en de oppervlaktekwaliteit van de werkstukken vandien. Daarnaast hebben een laag verspaningsvolume per tijdseenheid en overmatige gereedschapsslijtage en freesbreuk een nadelig effect op de productietijd en de efficiëntie van het proces. Online procesmonitoring en -regeling is van wezenlijk belang om het micro-freesproces te verbeteren. Indicatoren van problemen zijn nagenoeg onmogelijk waar te nemen zonder speciale instrumentatie. Er zijn technieken nodig om afwijkingen in het proces op te merken en waar mogelijk zelfs te voorspellen, en om de conditie van het freesproces online te monitoren.

Spillen met actieve magneetlagering ('Active Magnetic Bearings' — AMBs) hebben eigenschappen die hen bijzonder interessant maken voor toepassing voor het micro-freesproces. Dit is niet alleen vanwege de hoge rotatiesnelheden die gerealiseerd kunnen worden, maar ook vanwege de mogelijkheden voor het verrichten van online procesmonitoring en -regeling. Hierbij kan gedacht worden aan krachtschatting, monitoring van gereedschapsslijtage en freesbreuk, en regeling van chatter. Echter, de literatuur biedt op dit moment nog geen resultaten die deze mogelijkheden toepassen op het micro-frees proces.

Het doel van dit proefschrift is om de mogelijkheden te onderzoeken voor model-gebaseerde procesmonitoring en -regeling van het micro-freesproces, gebruikmakend van de intrinsieke eigenschappen van spillen met actieve magneetlagering.

Deze doelstelling is ingeperkt tot het doel om de freeskrachten te schatten uit de lagersignalen. Er is gekozen voor een aanpak bestaande uit drie stappen.

Als eerste is een methode ontwikkeld voor model-gebaseerde schatting van frees-krachten in micro-frezen, gebruikmakend van de signalen van de AMBs. Hiertoe is het krachtschattingsprobleem geformuleerd als een probleem van het schatten van een onbekende ingang, waarbij de freeskrachten voorgesteld worden als een onbekende ingang op het gesloten-lus AMB spilsysteem. Om dit probleem op te lossen, is een model-gebaseerde optimale ingangsschattingstechniek gepresenteerd die gebaseerd is op de Wiener filter theorie. Voor de situatie waarin geen kennis van de regelaar beschikbaar is, wordt een oplossing gegeven waarbij gelijke presta-tie van de schatter wordt gegarandeerd voor elke willekeurige regelaar. Oplossingen voor vertraagde schatters worden gegeven, die resulteren in lagere schattingsfouten wanneer het toelaatbaar is dat resultaten later beschikbaar komen.

De tweede stap is een techniek voor het identificeren van een hoge snelheids AMB spil in het frequentiegebied relevant voor freeskrachtschatting. Dit probleem is verdeeld in twee deelproblemen. Het eerste omvat identificatie van de dynamica van de stroomingang naar de verplaatsing van de rotoras ter plaatse van de lagers: de lagerdynamica ('bearing dynamics'). Het tweede deelprobleem is het identifice-ren van de dynamica van de krachten op de freespunt naar de verplaatsing van de rotoras ter plaatse van de lagers: de freespuntdynamica ('tooltip dynamics').

De gekozen aanpak voor systeemidentificatie van de lagerdynamica is door eerst een niet-parametrische schatting te maken van de multivariabele frequentierespon-sie. Een experimentontwerp is geformuleerd met als doel om zowel systematische als toevallige fouten in de schatting te minimaleren, alsmede fouten ten gevolge van niet-lineariteiten. Gebruikmakend van de geschatte frequentieresponsie wordt ver-volgens een parametrisch model geschat. Hier ligt de nadruk op het schatten van een parametrisch model van de systeemdynamica (en niet op de ruisdynamica), hetgeen resulteert in de keuze voor het minimaliseren van een Output Error (OE) criterium. Een instrumenteel variabele (IV) gebaseerd algoritme wordt gegeven voor het schatten van multi-input multi-output (MIMO) OE modellen in matrixfractie-beschrijving. Dit algoritme heeft de eigenschap dat convergentie impliceert dat een optimale oplossing verkregen is.

De voornaamste uitdaging bij het identificeren van de freespuntdynamica is om een bekende excitatiekracht aan te brengen. De route die gevolgd is in dit proefschrift is om de freespuntdynamica te identificeren uit data verkregen in een freesexperiment waarin de freeskrachten direct gemeten zijn. De hoeveelheid da-ta die op deze wijze gegenereerd kan worden is gelimiteerd. Daarnaast kan men ook beperkt invloed uitoefenen op de spectrale eigenschappen van het excitatie-signaal. Rekening houdend hiermee, wordt gebruik gemaakt van veronderstel-de observeerbaarheids- en regelbaarheidseigenschappen van het systeem, teneinde de complexiteit van het identificatievraagstuk te reduceren. Dit resulteert in een specifieke gesloten-lus parametrisatie van de model-set en een bekend, maar niet-minimum fase ruismodel. Voor dit identificatieprobleem is een oplossing geformu-

leerd.

