Title
Direct and Indirect Adaptive Feedforward Repetitive Control of Servo Systems

Author
Shahsavari, Behrooz

Publication Date
2015

Peer reviewed|Thesis/dissertation

Direct and Indirect Adaptive Feedforward Repetitive Control of Servo Systems

By

Behrooz Shahsavari

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering – Mechanical Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Roberto Horowitz, Chair
Professor Masayoshi Tomizuka
Professor Laurent El Ghaoui

Fall 2015

Direct and Indirect Adaptive Feedforward Repetitive Control of Servo Systems

Abstract

Direct and Indirect Adaptive Feedforward Repetitive Control of Servo Systems

by

Behrooz Shahsavari

Doctor of Philosophy in Engineering – Mechanical Engineering

University of California, Berkeley

Professor Roberto Horowitz, Chair

Control methodologies for deterministic disturbance rejection and trajectory tracking have been of great interest to researchers in the fields of controls, mechatronics, robotics and signal processing in the past two decades. The applications of these methods span a wide range from satellite attitude control requiring an accuracy of a few meters, to positioning of the read/write head in hard–disk drives with an accuracy of less than one nanometer. This dissertation addresses the problem of trajectory tracking and deterministic disturbance rejection in discrete time systems when the trajectory/disturbance is unknown, but can be realized as an affine combination of known basis functions. Despite the prior work on this problem that assumes known and time–invariant plant dynamics, we consider multi–input single–output systems with unknown dynamics. Moreover, we investigate the cases where the disturbance or system dynamics varies slowly or abruptly but infrequently. Within the broad class of disturbances/trajectories that satisfy the stated criteria, an elaborate study is conducted on periodic and superposition of multiple sinusoidal sequences.

We propose two novel adaptive control methods for the aforementioned problem. The first scheme can be classified as an indirect adaptive algorithm since it consists of two parts, namely a system identification mechanism that provides a dynamic model of the closed loop system, and the adaptive compensator which deploys the aforementioned dynamic model to synthesize the control law. The second proposed method is a direct adaptive controller, meaning that the control law is generated directly and the stated separation is not possible.

Besides providing theoretical guarantees, we experimentally evaluate our algorithms on a challenging control task for nano–scale positioning of the read–write head in a dual–stage hard disk drive (HDD). Even with the advent of NAND–flash based data storage devices, the HDD continues to thrive as the most cost effective, reliable solution for rewritable, very high density data storage. It remains a key technology particularly with the tremendously growing popularity of server–based cloud computing and novel hybrid enterprise storage solutions. We described that the control methodologies that can address the problem under our study are crucial for *Bit–Patterned Media Recording* which is one of the two breakthroughs in

magnetic recording that have been immensely investigated in the past few years. Extensive computer simulations and implementation on a digital signal processor unit are performed to validate the effectiveness of our proposed algorithms in full spectrum compensation of the repeatable runout in dual stage HDDs. This application introduces unique control challenges since it requires estimating a very large number of parameters that is order(s) of magnitude greater than prior work and frequency contents span from 120Hz to extremely large values (above 20KHz) where the plant dynamic uncertainties are large.

Contents

List of Figures

List of Tables

Acknowledgments

The past four years have brought many people across my path, and all have contributed to my journey and accomplishment in different ways. It is important for me to recount and recognize how so many people played a role in the completion of this project.

First and foremost, I want to thank my research advisor, Professor Roberto Horowitz, who guided my research and challenged me to keep my theoretical mind sharp. Over the years, he has always given me the freedom to explore my ideas, even if they were only tangentially related to the main research problem at hand; he always had complete faith in my ability as a researcher and trusted that my various projects would all further the study of the field of controls as a whole. Yet, at the same time, he was always interested, invested, and actively involved in my research and was always willing to give comments on, criticism of, and suggestions for improving any work I had done.

I also appreciate the unwavering behind-the-scenes support of Professor Masayoshi Tomizuka, my "academic grandfather". He has been there as a teacher, committee member for the qualifying exam and dissertation. His work on adaptive and repetitive control are among the references I have consulted the most throughout my Ph.D. I would also like to thank Professor Laurent El Ghaoui for serving on my qualification exam committee and my dissertation committee, Professor Claire Tomlin and Professor Liwei Lin for serving on my qualification exam committee. I have greatly benefited from Claire Tomlin's and Laurent El Ghaoui's classes in various different research projects.

I would like to especially thank Professor David Bogy who was my research advisor in the first semester of my Ph.D. program. I am indebted to him for giving me the freedom to explore other research fields and transferring to Prof. Horowitz's group. I am grateful for the level of care and support he gave me to find and pursue my interests.

I cannot thank enough Richard Conway who has been a tremendous help to me throughout my graduate career. Dr. Conway, as a postdoctoral researcher in our lab during my early years, as the instructor of advanced controls course, and later as my supervisor during two summer internships at Western Digital Co., has been an amazing mentor throughout my Ph.D. student career and taught me a great many things that shaped the foundations of my research. I am also thankful to Jianbin Nie, Siri Weerasooriya and Brian Rigney for their hospitality and advices during my internships at Western Digital.

The project described in this dissertation is unique in that it was supported by partners in the data storage industry. Hitachi Global Storage Technologies, Western Digital, Seagate,

other members of the Advance Storage Technology Consortium and Computer Mechanics Laboratory at UC Berkeley closely followed, and participated in, the progress of this research effort. Particular thanks go to Toshiki Hirano, Tetsuo Semba, Xiaotian Sun, Philip Steiner and Guoxiao Guo for providing the necessary hardware and data to perform the research, sharing with us their invaluable expertise, and providing useful feedback and suggestions.

I am thankful to my labmates, Omid Bagherieh, Fu Zhang, Josiah Wernow and especially Jinwen Pan who assisted greatly in the realization of my research. I cannot thank Jinwen Pan enough for the many hours he assisted me in implementing and experimentally verifying the control algorithms proposed in this dissertation.

In addition to the people that directly informed, guided, and funded my research, many other people greatly helped me in completing this dissertation. First of all, I would like to thank my life partner, Sayna Ebrahimi, for her constant understanding and support of my academic goals; she has been an enormous support for me and words cannot express my gratitude for her always being there for me. Last and foremost, I am thanking my parents and brother who have given me unconditional love, support and encouragement throughout my life.

Chapter 1

Introduction

1.1 Introduction

Control methodologies for disturbance rejection or trajectory tracking have been of great interest to researchers in controls and signal processing communities ever since 1930's. Vibration cancellation in electro–mechanical systems, acoustic noise reduction and trajectory tracking are relevant problems that have attracted a great deal of attention.

Disturbance compensation techniques can be categorized into "passive" and "active" methods. The traditional "passive" techniques are applicable to a confined class of systems and disturbances. Large scale mechanical systems such as vehicles – where active compensation is costly due to the need for powerful actuators – and high–frequency disturbance attenuation – through absorbers and mechanical dampers – are such applications Harris (1991); Beranek and Ver (1992); Nair and Keane (2001); Herzog (1994); Konstanzer et al. (2008). These passive methods are not in the scope of this study and will not be discussed. Rather, we focus on active compensation techniques that are more advanced, flexible and applicable to a considerably broader class of applications.

Regardless of the target application, all active methods are hinged to a simple and natural idea based on superposition of signals.That is, a signal correlated to the disturbance is taken in by the algorithm, transformed and then injected to the environment (i.e. a dynamic system) through an apparatus such that the injected signal makes an effect similar in magnitude but opposite in phase to the disturbance effect. This idea is inherited from early work such as Lueng's intuitive method (Lueng (1936)) that is illustrated in Fig. 1.1.

The active compensation and, similarly, trajectory tracking problems can be split into four categories based on whether the plant dynamics is known or unknown and whether the disturbance (or equivalently the trajectory) characterization is available or not. We will briefly review these four classes and applicable techniques for the case of broad band and narrow–band disturbance rejection to clearly determine what type of problems will be addressed in this dissertation and what classes are irrelevant. The related work will be discussed in detail later throughout the dissertation. Note that a "disturbance" is inherenelty unmeasurable and the notion of "disturbance characterization" only refers to some information

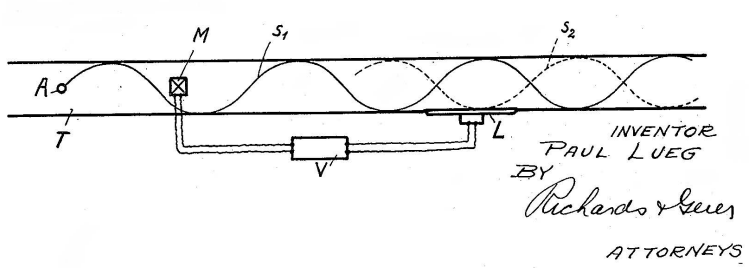


Figure 1.1: Schematic of a single frequency sound wave rejection method proposed by Lueng in 1934 Lueng (1936). The microphone M captures some information about noise S_1. The block V$'$ through speaker L attempts to reproduce S_1 based on information from M and inject it to the environment with an opposite phase. *(The figure is extracted from the original patent filed by Lueng.)*

about the disturbance spectrum or a transfer function that can represent its power spectral density. It is also worth noting that tracking problems, especially when the trajectory is unknown, can be treated as a "rejection" problem by defining an appropriate performance measure that bolds tracking error. Therefore, in the sequel, we use the *trajectory tracking* and *disturbance rejection* interchangeably. The four aforementioned problem categories are as follows:

1. **Problems with known plant dynamics and known disturbance models:** Both feedback and feedforward control algorithms have been applied to the problems in this framework. The use of feedback control for this type of problems is of fundamental importance since an optimal controller can be designed offline based on priori knowledge about the system and disturbance. In broad band disturbance rejection, a causal linear time invariant (LTI) system that represents the disturbance spectral density along with a well–modeled plant can be exploited to pose a classical linear quadratic Gaussian (LQG) problem with weighting filters to shape the controller effort in accordance with the disturbance spectrum and performance objectives Gupta (1980); Moore and Mingori (1987); Connolly et al. (1995); Shahsavari et al. (2013a).

   On the other hand, for the case of multiple narrow band disturbances with known frequencies, a model of the exogenous signal can be embedded in the controller to produce high–gain feedback at frequencies that comprise the disturbance spectrum. The resulting *internal model controller* applies high gain at the disturbance frequencies to obtain high level of disturbance attenuation Francis and Wonham (1976); Bengtsson (1977); Tsypkin (1997).

   In a feedback control framework, the disturbance rejection at a frequency – by the classical Bode integral theorem Bode et al. (1945) – causes amplification at other frequencies and full spectrum rejection is impossible. Since feedforward control strategies do not have this limitation, a great deal of research effort has been focused on deriving feedforward methods that have "zero spillover". The relationship between these

two paradigms is investigated in MacMartin (1994); Hong and Bernstein (1998). In general, the feedforward control is advantageous over the feedback control when the controller can sense the disturbance, through an appropriately located sensor, before the disturbance propagates through the system and affects the *performance* signal.

2. **Problems with known plant dynamics and unknown/partially known disturbance models:**
Adaptive algorithms are crucial in control problems where the plant dynamics or noise model are not known. In broad band disturbance rejection where the noise statistics is not known, these methods are usually exploited to estimate the noise statistics or to design state estimators Odelson et al. (2006); Mehra (1970); Myers et al. (1976); Nummiaro et al. (2003) that can be exploited in a state feedback control paradigm. Adaptive methods that suppress the effect of unknown and time–varying narrow band disturbances have been recently revisited by many researchers Bodson and Douglas (1997); Marino et al. (2003); Landau et al. (2011a). This type of problems is named "Adaptive Regulation" by Landau and a survey of relevant literature till 2011 was reported Bobtsov et al. (2012). Moreover, a benchmark problem for suppression of multiple unknown/time–varying narrow–band vibrations and a comparison of multiple adaptive feedback control systems has been recently carried out Landau et al. (2013). The term "partially known" on the title refers to the algorithms that benefit from knowing that the disturbance includes only narrow–band contents. For instance, Chen and Tomizuka (Chen and Tomizuka (2012, 2013)) embed parametric models for narrow–band disturbances in an adaptive Q filter to form a novel disturbance observer for selective disturbance cancellation.

3. **Problems with unknown plant dynamics and known disturbance models:**
Indirect adaptive control deals with this type of problems by converting them to the first type through an online system identification mechanism that attempts to learn a plant model (chapter 12 in Landau et al. (2011c)). Indirect adaptive methods for pole placement Giri et al. (1987), generalized predictive control M'Saad et al. (1993) and linear quadratic control M'saad and Sanchez (1992) are among the most established techniques for broad band disturbance rejection Zhang et al. (2014b,a); Shahsavari et al. (2014a); Bagherieh et al. (2015). There are also *direct adaptive control* schemes that directly update the parameters of the controller from a signal error (adaptation error) reflecting the difference between attained and desired performance M'saad et al. (1985). Pioneer research on direct adaptive schemes for pole placement includes Elliott (1981); Leal and Landau (1982); Åström (1980). For the case of narrow band and sinusoidal disturbances acting on unknown plants, adaptive feedforward methods are developed that generate the control signal based on an online estimation of the plant frequency response and the disturbance parameters Chandrasekar et al. (2006); Pigg and Bodson (2006, 2010).

Beside the adaptive control algorithms, numerous methods have been presented by

the signal processing community with an aim to combine a gradient algorithm (i.e., adaptive least mean–squares or LMS algorithm) with an online identifier of the plant impulse response. The identification results are then used in an active noise cancellation method that require a plant model Kuo and Morgan (1995); Zhang et al. (2000, 2001).

4. **Problems with unknown plant dynamics and unknown disturbance models:** This is the most general type of regulation/rejection problems that is usually treated by adaptive control methods mentioned in the second category. Moreover, self–tuning regulators Åström (1975) are flexible control schemes that can be applied to this type of problems.

Note that these four categories only include problems related to this dissertation and various other control methodologies, such as *model predictive control* Camacho and Alba (2013); Shahsavari et al. (2015b, 2016) are not considered here.

This work considers the problem of trajectory tracking or equivalently deterministic (but unknown) disturbance rejection in discrete time systems when the trajectory or the disturbance can be realized (exactly or approximately) by the dot product of a known and an unknown vector. Let d_k denote the disturbance or trajectory signal sampled at time step k

$$d_k = \theta^T \phi_k \tag{1.1}$$

where $\theta \in \Re^n$, $(n \geq 1)$ stands for an unknown vector of parameters. The real value vector $\phi_k \in \Re^n$ is known and can be thought of as a *basis function* or a *kernel* for defining the disturbance. In other words, ϕ represents what we know about the "noise characterization" and θ is what makes the signal unknown to us. Problems with slowly varying parameter vector θ_k are relevant and will be discussed too.

Different type of disturbances can be realized by (1.1). The simplest case is an unknown DC bias $d_k = d_0$ that can be realized by $\theta = d_0$ and $\phi_k = 1$. This idea can be generalized to periodic signals with known periods. Let $d_k = d_{k+n}$ be an n–periodic signal and $\theta^T = \begin{bmatrix} d_0, d_1, \cdots, d_{n-1} \end{bmatrix}$. Expression (1.1) can be satisfied by choosing ϕ_k as a sparse vector that has only one nonzero value which equals 1 and is located at the i-th element where $i = \text{mod}(k, n) + 1$. Signals that are superposition of multiple sinusoids

$$d_k = \sum_{i=1}^{n} \gamma_i \sin(\omega_i k + \delta_i)$$

fall in this class of disturbances/trajectories too. One needs to include $\sin(\omega_i k)$ and $\cos(\omega_i k)$ in $\phi_k \in \Re^{2n}$ and the pairs $\gamma_i \cos(\delta_i)$ and $\gamma_i \sin(\delta_i)$ in θ in order to have a realization like (1.1). Applications with an extra transducer – such as a microphone or an accelerometer – that provides measurements that can be transformed to the disturbance (or trajectory) through a (finite dimension) FIR filter can also be treated in this framework. In this case, a finite horizon of historical measurements should be included in ϕ_k and the tap coefficients in θ.

As for the plant dynamics, we consider both cases of known and unknown systems. Therefore, the methods discussed in this dissertation fall in category 2 and 3. We will discuss

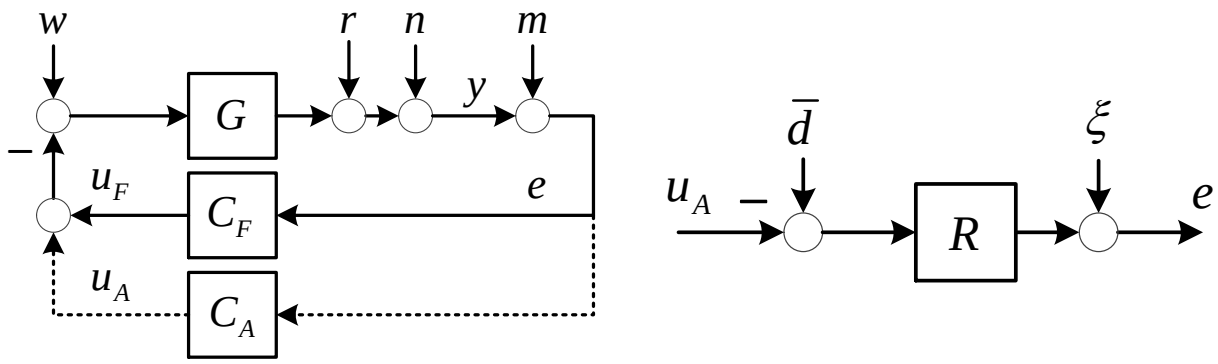


Figure 1.2: Left: Closed loop system augmented by a *plug–in* adaptive controller. Right: Succinct representation of the closed loop system.

an indirect adaptive algorithm for the problems that satisfy hypothesis (1.1) and a direct adaptive algorithm for the disturbances (trajectories) that are periodic or the superposition of multiple sinusoids with known frequencies.

Within the broad class of disturbances/trajectories that satisfy hypothesis (1.1), an elaborate study will be specified to the periodic and superposition of multiple sinusoidal signals. Although, from a theoretical point of view, these disturbances might seem to be a very special and constraining case of (1.1), there is a multitude of applications – especially due to the dominating role of rotary actuators and power generators – that crucially depend on this type of regulation. A non–exhaustive list of these applications include aircraft interior noise control Emborg (1998); Wilby (1996), vibration rejection in helicopters Malpica (2008); Patt et al. (2005); Bittanti and Moiraghi (1994), periodic load compensation in wind turbines Stol and Balas (2003); Houtzager et al. (2013), gearbox systems Li et al. (2005); Guan et al. (2004), optical disk drives Moon et al. (1998); Doh et al. (2006), wafer stage platforms De Roover and Bosgra (2000); Dijkstra (2004), tape drives Panda and Lu (2003); Pantazi et al. (2012), magnetic bearings Knospe et al. (1997); De Wit and Praly (2000), steel casting processes Tsao and Bentsman (1996), spacecrafts Lau et al. (2006); Goodzeit and Phan (1997), laser systems McEver et al. (2004) and milling machines Rober and Shin (1996); Tsao and Pong (1991). We will elaborate the application of the proposed algorithms to hard disk drives, especially to an emerging technology called *bit patterned media recording*.

In the following section, the problem of tracking unknown trajectories, or compensating unknown deterministic disturbances that was intuitively described above is mathematically formalized and the system under our study is described.

1.2 Problem Statement

The adaptive controllers proposed in this work are aimed to be implemented in a *plug–in* fashion. That is, an adaptive controller is used to augment an existing robustly stable closed loop system in order to compensate a special type of disturbance that is not well attenuated

by the existing controller. In this architecture, the original controller can be designed without consideration of the special disturbances. Moreover, the adaptive controller does not alter the performance of the original control system. To clarify this notion, we use a common plant–controller interconnection shown in Fig. 1.2(left) as a running example. The blocks G and C_F in the figure respectively denote a linear time invariant (LTI) plant and a nominal LTI feedback compensator that is used to stabilize the plant. This nominal controller can be continuous or discrete time, and it generally provides disturbance rejection across a broad frequency spectrum. On the other hand, the plug–in adaptive controller, denoted by C_A, is a non–linear discrete time system that provides compensation for the disturbance r_k that can be decomposed to

$$r_k = \theta^T \phi_k. \tag{1.2}$$

As mentioned in the introduction, there are numerous classes of signals that can be decomposed in this format exactly and many other types can be approximated by incorporating the coefficients of an (unknown) FIR filter in θ and tapped values of a signal correlated to r_k in the regressor ϕ_k. Since our design does not depend on whether the plant/nominal controller are continuous or discrete time, we assume that both G and C_A are discrete time systems to make notations simpler.

We consider a general stochastic environment for the system by appending input disturbance w, output disturbance n, and contaminating measurement noise m to our framework. Generally, the nominal feedback controller is designed to compensate for these input and output noises. The special disturbance that should be compensated by the adaptive controller is denoted by r, and without loss of generality, we assume that it is applied to the plant output.