De derde en laatste stap naar de doelstelling van het proefschrift heeft betrekking op model-gebaseerde correctie van runoutverstoringen in de metingen van de stroom- en positiesignalen van de AMB spil. Dergelijke verstoringen zijn synchroon met de rotatie van de spil en dus quasi-periodiek. Een geparametriseerd eindige-Fourier-reeks-model wordt gebruikt om de runoutverstoring als functie van de hoekpositie te modeleren. Hiermee kan runoutidentificatie geformuleerd worden als een parameterschattingsprobleem. Een centraal vraagstuk in het runout-correctieprobleem is hoe om te gaan met de onzekerheid in de metingen van de hoekpositie, of zelfs de afwezigheid van zulke metingen. Oplossingsrichtingen worden geformuleerd die enerzijds de fouten ten gevolge van deze onzekerheden compenseren, en die anderzijds de hoekpositie schatten uit de beschikbare data, hierbij gebruikmakend van een Extended Kalman Filter.
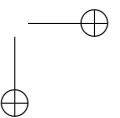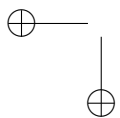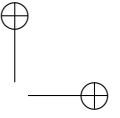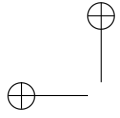
Lists of symbols and abbreviations

General notation

$\otimes$	Kronecker matrix product
$\mathbb{C}^-, \mathbb{C}^O, \mathbb{C}^+$	Open unit disc, unit circle, area outside unit disc in complex plane
$\mathrm{COV}\{A\}$	Covariance of matrix A
$\mathrm{COV}\{A, B\}$	Cross-covariance of matrix A and B
$\mathbb{E}$	Expectation operator
$\|\cdot\|_F$	Frobenius norm
$\mathbb{F}$	Set of all filters that yield a bounded estimate
$\mathbb{F}_{\mathrm{caus,N}}$	Set of all filters that are causal and stable if delayed by N time steps
vec	Vectorization operator
$\hat{x}$	Estimate of x

Roman symbols

a_e	Width of cut [mm]
a_p	Depth of cut [mm]
A	Pole surface of AMB [m^2]
A, B, C, D	State space realization matrices
A_c, B_c, C_c, D_c	State space realization of cascaded or compound system
A_K, B_K, C_K, D_K	State space realization of AMB controller
$A_P, [B_{P,1}\, B_{P,2}], C_P, 0$	State space realization of AMB spindle dynamics

BIBLIOGRAPHY

$A_u, B_u, C_u, 0$	State space realization of input process
$A_v, B_v, C_v, 0$	State space realization of noise process
A_k	Amplitude of sinusoid k
d	Tool diameter [mm]
f_z	Feed per tooth [μm]
F	Vector of generalized forces
F_t	Cutting force in tangential direction [N]
F_r	Cutting force in radial direction [N]
F_{amb}	Vector of AMB forces [N]
F_0	Optimal input estimator
$F_{\text{caus},N}$	Optimal N-causal input estimator
F_0^K	Optimal controller independent input estimator
$F_{\text{caus},N}^K$	Optimal controller independent N-causal input estimator
g	Nominal gap length [m]
$G(\xi, \theta)$	Transfer function model parametrized in θ
h_i	Uncut chip thickness of tooth with index i [μm]
i_{bias}	Bias current [A]
$K(s), K(z)$	Controller transfer function matrix
K	Number of periods
K_x	Negative stiffness matrix of AMB spindle [N/m]
K_i	Actuator gain matrix of AMB spindle [N/A]
m_i	Dimension of input with index i
M, D, G, K	Mass, damping, gyroscopy and stiffness matrix
M	Number of experiment repetitions
M	Cutting dynamics
n	used for: Rotational speed [rpm], state dimension of plant
n_a	Order of denominator polynomial
n_b	Order of numerator polynomial
n_K	State dimension of controller
N	Number of {teeth, windings, delay steps, samples}
N_K	Period length of excitation signal
p	Output dimension
$P(s), P(z)$	AMB plant transfer function matrix