An important point to make here is that our plug–in controller design is not limited to this particular interconnection, and in general, it does not require any details about the individual components of the closed loop system and their interconnections. Rather, our design is based on an abstract LTI dynamics from the adaptive control (u_A) injection point to the *error* signal (e). We will show by simulation and experiments that the algorithms, in practice, can be applied to slowly time varying systems and in some special cases to switching systems when the time interval between switchings is considerably larger than the controller sampling time.

A succinct representation of this framework is shown in Fig. 1.2(right). Indeed, our design is only based on the dynamics of $R(.)$ which is the transfer function from u_A to e, and without loss of generality we assumed that it is discrete time. Returning to our running example, these blocks and signals shown in the abstract form are defined by

$$\begin{aligned} S(z) &:= \frac{1}{1 + G(z) C_F(z)} \\ R(z) &:= G(z) S(z) \\ \xi_k &= R[w_k] + S[n_k + m_k]. \end{aligned} \tag{1.3}$$

Here, the standard discrete–time z variable notation is used for the transfer functions, and the time functionality of the signals is shown by the step index k. We only focus on multi–input single–output (MISO) plants throughout this dissertation. The results, however, can be easily extended to multi–input multi–output (MIMO) systems with minor modifications. For a general n_i–input n_o–output transfer function $T(z) \in \Re_z^{n_o \times n_i}$ and an input sequence $i_k \in \Re^{n_i}$, the bracket notation $T[i_k] \in \Re^{n_o}$ represents the time domain response of the system. For instance, the response of the transfer function $R(z)$ to the input disturbance $w(k)$ is represented by $R[w_k]$. When the same transfer function filters multiple input signals $i_k^1, i_k^2, \cdots, i_k^m$, we abuse the notation and use

$$T\left[\begin{bmatrix} i_k^1 \\ i_k^2 \\ \vdots \\ i_k^m \end{bmatrix}\right] = \begin{bmatrix} T[i_k^1] \\ T[i_k^2] \\ \vdots \\ T[i_k^m] \end{bmatrix}.$$

The special disturbance r_k in this block diagram can be replaced by an equivalent disturbance, $\bar{d}$, which has the same effect as r on the error signal,

$$R\left[\bar{d}_k\right] = S\left[r_k\right].$$

From a control point of view, this replacement is admissible since the disturbance is bounded, (cyclo)stationary, and the system is linear; hence, it is possible to consider the disturbance at any other point, or break it to portions injected at multiple points to the closed loop system. For instance, for a dual–stage system – i.e. a system with two inputs and one output – we can split r_k into low and high frequency portions based on its spectrum and then transfer these two signals to the two inputs of plant $G(z)$. We will show that this separation and transformation is beneficial when different input–output channels of the plant have different characteristics. For example, if one channel has a higher gain than the others at certain frequencies, or it has less dynamics uncertainty in that frequency region, it is beneficial to associate the disturbance compensation at that frequency interval to this input channel.

Suppose that $R(z) \in \Re_z^{1 \times n_i}$ is a multi–input single–output system and $\bar{d}_k \in \Re^{n_i}$ is the multi–dimensional unknown disturbance/trajectory that is considered on (or transformed to) the input side. Our objective is to synthesize an adaptive controller that only uses the scalar–valued error signal e_k to generate a vector–valued control $u_{A,k} \in \Re^{n_i}$ such that it compensates the contribution of $\bar{d}$ in the variance of error signal.

1.3 Contributions per Chapter

The following contributions are presented in the respective chapters of this dissertation:

- **Chapter 2:** Among various applications that were mentioned for the adaptive control algorithms proposed in this dissertation, we are particularly interested in nanopositioning in hard disk drives (HDDs). The importance of HDDs compared to other types

of data storage technologies such as *NAND–flash*–based memory devices and *Tapes* is explained and their technology road maps are compared briefly. The recent evidences show that fast read access of NAND flash technology, especially for random reads, has replaced the hard disk drives by flash storage devices, especially Solid State Drives (SSD), in majority of personal electronic devices. However, we will show that the increasing need to data storage capacity in data centers and cloud computing has caused ever increasing demand for HDDs.

We review the fundamental aspects of HDD servo mechanism that is responsible for nanopositioning of the read–write head. This is followed by introducing *Bit–Patterned Media Recording* technology that necessitates precise control algorithms for tracking unknown but periodic trajectories. The control problem for this application is formalized, the system dynamics for both single and dual stage servo systems are provided and different types of disturbance/noise that contaminate such a system are described. Lastly, the experimental setup that we prepared for DSP implementation of the algorithms in order to apply them to a dual–stage hard disk drive servo system is presented. This setup is used for extensive experiments that will be discussed in detail in two other chapters.

- **Chapter 3:** Control methodologies for coping with periodic signals, commonly known as repetitive controllers, and their applications are briefly reviewed in this chapter. In particular, we compare the feedback control and feedforward control methods and show that for the class of problems under our study the latter methods are superior. We deploy a stochastic gradient descent method, to develop an adaptive feedforward control algorithm for compensating multitude narrow–band disturbances or trajectory tracking. Using the averaging theory, we derive a set of conditions on the adaptation step size to guarantee the algorithm convergence and perfect compensation. We propose a novel adaptive step size and integrate it to the adaptive control algorithm to enhance the convergence rate and decrease the steady state error.

  The analysis are initially carried out in a spectral framework where trigonometric functions are chosen to form an orthogonal basis for the space of real valued square integrable trajectories (deterministic disturbances) that should be tracked (compensated). Two alternatives to this approach, namely decomposition of the disturbance in time–domain and in time–frequency–domain, are also discussed. It is shown that time–domain approaches can be computationally more plausible than frequency–domain methods. However, the latter methods can be more robust to system dynamics uncertainty.

- **Chapter 4:** The parameter adaptation algorithm proposed in chapter 3 requires a model for the system dynamics. In general, an exact dynamics of the system, especially at high frequencies, is not known in many practical applications. The robustness of the adaptive feedforward controller to dynamic mismatch between the actual plant and the model deployed in the algorithm is analyzed in this chapter and it is shown that unlike

many existing methods, the mismatches are only important at excitation frequencies.

An online system identification architecture is proposed to provide an accurate model of the system dynamics in case a model is not available or accurate. It is shown that under a set of assumptions, the proposed scheme is able to obtain a model that satisfies convergence criterion outlined in the aforementioned robustness analysis. As a result, the identification and compensation mechanisms together form an "indirect" adaptive controller for the class of problems under our study.

The proposed identification scheme requires an exogenous excitation signal. In general, this type of extra excitations are not desired from a practical point of view. We propose a special excitation that is extremely low power, effective in our problem since its energy is focused around frequencies that are important to us, and easy to generate in DSP implementation. Besides, we suggest using adaptive band–pass filters on the inputs to the identification unit in order to further reduce the required excitation power. A design method for synthesizing these adaptive filters through *frequency transformation* of a prototype filter is proposed. It is shown that explicit "parametric" solutions for the filter coefficients can be obtained. The "parameters" of these relations depend on the pass–band of the filter and can be evaluated easily in real–time.

- **Chapter 5:** Besides providing theoretical guarantees, we experimentally evaluated our algorithm on a challenging control task for nanopositioning of the read–write head in a dual–stage HDD. It is illustrated how the proposed adaptive control algorithm can be efficiently integrated with the servo controller of an HDD for following/compensating repeatable runout (RRO), a problem that was introduced and formalized earlier in chapter 2.

  We describe our modeling procedure for computer simulations and provide remarks for efficient implementation of the algorithms on an embedded processor. Computer simulation in MATLAB and implementation on a digital signal processor (DSP) unit are performed to compensate for RRO that has narrow–band contents at the HDD spinning frequency (120Hz) and its 173 higher harmonics (up to 20'880Hz). This is a challenging task since it requires estimating a very large number (348) of parameters which is order(s) of magnitude greater than other results reported in the literature. These frequencies span from 120Hz to extremely large frequencies (above 20KHz) where the plant dynamics uncertainties are large and feedback controller amplifies disturbances.

- **Chapter 6:** In this chapter we propose a direct adaptive control for the problem of compensating (tracking) periodic disturbances (trajectories) in systems with unknown dynamics. We denote this method as a "direct" algorithm since the control parameters are directly updated by the parameter adaptation algorithm and a "control design" block that commonly exists in indirect methods is avoided. The analysis is carried out in a spectral framework which makes the algorithm applicable to non–minimum phase systems without requiring any assumptions. Moreover, the method is applicable in situations with slowly time–varying systems and disturbances. Furthermore, unlike

most of the existing methods in the literature it does not require batches of data to update its parameters.

It is shown that the number of estimated parameters in our proposed algorithm is slightly larger than $2n$ where n is the number of disturbance/trajectory frequency contents, while other methods require $4n$ parameters ($2n$ for system dynamics and $2n$ for control synthesis). As a result, our method estimates significantly less parameters when n is large (e.g. as in the HDD problem where $n = 174$). The convergence of adaptive algorithm parameters to actual values is rigorously analyzed and a set of practical remarks are made for reducing the transient error.

We will show that this adaptive controller scheme is "modular", meaning that it can be split to "smaller" controllers such that each one estimates a portion of the original unknown parameter vector. This is a very appreciated property because: (1) identification of systems with complex frequency responses is, in general, very difficult and may be impossible in noisy systems. On the other hand, breaking the spectrum to small partitions and learning a model for each one is a significantly easier task. (2) large amount of computation can be split between nonconcurrent modules.

Partitioning the spectrum among a set of controller modules can be manually or automatically. We will propose a heuristic algorithm inspired by divisive hierarchical clustering in conjunction with a monitoring mechanism to perform the partitioning in an automated fashion.

- **Chapter 7:** The direct adaptive control algorithm is experimentally evaluated on the HDD setup described in the previous chapters. In addition to testing the control tasks presented in chapter 5, the algorithm behavior in the cases of slow and abrupt changes of the system dynamics and trajectories is studied. Moreover, the frequency partitioning algorithm proposed in chapter 6 is implemented and it is illustrated how effectively this mechanism is able to split full spectrum compensation among several small automatically defined adaptive control modules.

- **Chapter 8:** In chapter 8, we make some concluding remarks and mention some areas of future work.

Chapter 2

Nano–positioning in Hard Disk Drives

2.1 Introduction

Nanotechnology is the science of understanding and manipulation of matter with at least one dimension in the range of 1 to 100 nanometers. Nanopositioning is a crucial aspect of nanotechnology that involves precision control and manipulation of devices and materials at the above dimension range. Nanopositioners are indeed precision mechatronic systems designed to move objects with a resolution down to a fraction of an atomic diameter. The desired attributes of these mechanisms are extremely high resolution, robustness, and fast response. The key to successful nanopositioning is accurate position sensing and highly responsive motion control mechanism. Hard disk drives, scanning probe microscopes, lithography tools and nano-assembly tools are among the various applications of nanopositioning. In this section we focus on nanopositioning mechanism in hard disk drives which is controlled by the HDD servo mechanism.