$\hat{P}^{\mathrm{H}_1}(\omega_k)$	H_1 estimate of the FRF of P
$\hat{P}^{\mathrm{ARI}}(\omega_k)$	Arithmetic mean estimate of the FRF of P
$\hat{P}^{\mathrm{JIO}}(\omega_k)$	Joint input-output estimate of the FRF of P
r	Reference input of closed-loop
$R_x(\tau)$	Auto-correlation function of $x(t)$
$R_{yx}(\tau)$	Cross-correlation function between $y(t)$ and $x(t)$
RH	Set of all rational transfer functions
RH_∞	Set of all stable rational transfer functions
$\mathbf{R}(\omega_k)$	Matrix of DFTs of r (experiments collected column-wise)
$\mathbf{R}^{[i]}(\omega_k)$	Block i of $\mathbf{R}(\omega_k)$
S	Sensitivity function
S_F	Filtering sensitivity
S_V	Noise sensitivity
T	Complementary sensitivity function
T_F	Filtering complementary sensitivity
T_{yu}	Transfer function from u to y
T_s	Sample time
$u_{1,0}$	Vector of control currents [A]
u_1	Noisy measurement of $u_{1,0}$
$u_{2,0}$	Vector of cutting forces [N]
u_2	Noisy measurement of $u_{2,0}$
$\mathbf{U}(\omega_k)$	Matrix of DFTs of u (experiments collected column-wise)
$\mathbf{U}^{[i]}(\omega_k)$	Block i of $\mathbf{U}(\omega_k)$
v	Measurement noise
v_c	Cutting speed [mm/min]
v_f	Feed speed [mm/min]
v_r	Runout disturbance
$V(\theta)$	Cost criterion function
x	State signal
x_{amb}	Vector of bearing coordinates [mm]
$y_{1,0}$	Vector of displacements of rotor shaft at AMBs [mm]
y_1	Noisy measurement of $y_{1,0}$
$y_{2,0}$	Tool tip displacement vector [mm]

| $\mathbf{Y}(\omega_k)$	Matrix of DFTs of y (experiments collected column-wise)
$\mathbf{Y}^{[i]}(\omega_k)$	Block i of $\mathbf{Y}(\omega_k)$
Z^N	Set of input-output data

Greek symbols

$\epsilon(t)$	Estimation error
θ	Parameter vector
$\tilde{\theta}$	Parameter vector estimation error
σ_x	Standard deviation of $x(t)$
$\tau_F(\omega)$	Group delay at frequency ω
ϕ	Phase angle
$\phi_{k,l}$	Random phase of sinusoid k in signal l
$\Phi_x(z)$	z-spectrum of $x(t)$
$\Phi_{yx}(z)$	Cross z-spectrum between $y(t)$ and $x(t)$
Ω	Rotational speed [s^{-1}]

List of Abbreviations

ACC	Adaptive Control of Constraints
ACO	Adaptive Control with Optimization
AMB	Active Magnetic Bearing
ARI	Arithmetic Mean
BCLS	Bias Compensated Least Squares
BL	Band Limited
CTFT	Continuous-Time Fourier Transform
DARE	Discrete-Time Algebraic Ricatti Equation
DFT	Discrete-Time Fourier Transform
DTFT	Discrete-Time Fourier Transform
ECM	Micro-electrochemical Machining
EDM	Micro-Electric Discharge Machining
EKF	Extended Kalman Filter

EIV	Errors-In-Variables
FFT	Fast Fourier Transform
FRF	Frequency Response Function
IV	Instrumental Variable
JIO	Joint Input-Output
LBM	Laser Beam Machining
LS	Least Squares
MCT	Minimum Chip Thickness
MEMS	Micro-Electro-Mechanical Systems
MFD	Matrix Fraction Description
MIMO	Multiple Input Multiple Output
MOEMS	Micro-Opto-Electro-Mechanical Systems
MSE	Minimum Squared Error
OE	Output Error
SISO	Single Input Single Output
SNR	Signal-to-Noise Ratio
SRIV	Simplified Refined Instrumental Variable
SVD	Singular Value Decomposition
UIO	Unknown Input Observer
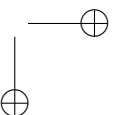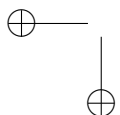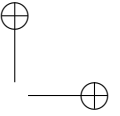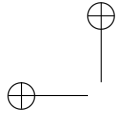
About the author

Rogier S. Blom was born on September 30th, 1975 in Haarlem, the Netherlands. After having completed secondary eduction, he studied Electrical Engineering at the University of Twente and obtained the M.Sc. degree in 2000 *cum laude*. From 1998-1999 he was Guest Researcher at the National Institute of Standards and Technology (USA), where he developed a trajectory planner for the Enhanced Machine Controller, an open architecture CNC. The topic of the M.Sc. project was Digital Elevation Model estimation using sequences of airborne video images.