TAPE, NAND Flash, and hard disk drives (HDD) are the three major device technology types that are used for storage class memory applications. A measure of the progress of these technology types has been areal density, i.e. the number of bits stored per unit area. In recent years, all three technologies have been characterized by annual areal density increases of approximately 30% Fontana et al. (2012). Fast read access of NAND flash technology, especially for random reads, has replaced the hard disk drives by NAND flash memory-based storage devices, especially Solid State Drives (SSD), in majority of personal electronic devices such as laptops and desktops. However, the increasing need to data storage capacity in data centers and cloud computing has caused ever increasing demand for HDDs since the price (per data volume) of SSD is several times more expensive than the HDD. Moreover, SSD has several weak points caused by the nature of NAND flash memory. The "erase-before-write" characteristic that causes slow and non-uniform write latency and wear-out problem are other factors that have limited the use of SSDs in hyper-scale data storage infrastructures. Recently, Western Digital Corporation – one of the largest computer data storage companies – reported the statistics shown in Fig. 2.1 for the data capacity demand prediction till 2020

and claimed that 75% of the total volume will be stored on HDDs in 2020[1].

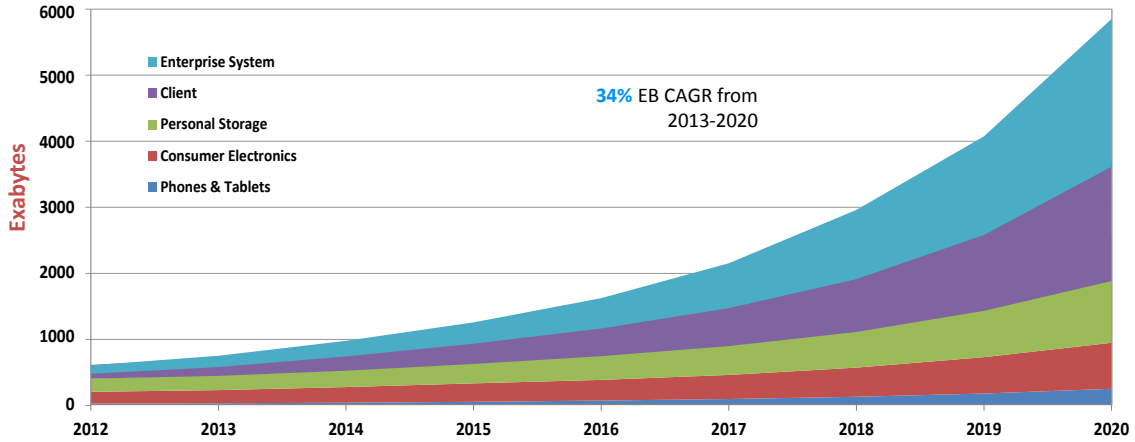


Figure 2.1: Prediction of data capacity demand till 2020.(Courtesy of Western Digital Corporation[2].)

Ever increasing aerial density in HDDs that utilize classical perpendicular magnetic recording (PMR) Khizroev and Litvinov (2006) has arose potential limitations that impede maintaining the same growth rate in the future. This is mainly due to the fact that as the magnetic "bit" area shrinks, the data bits become thermally unstable since they approach thermal energy fluctuation limits associated with the media grains Eisenmenger and Schuller (2003). In the past 5 years, in order to overcome this limitation, a great deal of research has been focused in two directions, namely *Bit Patterned Media Recording (BPMR)* Albrecht et al. (2015) and *Heat Assisted Magnetic Recording (HAMR)* Challener et al. (2009). *Advanced Storage Technology Consortium (ASTC)* as a hub for collaborative joint R&D effort among university researches and the main computer data storage companies – e.g. Seagate, Western Digital Co. and Hitachi GST – recently published a technology road–map shown in Fig. 2.2. BPMR as one of the two breakthroughs in data storage technology emerges specific challenges in terms of nanopositioning and requires a precise control mechanism. We will describe that precise tracking of unknown but periodic trajectories is a crucial control requirement in BPMR. Moreover, we will validate the effectiveness of our proposed control algorithms by conducting comprehensive simulations and experiments on such a system.

We begin by reviewing the fundamental aspects of HDD servo mechanism and its control requirements. This is followed by introducing BPMR technology that necessitates precise control algorithms for tracking unknown but periodic trajectories. We will formalize this control problem, provide the system dynamics for both single and dual stage servo systems

[1]The plot was presented in one of the Advanced Storage Technology Consortium technical meetings. It is reprinted here by permission from the author.

[2]Figure reprinted by permission from corresponding author in Western Digital Co.

[3]Figure is available to the public on http://www.idema.org/wp-content/plugins/download-monitor/download.php?id=2244

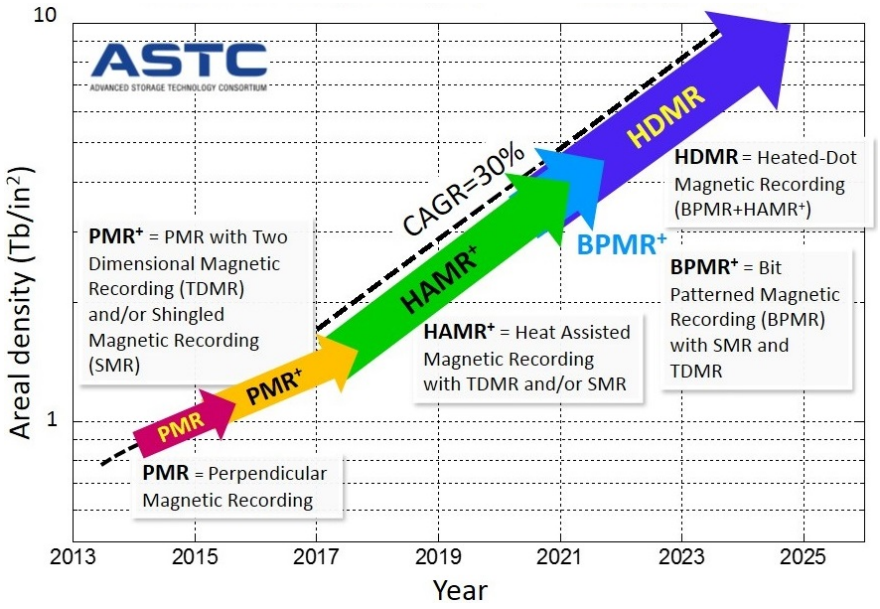


Figure 2.2: Technology roadmap for data storage technologies reported by Advanced Storage Technology Consoritium (ASTC) [3]. Compound Annual Growth Rate (CAGR) predicted to be achieved by the emerging technologies is 30%.

and describe different types of disturbance/noise that contaminate such a system.

2.2 Hard Disk Drive Servo Mechanism

The main mechanical parts of a hard disk drive are shown in Fig. 2.3. An HDD stores data on a set of spinning magnetic disks by using a few electronic *read-write heads* (shortly called the *heads*) that fly above or under the disk surfaces and write data bits along a set of "tracks". Figure 2.4 depicts an HDD with 2 disks – 4 magnetic surfaces – and 4 heads. The head positioning servomechanism moves the magnetic read/write head as quickly as possible from one track to another when asked by the host system using track-seeking and track-settling control systems. Once the head reaches the target track, its position relative to the track's center is controlled by a track-following servo system during the data reading and writing process. In this process, a voice coil motor driven by the servo controller is responsible for moving the head over the span of the disk. Such a mechanism that solely relies on the VCM for actuation is called *Single Stage* servo. The bandwidth of a single-stage servo system is limited by the multiple mechanical resonances of the VCM as well as the control effort saturation bounds Miu and Bhat (1991). Many hard drives are equipped with secondary actuators that have a higher bandwidth and a smaller stroke Horowitz et al. (2007). This architecture is known as "Dual Stage" servo mechanism and these actuators are commonly called *micro* or *milli actuators* based on their strokes Zheng et al. (2015). Here, we use "MA" to refer to both cases. There are different possible locations between

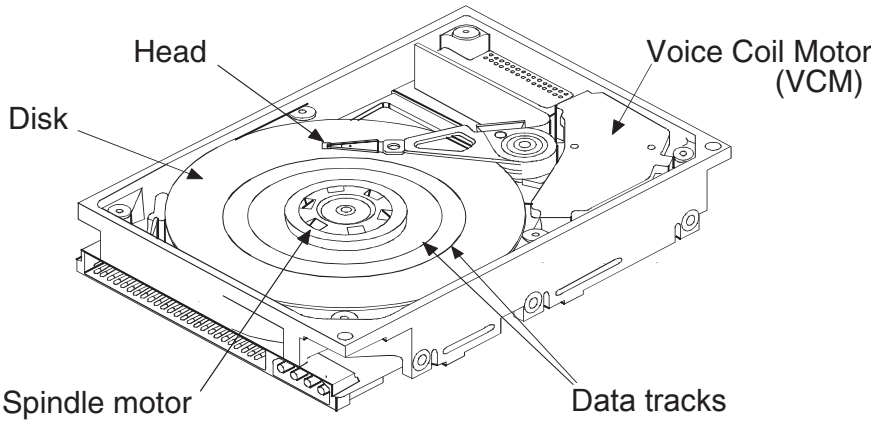


Figure 2.3: Disk drive configuration. (Reprinted by permission from Horowitz et al. (2007).)

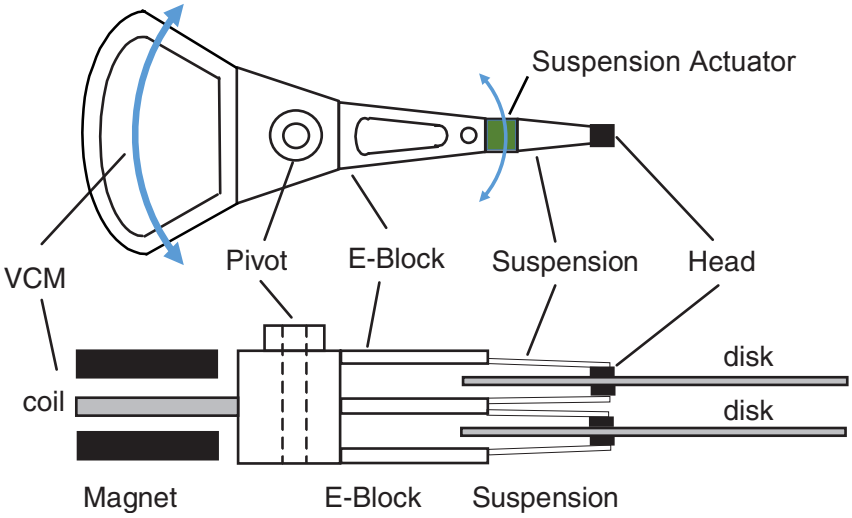


Figure 2.4: Actuators and arm assembly in a dual stage HDD with two disks and four heads. (Reprinted by permission from Horowitz et al. (2007).)

the E-Block and head to mount the MA. Figure 2.4 shows a so called *suspension actuated* configuration. Researchers have also proposed slider-based actuators – i.e. actuators that moves the slider relative to the suspension – and head-based actuators – actuators that move the head relative to the slider Hirano et al. (2003). Regardless of the MA type and its location, all configurations are aimed to move the head in a direction parallel to the disk surface and locally normal to the data tracks. Since the disk diameter is orders of magnitude larger than the track pitch (distance between the center of two neighboring tracks), the curvatures in track shapes and actuator movements are negligible.

The position of read-write head is obtained from a set of *servo patterns* whenever the read-write head crosses them Abramovitch and Franklin (2002); Al Mamun et al. (2007). Servo patterns consist of several magnetic sequences, including track number, sector number,

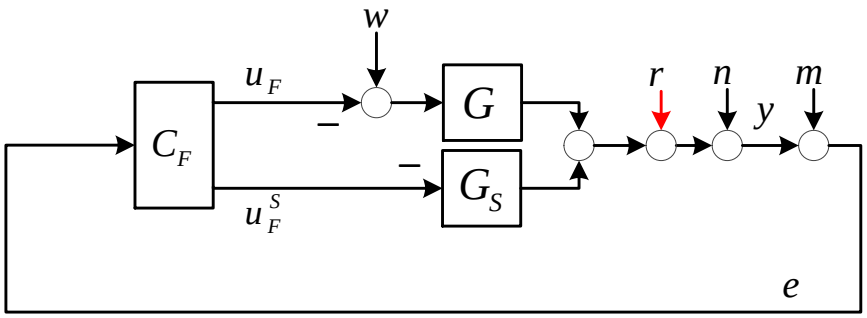


Figure 2.5: Feedback control of an HDD dual-stage system in track-following mode.

and sub-track position reference pattern (usually called a burst pattern). The servo sectors are ideally distributed on equidistant points on the perimeter of concentric circles. Suppose that the disk is spinning at a constant speed $S(rev/s)$ and there are N servo sectors on circumferential direction. The *Position Error Signal (PES)* is obtained $S \times N$ times per second which implies that the digital servo controller has a sampling frequency of $S \times N$(Hz). For example, a 7200RPM HDD with 348 wedges (distributed circumferentially) has a sampling frequency of 41.760KHz.

The servo system is subjected to both repeatable (periodic) and non-repeatable (random) disturbances/noises that are due to the imperfection in fabrication and assembly processes, internal and external vibrations Sun et al. (2014, 2013); Zheng et al. (2014a,b), and electronic interferences. Fig. 1.2 (left) can be adopted to abstract the block diagram of a single stage HDD servo system in track-following mode. The blocks G and C_F refer to a voice coil motor (VCM) and the nominal feedback controller respectively. The signals w, r, n and m in Fig. 1.2 denote the airflow disturbance known as *windage*, *repeatable runout* (RRO), *non-repeatable runout* (NRRO) and *measurement noise* respectively. The design of the feedback controller C_F is not discussed in this dissertation and it is assumed that this compensator can robustly stabilize the closed loop system. The interested reader can refer to Shahsavari et al. (2013c,b, 2012); Keikha et al. (2013); Bagherieh et al. (2014) for further details of designing the feedback controller for an HDD servo system. The actual *position error signal* (PES) and measured PES are respectively referred by y and e in the figure. In the remaining, the term "PES" is used for referring to the measured PES signal. NRRO is the random lateral movement of the disk caused by the mechanical contacts in the spindle bearing, and windage is the off track motion at the head caused by the turbulent nature of the air between the disk and the actuator. The block diagram of a dual stage servo system is shown in Fig. 2.5. In the figure, G_S represents the "S"econdary actuator (i.e. the MA). As is shown in the figure, the position of head relative to the track is due to the contributions from both actuators. Note that this is the only feedback signal and the controller does not have access to the position of each actuator individually. The windage affects both actuators; but, since this does not change our analysis, we have considered it only on the VCM.

The performance metric of track-following controller is usually evaluated by the 3σ (3

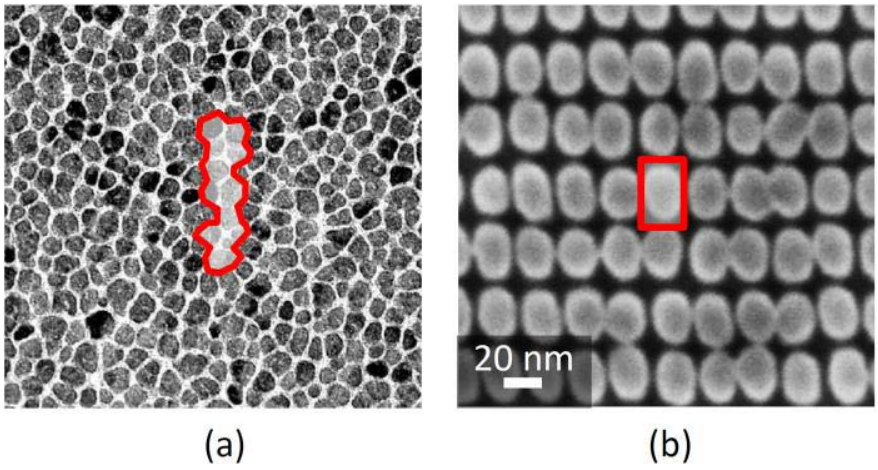


Figure 2.6: Comparison of (a) traditional granular media and (b) BPM. On granular media, an individual bit is recorded on an ensemble of grains (red outline), while on BPM, each island stores 1 bit. (Reprinted by permission from Albrecht et al. (2015) © 2015 IEEE.)

times the standard deviation) of the PES, which is usually called the *Track Mis-Registration* (TMR) budget. A simple rule of thumb is that the TMR in current HDDs is desired to be approximately 8% of the track pitch. That means in a HDD with 500'000 tracks per inch, the 3σ value of the PES should be kept around 4 nanometers (nm).

2.3 Application to Bit Patterned Media Recording

Annual increase in data aerial density (AD) has slowed down to less than 20% in recent years due to the challenges with thermal stability of granular magnetic material when the size of recorded bits shrinks. In order to be able to read recorded data from a traditional granular media it is required to maintain an adequate signal to noise ratio (SNR) by scaling the grain sizes with the size of recorded bits. This can cause thermal instability as the grain sizes become smaller. Bit patterned media (BPM) is a breakthrough in data storage since it can address many of these issues and facilitate thermally stable magnetic recording at higher than 1Tb/in^2 AD Yang et al. (2011); Albrecht et al. (2009); Chou et al. (1994); New et al. (1994). The key idea of BPM is that each data bit is recorded on a single isolated magnetic island in a gigantic array of islands patterned by lithography on the magnetic disk.

Figure 2.6 compares one bit recorded on an ensemble of grains on granular media with another bit recorded on a single island on BPM. The SNR in granular data depends on the number of grains, whereas in BPMR, it relies on the fabrication tolerances and servo accuracy to position the head – in both reading and writing – exactly on top of an island. The resulting small track pitch (TP) in BPMR makes the TMR budget a large fraction of the TP and makes the servo system a crucial component that can limit the achievable aerial density. This introduces significant new complexity by requiring a write synchronization

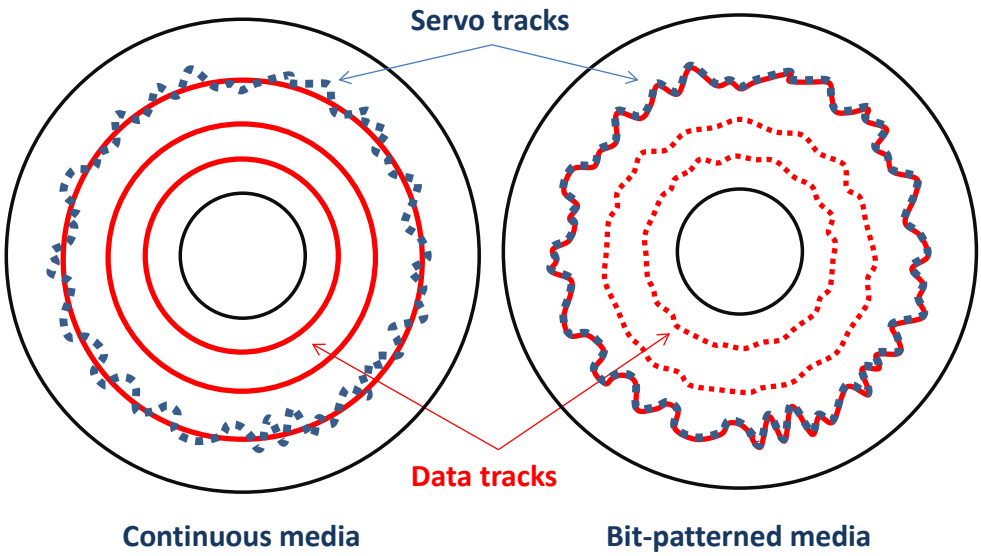


Figure 2.7: Schematic of servo tracks (dotted lines) and data tracks (solid lines) in conventional and bit-patterned media HDDs.

system and servo technology suitable for following eccentric, prepatterned tracks at track pitch of 20 nm.

BPMR requires that the data tracks be followed with significantly more accuracy than what is required in conventional continuous media recording, since the read/write head has to be accurately positioned over the single-domain magnetic islands to read or write data. In traditional magnetic recording, data is (ideally) written on concentric circular tracks since the media is continuous, whereas in BPMR Shahsavari et al. (2014b, 2015a), data should be written on data tracks with unknown shapes, which are created by lithography on the disk. Accordingly, the servo control methodologies used for conventional drives Kempf et al. (1993); Sacks et al. (1995); Wu and Tomizuka (2006); Chen et al. (2006) cannot be applied to BPMR. A schematic of the ideal trajectory for these two types of magnetic recording is shown in Fig. 2.7. In the figure, servo tracks (in circumferential direction) determine the desired trajectories to be tracked in BPMR. The shape of each individual BPMR servo/data track is patterned on the disk using electron-beam lithography (EBL) or some form of nano-lithography process, and its variations relative to a perfect circular track result in written-in runout which becomes repeatable (RRO) due to the disk spinning. The BPMR written-in RRO caused by EBL, assembly eccentricity, spindle vibration and disk normal fluctuation, usually contains high frequency components, and must be accurately tracked by the servo-system. Challenges in regards to control design for BPMR mainly arise due to the following RRO specifications:

- The RRO profile is unknown and its frequency spectrum spreads beyond the bandwidth of the servo system. Therefore, the tracking error will be amplified by the feedback controller at high frequencies.