From 2000-2005 he was employed by Royal Dutch KPN in its management development program and stationed at Xantic, a joint venture of KPN and Telstra (the Australian incumbent telecom operator). Here he had several positions, including those of Program Manager and Company Secretary.

In 2005 he started in a Ph.D. research project at the Delft University of Technology under the supervision of professors Paul Van den Hof and Rob Munnig Schmidt. The results of this research are described in this thesis.

In September 2010 he joined General Electric Global Research in Munich, Germany as a Research Engineer with main focus on Controls.

Publications

Refereed conference proceedings

- R.S. Blom, P.M.J. Van den Hof, Multivariable Frequency Domain Identification using IV-based Linear Regression, Proc. 49th CDC, Atlanta, 2010.
- R.S. Blom, P.M.J. Van den Hof, H.H. Langen, R.H. Munnig Schmidt, Multivariable Frequency Response Function Estimation of a Micro-Milling Spindle with Active Magnetic Bearings, Proc. Euspen conf., Delft, 31 May - 3 June, 2010.
- H.H. Langen, M.H. Kimman, R.S. Blom, P. Li, E.S. Buice, R.H. Munnig Schmidt, Towards Micro (Hard)Milling using Miniature Active Magnetic Bearings, Proc. 6th International Workshop on Microfactories, Evanston, IL, October 5-7, 2008.

BIBLIOGRAPHY

- R.S. Blom, P.M.J. Van den Hof, Estimating Cutting Forces in Micromilling by Input Estimation from Closed-loop Data, Proc. 17th IFAC World Congress, July 6-11, 2008, Seoul, 468-473.

- R.S. Blom, M.H. Kimman, H.H. Langen, P.M.J. Van den Hof, R.H. Munnig Schmidt, Effect of miniaturization of Magnetic Bearing Spindles for micro-milling on actuation and sensing bandwidths, Proc. Euspen conf, Zürich, May 18-22, 2008, 155-159.

- R.S. Blom, P. Li, H.H. Langen, P.M.J. Van den Hof, R.H. Munnig Schmidt, Micro-milling with Active Magnetic Bearing Spindles, Proc. Euspen conf., Zürich, May 18-22, 2008, 526-529.

- R.S. Blom, A.M. Hoogstrate, P.M.J. Van den Hof, H.H. Langen, Indirect Force Measurement of Micromilling with an Active Magnetic Bearing Spindle, Proc. 10th CIRP Int. WS on Modeling of Machining Operations, Reggio Calabria, August 27-28, 2007, 525-532.

- P. Li, R.S. Blom, J.A.J. Oosterling, M. Achtsnick, Design Requirements of an Advanced Micromilling Setup, Proc. 4th International Conference on Manufacturing Research, Liverpool UK, September 5-7, 2006.

Non-refereed conference proceedings

- R.S. Blom, P.M.J. Van den Hof, H.H. Langen, R.H. Munnig Schmidt, Frequency domain identifhcation of Output Error models in Matrix Fraction Description, 29th Benelux Meeting on Systems and Control, Heeze, March 30 - April 1, 2010.

- R.S. Blom, P.M.J. Van den Hof, Estimating Cutting forces in Micromilling by Input estimation from Closed-loop Data, 27th Benelux Meeting on Systems and Control, Heeze, March 18-20, 2008.

- R.S. Blom, H.M.N.K. Balini, A. Borisavljevic, G. M.J. Delhaes, M.H. Kimman, H. H. Langen, Development of 3D microstructuring technologies and machinery, National MicroNano Conference, Wageningen, November 16-17, 2007.

- R.S. Blom, P.M.J. Van den Hof, M. Achtsnick, H.H. Langen, Model-based indirect force measurement for micro-milling using active magnetic bearings spindles, Philips Precision Technology Conference, Hilvarenbeek, October 3-4, 2006.

- R.S. Blom, P.M.J. Van den Hof, M. Achtsnick, H.H. Langen, Model-based indirect force measurement for micro-milling using active magnetic bearings spindles, NanoMicroNED Symposion II, Eindhoven, November 16-17, 2006.

- R.S. Blom, P.M.J. Van den Hof , A.M. Hoogstrate, H.H. Langen, Monitoring and Control of Micro-Manufacturing Processes, 25th Benelux Meeting on Systems and Control, Heeze, March 13-15, 2006.

Other contributions

- R.S. Blom, P. Li, H.H. Langen, A.H. Hoogstrate, J.A.J. Oosterling, P.M.J. Van den Hof, R.H. Munnig Schmidt, Improvement of micro-milling technology towards industrial applications, Mikroniek, Vol. 50 (1), 2010, 6-11.