- The RRO spectrum contains many harmonics – approximately 200 in current HDDs – of the spindle frequency that should be attenuated. This requires computationally intensive control methods.
- RRO profile varies across the span of the disk and the controller should be able to quickly adapt to a new track.
- The actuators dynamics vary across the HDDs of the same product line. Even the dynamics of different MAs in the same drive can be considerably different. The controller should be robust to these uncertainties.
- The actuators dynamics, especially the phase response, significantly changes by temperature at high frequencies.

The proposed control methodologies in this dissertation not only come with theoretical guarantees, but have also enabled us to address all aforementioned issues. All proposed adaptive control algorithms are implemented on a digital signal processor (DSP) unit and applied to an actual HDD. We have achieved full spectrum compensation of RRO that involves 174 frequency contents from 120Hz to over 20KHz. Our design is "modular", meaning that in case enough computation power is not available in the embedded processor, the controller can be split to "smaller" nonconcurrent controllers such that each one compensates a portion of the spectrum. As for the system dynamics variations, and uncertainties, our proposed algorithms are able to adapt to the system dynamics when it does not match our models or when it changes over time.

2.4 Experimental Setup

A block diagram of the experimental setup we use for implementing the proposed algorithms and applying them to a dual–stage hard disk drive is shown in Fig. 2.8. The HDD under our study is modified by our industry partner such that it provides position error signal through a serial peripheral interface (SPI) in real–time. The drive has 9 actuators, namely 1 VCM and 8 MAs that are located on both sides of 4 magnetic disks. Only one of the MAs is active at a time. The modifications in the HDD printed circuit board provide two gates for plugging our external adaptive controller to the internal closed loop system. Through these two gates, the plug–in control signals can be added, individually, to the input of VCM and only one of the MAs that is selected through the "special commands" sent from the "Host computer for the HDD". Indeed, this architecture is similar to the block–diagram that was earlier shown in Fig. 5.1. It is also possible to change the track that we servo on by sending a special command from the HDD host computer. This is equivalent to changing the RRO (disturbance r in Fig. 5.1).

We use a LOGIC PD Zoom^{TM} OMAP–L138 EVM Development Kit as the baseboard for a digital signal processor (DSP) that will execute the algorithms. OMAP–L138 EVM

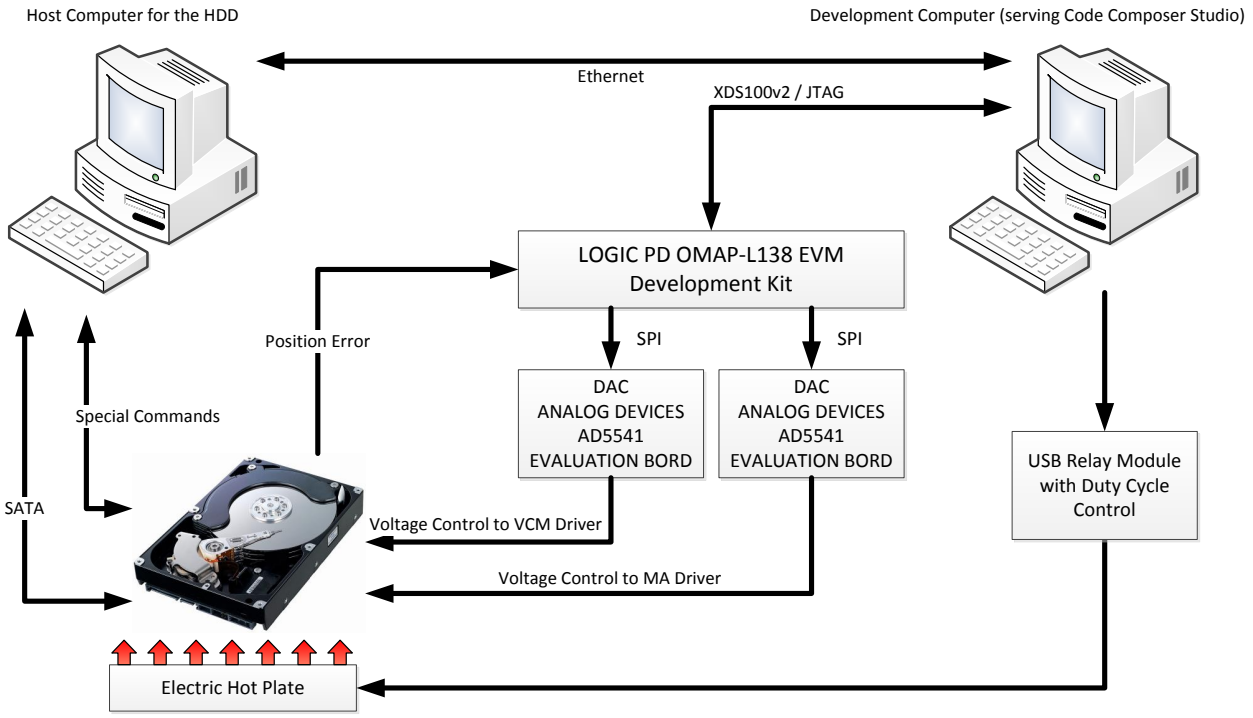


Figure 2.8: Block diagram of the experimental setup.

Development Kit is a high–performance application development kit for evaluating the functionality of Texas Instruments' (TI) energy–efficient OMAP–L138 applications processor, TMS320C6748 digital signal processor, and Logic PD's System on Module (SOM). The kit includes two SOMs, namely OMAP–L138 SOM–M1 and C6748 SOM–M1. The TMS320C6748 fixed– and floating–point DSP is a low–power applications processor based on a C674x DSP core. Our application development is performed on C6748 SOM–M1 and software Board Support Package included in the kit. The CPU frequency for this DSP can be set to 375 or 456MHz and the higher rate is chosen in our setup. The device DSP core uses a 2–level cache–based architecture. The level 1 program cache (L1P) is a 32–KB direct mapped cache, and the level 1 data cache (L1D) is a 32–KB 2–way, set–associative cache. The level 2 program cache (L2P) which is the target location for our codes consists of a 256–KB memory space. The algorithms are implemented in `C` and `C++` languages in Code Composer Studio^{TM} integrated development environment that interfaces with the development board through an XDS100v2 emulator.

The development kit provides an interface for digital to analog conversion (DAC). However, this interface is customized for audio processing applications and can only support special sampling rates which are not equal to the sampling rate of the PES specified by the HDD. Accordingly, we use two external DAC evaluation boards (AD5541) from Analog Devices to actuate the VCM and MA. The AD5541 is a single, 16–bit, serial input, voltage output DAC with $1\mu s$ settling time. It operates from a single 5V$\pm$ 10% supply and utilizes a versatile

3–wire interface. Communication between these boards and the DSP is performed through one serial peripheral interface with individual chip–select commands for each of the DAC boards. The control signal is passed to the HDD in voltage and the internal proportional drivers in the HDD convert it to current for the VCM and a scaled voltage for the MA.

We will show in figures 4.1 and 5.2 that the HDD system dynamics changes by temperature which can degrade the controller performance if not compensated. In order to study the behavior of our algorithms in such cases, we use an electric hot–plate under the HDD to vary its temperature. The hot–plate generated heat energy is controlled (in open loop) by a relay module that can set the duty cycle of its supply power.

Chapter 3

Adaptive Feedforward Repetitive Control for Systems with Known Dynamics

3.1 Introduction

Control methodologies for coping with periodic signals, commonly known as repetitive controllers, were first introduced in 1980's, and since then have been widely used in applications in which a task should be performed repeatedly, a periodic disturbance must be attenuated or a periodic trajectory must be tracked Chew and Tomizuka (1989). These methods have been applied in many robot manipulators applications, thermal cycling, milling machines and satellite altitude control. For instance, repetitive control has been applied in hard disk drives to follow periodic trajectories (or reject periodic disturbances), in order to read/write data on a magnetic disk surface. A more detailed list of applications and references to prior work in this field is provided in chapter 1.

Repetitive controllers are typically categorized into two types, namely feedback methods that are based on internal model principle (IMP) Francis and Wonham (1976) and feedforward algorithms that usually use an external model Tomizuka et al. (1990) or a *reference* signal correlated to the disturbance. The former class applies the internal model principle within a model reference or pole–placement control strategy. When the plant dynamics or disturbance frequency (in narrow–band case) is unknown or slowly time varying, adaptive versions of these methods are used. Among the various approaches that have been taken for solving this problem, one can mention: (1) indirect feedback adaptive control/regulation scheme that estimates in real time a model for the disturbance or plant dynamics and then recomputes the controller parameters Feng and Palaniswami (1992). (2) direct feedback adaptive regulation that utilizes the Youla–Kucera parametrization (also known as the Q–parametrization) of the controller to insert the internal model to the controller and adjust it by adapting the parameters of the Q–polynomial Landau et al. (2005); Chen and Tomizuka (2012). The effectiveness of these methods in rejecting a moderate number (e.g. less than 10) of narrow–band disturbances has been proved in practice Landau et al. (2013).

However, to the best of our knowledge, the practical viability of these algorithms for

rejecting multitude disturbances (e.g. 50 narrow band disturbances), especially with high frequency spectrum (i.e. very close to the Nyquist frequency of the digital system) has not been reported. Most, if not all, of these methods introduce poles on the stability boundary which is not desirable and can cause poor numerical properties and instability when implemented on an embedded system with finite precision arithmetic. Instability can also happen due to unmodeled dynamics when the poles are on or very close to the stability boundary. A practical solution to avoid marginally stable poles is to push them inside the unit circle by utilizing *acausal* filters or widening the bandwidth of peak filters. However, according to Bode's sensitivity integral theorem Bode et al. (1945), this approach will change the baseline loop shape considerably when the number of narrow–band disturbances is large. Hence, it may result in an unsatisfactory performance against other disturbances. Furthermore, perfect (deterministic) disturbance rejection is not achieved by applying this consideration. For disturbances with unknown, but periodic profiles, these methods are not efficient since they generate the control signal in a feedback interconnection. This means that even though the required controller is deterministic and the controller can learn it quickly, the controller has to be always kept in the loop to generate the control signal.

In general, the adaptive feedforward algorithms applied to this class of problems do not have the above limitations since their dependency on the error signal becomes less and less as their parameters converge. Indeed, when the system is not stochastic or the adaptation gain is vanishing in a stationary (or cyclostationary) and stochastic system, the control signal becomes a pure feedforward action that can be stored and then generated without a need to feeding back the error to the controller. Two general approaches exist for adaptive feedforward cancellation of periodic disturbances: (1) an approach that uses an additional sensor located properly in "upstream". This sensor picks up a signal highly correlated to the disturbance before the disturbance propagates through the system. The controller then generates a destructive signal to compensate for the disturbance. The drawback with these algorithms is possible instability caused by positive feedback when the compensation propagates back to the reference sensor (c.f. chapter 15 in Landau et al. (2011b)). (2) an approach that generates a "reference" signal by knowing the frequency of narrow–band/sinusoidal disturbances. In most applications the disturbance frequencies are known or can be measured by a sensor that is not influenced by the control filed, e.g. by a tachometer or accelerometer. When neither of these information types is in hand, the frequencies can be estimated by an algorithm such as adaptive notch–filter based frequency estimation Bodson and Douglas (1997) or a phase–locked–loop–like method Wu and Bodson (2003, 2004).

The adaptive feedforward compensation algorithms in the second approach, especially the algorithms used for Active (acoustic) Noise Cancellation (ANC), are mostly based on Least Mean Squares (LMS) algorithm. LMS is a gradient descent method that was first developed by Widrow and Hoff WIDROW et al. (1960) in 1960's and began to thrive quickly due to its simplicity and stable behavior when implemented with finite–precision arithmetic Diniz (1997). However, the algorithm may diverge when the gradient of cost function with respect to the parameters is not accurate. More explicitly, in applications that the control signal traverses a secondary path, as the one in Fig. 3.1, to affect the error signal, the true gradient

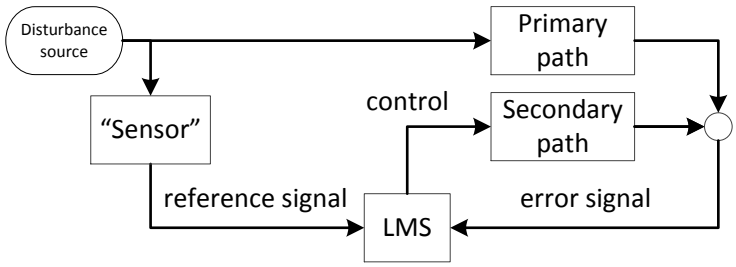


Figure 3.1: LMS-based adaptive controller

of mean squared error (w.r.t the control parameters) is a function of the system dynamics – a fact that is ignored in the original LMS algorithm. Therefore, when the phase response of system dynamics increases, the gradient in LMS algorithm becomes less and less informative and even misleading. More formally, if the secondary path has a phase response above 90 degrees in absolute value, the gradient used in LMS updates the controller coefficients in a direction that increases the cost function. The well known *Filtered–X Least Mean Squares (FX–LMS)* algorithm and modified versions of it attempt to overcome this issue by aligning the gradient used in the LMS algorithm with the real gradient through filtering the reference signal. This technique will be described in detail later in the section.

Our work in this section is based on the second type of adaptive feedforward algorithms. We deploy a stochastic gradient descent method, adopted from FX–LMS algorithm, to develop the parameter adaptation algorithm for an adaptive multiple narrow–band disturbance compensator (trajectory tracker). We first introduce the prior work on multiple narrow–band disturbance rejection inherited from FX–LMS algorithm. It is followed by developing a Modified FX–LMS (MFX–LMS) algorithm for multiple narrow–band or periodic disturbance rejection. Our main contribution in this section can be summarized as

- A novel adaptive controller is developed based on a modified FX–LMS algorithm to perfectly reject (track) disturbances (trajectories) that are periodic or equivalently consisted of multiple sinusoids. The controller is a "plug–in" device to an existing control system. Therefore, the original controller can be designed without consideration of the periodic disturbances (trajectory). Moreover, the adaptive controller does not alter the performance of the original control system.

- The stability of the algorithm is analyzed and guaranteed by averaging theory.

- A novel variable adaptation step size is proposed and integrated to the adaptive control algorithm for convergence rate enhancement and steady state error reduction.

- Four variants of the controller that factorize the disturbance/trajectory in frequency–domain, time–domain (2 cases), and time–frequency–domain are described and compared. To the best of our knowledge, the basis functions used in two of these variants have not been reported in related work.

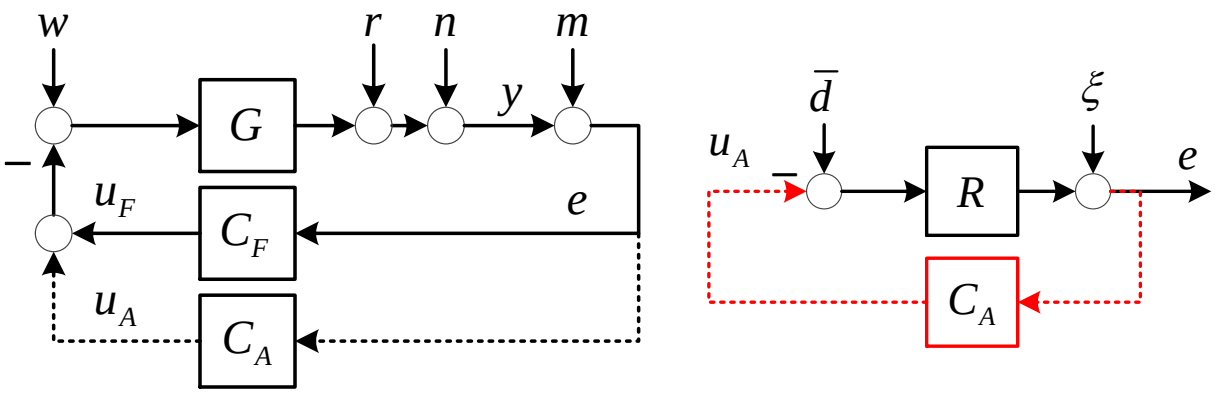


Figure 3.2: Left: Closed loop system augmented by a *plug–in* adaptive controller denoted by C_A. Right: Succinct representation of the closed loop system.

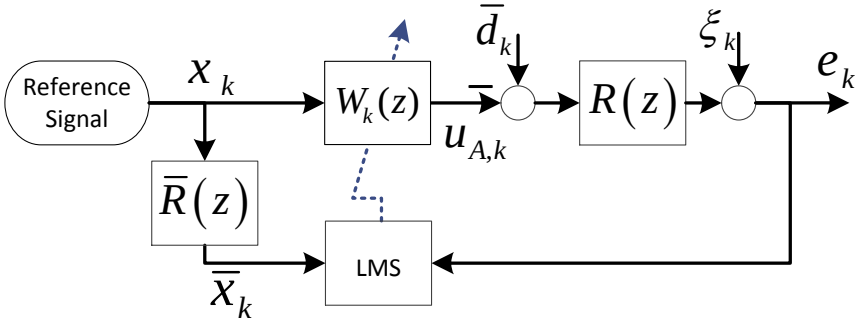


Figure 3.3: Adaptive feedforward control scheme based on FX–LMS algorithm.

3.2 Preliminaries: Filtered–X LMS Algorithm

The block diagram for the system under our study that was introduced by Fig. 1.2 is repeated in Fig. 3.2 for convenience. It has been shown that when the secondary–path $R(z)$ — i.e. the transfer function from the control input to the error signal — is not simply a static gain, the standard LMS algorithm will generally cause instability Elliott and Nelson (1985). This is because the error signal is not correctly "aligned" in time with the reference signal and as a result the update direction of parameters is not aligned with the negative of the actual gradient of the mean squared error with respect to the parameters.

The key idea of FX–LMS algorithm is to "align" the update direction with (the negative of) the actual gradient. The block diagram for an adaptive controller based on FX–LMS algorithm is shown in Fig. 3.3. The adaptive filter attempts to minimize the instantaneous square error (e_k^2) rather than the expected square error $\mathbf{E}[e_k^2]$. Therefore, the cost function is

$$J_k = e_k^2,$$

and the parameters of the adaptive digital controller $W_k(z)$ should be adjusted by the LMS

algorithm such that

$$J_k = \left(R\left[\bar{d}_k - W_k\left[x_k\right]\right] + \xi_k\right)^2$$

is minimized. The digital controller $W_k(z)$ is an FIR filter

$$W_k(z) = w_{k,0} + w_{k,1}z^{-1} + \cdots + w_{k,n_w}z^{-n_w}$$

that should ideally converge to $\mathcal{Z}\{\bar{d}_k\}/\mathcal{Z}\{x_k\}$ to perfectly cancel the effect of $\bar{d}_k$. The response of $W_k(z)$ to the reference signal x_k is

$$u_{A,k} = W_k\left[x_k\right] = \underbrace{\begin{bmatrix} x_k \\ x_{k-1} \\ \vdots \\ x_{k-n_w} \end{bmatrix}^T}_{\phi_k} \underbrace{\begin{bmatrix} w_{k,0} \\ w_{k,1} \\ \vdots \\ w_{k,n_w} \end{bmatrix}}_{\hat{\theta}_k} = \phi_k^T\hat{\theta}_k.$$

The parameters of the controller, $\hat{\theta}$, are supposed to be updated by the steepest descent method which moves the parameters in the negative gradient direction

$$\hat{\theta}_{k+1} = \hat{\theta}_k - \frac{\mu}{2}\nabla_{\hat{\theta}_k} J_k.$$

Here, μ is a scalar step size and $\nabla_{\hat{\theta}_k} J_k$ is the cost function gradient w.r.t to the parameters

$$\begin{aligned}\nabla_{\hat{\theta}_k} J_k &= \nabla_{\hat{\theta}_k} e_k^2 \\ &= 2e_k\nabla_{\hat{\theta}_k}\left(R\left[\bar{d}_k - W_k\left[x_k\right]\right] + \xi_k\right) \\ &= -2e_k\nabla_{\hat{\theta}_k}\left(R\left[\phi_k^T\hat{\theta}_k\right]\right)\end{aligned}$$

and the update rule is

$$\hat{\theta}_{k+1} = \hat{\theta}_k + \mu\nabla_{\hat{\theta}_k}\left(R\left[\phi_k^T\hat{\theta}_k\right]\right)e_k. \tag{3.1}$$

The FX–LMS algorithm adapts this update rule and makes the assumption that the step size μ is small compared to

$$\frac{\|\hat{\theta}_k\|}{\|\nabla_{\hat{\theta}_k}\left(R\left[\phi_k^T\hat{\theta}_k\right]\right)e_k\|}$$

for all values of k. Since the step size is small, the parameters are adapting slowly and $\hat{\theta}_k$ can be brought out of the brackets in 3.1

$$R\left[\phi_k^T\hat{\theta}_k\right] \approx R\left[\phi_k\right]^T\hat{\theta}_k.$$

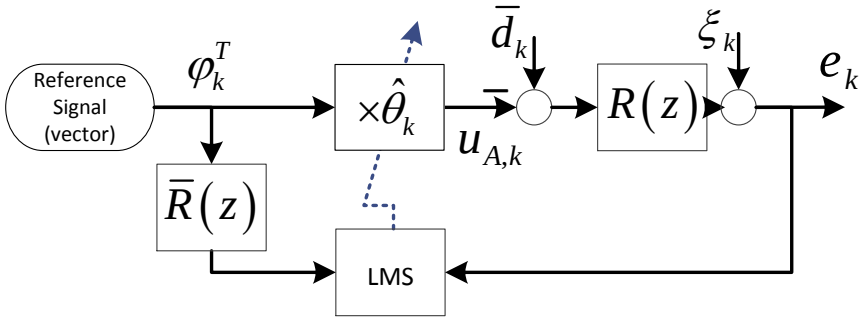


Figure 3.4: Adaptive feedforward control scheme based on FX–LMS algorithm (vector representation).

Note that in the original update rule, the term $\nabla_{\hat{\theta}_k}\left(R\left[\phi_k^T\hat{\theta}_k\right]\right)$ is zero when the transfer function $R(z)$ has a positive relative degree, meaning that $R\left[\phi_k^T\hat{\theta}_k\right]$ does not depend on $\hat{\theta}_k$. However, as will be explained later, by choosing a small enough step size that is inversely related to the delay of $R(z)$, we have the following approximation

$$\nabla_{\hat{\theta}_k}\left(R\left[\phi_k^T\hat{\theta}_k\right]\right) \approx \nabla_{\hat{\theta}_k}\left(R\left[\phi_k\right]^T\hat{\theta}_k\right) = R\left[\phi_k\right].$$

In many practical applications the exact dynamics of $R(z)$ is not known and an approximate model $\bar{R}(z)$ should be used. Let

$$\bar{\psi}_k := \bar{R}\left[\phi_k\right] = \begin{bmatrix}\bar{x}_k & \bar{x}_{k-1} & \cdots & \bar{x}_{k-n_w}\end{bmatrix}^T$$

where $\bar{x}_k = \bar{R}\left[x_k\right]$. By the above two approximations, the update rule of FX–LMS algorithm can be written as

$$\hat{\theta}_{k+1} = \hat{\theta}_k + \mu\bar{\psi}_k e_k$$

which is illustrated in a block diagram in Fig. 3.4.

This algorithm has been applied to narrow band disturbance rejection mostly for acoustic noise cancellation. When the frequency of disturbance is known, the reference signal x_k or equivalently the regressor ϕ_k can be generated without needing an extra sensor. Two types of reference signal have been considered for this purpose in the literature. The first signal is a periodic train impulse that has a period equal to the inverse of the fundamental frequency of the periodic noise Elliott and Darlington (1985); Chaplin (1980). This approach cannot be applied when the disturbance frequencies are not multiple integers of a fundamental frequency, or when selective frequencies should be compensated except that the least common multiple of all frequencies is considered as the fundamental frequency. However, this results slow convergence rate when the least common multiple is very large. The second type of reference signal is consisted of sinusoidal signals which does not have these limitations. However, its computational complexity is higher. The accuracy of the aforementioned

approximations depends on the step size and as is shown by Elliott and Nelson (1993) the maximum step size that can be used is approximately

$$\mu_{\max} \approx \frac{1}{\mathbf{E}\left[\bar{x}_k^T \bar{x}_k\right](n_w + \Delta)}$$

where Δ is the number of samples corresponding to the overall delay in the secondary path. This limitation originates from the step we took to approximate $e_k = R[d_k] - R[\phi_k^T \hat{\theta}_k]$ by $R[d_k] - R[\phi_k]^T \hat{\theta}_k$. The MFX–LMS algorithm that will be discussed in the next part does not need this approximation since it directly considers $\bar{e}_k := R[d_k] - R[\phi_k]^T \hat{\theta}_k$ as the error signal. Motivated by the fact that MFX-LMS algorithm has benefit of being robust as FX–LMS, and fast as LMS algorithm, we pursue this method to synthesis an adaptive repetitive controller for rejecting periodic disturbances at desired frequencies.

3.3 Adaptive Control Synthesis

Recall that our main objective in the framework outlined in Fig. 3.2 is to synthesize an adaptive control law for generating u_A such that the effect of disturbance $\bar{d}$ on the error signal e is minimized. This objective can be ideally achieved when the control signal is equal to the disturbance $\bar{d}$. As mentioned in the introduction, our analysis revolves around a special factorization of the disturbance

$$\bar{d}_k = \theta^T \phi_k$$

where ϕ_k forms a basis function and can be generated in different manners for the special case of periodic and narrow–band disturbances. This factorization implies that perfect compensation can be achieved if the control algorithm learns the unknown parameter vector θ perfectly. We begin our analysis in a spectral framework, where trigonometric functions can form an orthogonal basis for the space of real valued square integrable functions. Nevertheless, after outlining the algorithm, we revisit the problem for other choices of basis function and compare them.

In a spectral analysis framework, the disturbance signal can be represented as a summation of orthogonal pairs of sinusoidal and cosinusoidal functions with zero initial phases

$$\bar{d}_k = \sum_{i=1}^{n} \left[a_i \left(\alpha_i \sin\left(\omega_i k T\right)\right) + b_i \left(\alpha_i \cos\left(\omega_i k T\right)\right)\right] = \theta^T \phi_k \tag{3.2}$$

where ω_i's and n respectively denote the frequencies and the number of components that are desired to be compensated. The constant T is the sampling time of the discrete time system. The vector of unknown parameters θ, and the regressor vector ϕ_k at time step k are defined as

$$\theta^T := [a_1, \cdots, a_n, b_1, \cdots, b_n] \tag{3.3}$$

$$\begin{aligned} \phi_k^T := [&\alpha_1 \sin\left(\omega_1 k T\right), \cdots, \alpha_n \sin\left(\omega_n k T\right), \\ &\alpha_1 \cos\left(\omega_1 k T\right), \cdots, \alpha_n \cos\left(\omega_n k T\right)] \cdot \end{aligned} \tag{3.4}$$

Here, α_i's are a set of positive weighting parameters that should be chosen by the designer. We will suggest an explicit method for choosing these parameters later in this chapter. Note that the regressor vector ϕ_k is known since the compensation frequencies, ω_i's, and weighting coefficients, α_i's, are known. Based on the factorization in (3.2), we propose an adaptive algorithm that at any time step k obtains an estimate of the unknown parameter vector, say $\hat{\theta}_k$, and constructs the control signal u_A by

$$u_{A,k} = \hat{\theta}_k^T \phi_k. \tag{3.5}$$

This control law can achieve perfect compensation if $\hat{\theta}_k$ converges to θ. A parameter adaptation algorithm for this purpose is given in the following.

3.3.1 Parameter Adaptation Algorithm

We deploy a gradient descent method inspired by the Filtered–X LMS algorithm to adaptively identify the parameter vector θ based on the information provided by the error signal. Based on Fig. 3.2, the error signal is

$$\begin{aligned} e_k &= R\left[\bar{d}_k - u_{A,k}\right] + \xi_k = R\left[\phi_k^T\theta - u_{A,k}\right] + \xi_k \\ &= \underbrace{R\left[\phi_k\right]}_{\psi_k}{}^T\theta - R\left[u_{A,k}\right] - \underbrace{R[\phi_k]}_{\psi_k}{}^T\hat{\theta}_k + \underbrace{R[\phi_k]}_{\psi_k}{}^T\hat{\theta}_k + \xi_k \\ &= \psi_k^T\theta - \psi_k^T\hat{\theta}_k + \xi_k - R\left[u_{A,k}\right] + \psi_k^T\hat{\theta}_k. \end{aligned} \tag{3.6}$$

We denote the signal $\psi_k := R[\phi_k] \in \mathbb{R}^{2n}$ as the *filtered regressor*. Suppose that the transfer function $R(z)$ is known; then, the filtered regressor is known as well, and the error signal is an affine function of our estimated parameter vector, $\hat{\theta}_k$, if the last term on the right hand side of (3.6) is omitted. This suggests defining an *auxiliary error* signal

$$\bar{e}_k := e_k + R\left[u_{A,k}\right] - \psi_k^T\hat{\theta}_k \tag{3.7}$$

which is explicitly an affine function of our estimated parameter vector

$$\bar{e}_k = \psi_k^T\theta - \psi_k^T\hat{\theta}_k + \xi_k.$$

The *auxiliary error* signal is a good measure of the original *error* signal because they are equivalent when the parameters converge. Suppose that the estimated parameter vector converges to $\hat{\theta}_{ss}$. The two signals will be equal because

$$\begin{aligned} \bar{e}_k - e_k &= R[\phi_k]^T\hat{\theta}_{ss} - \psi_k^T\hat{\theta}_{ss} \\ &= \psi_k^T\hat{\theta}_{ss} - \psi_k^T\hat{\theta}_{ss} = 0\,. \end{aligned} \tag{3.8}$$

Accordingly, rather than minimizing the instantaneous square *error* signal e_k^2, we use a gradient descent algorithm to minimize a cost function $\bar{J}_k$ that equals to the square of

instantaneous *auxiliary error* signal $\bar{e}_k$

$$\begin{aligned} \bar{J}_k &= \bar{e}_k^2 \\ \hat{\theta}_{k+1} &= \hat{\theta}_k - \frac{\mu}{2}\nabla_{\hat{\theta}_k}\bar{J}_k. \end{aligned} \tag{3.9}$$

Note that this update rule is analogous to the LMS algorithm. Accordingly, one can show that it has similar good properties such as as low computational complexity, convergence in stationary environment, and under independence theory assumption – which is satisfied here – the parameters converge in the mean to the Wiener solution. Moreover, this algorithm is suitable for DSP implementation of the specific application we discussed in section 2.3 since it is shown that the LMS algorithm has stable behavior when implemented with finite–precision arithmetic. The update rule in 3.9 can be simplified further

$$\begin{aligned} \hat{\theta}_{k+1} &= \hat{\theta}_k - \frac{\mu}{2}\nabla_{\hat{\theta}_k}\bar{J}_k \\ &= \hat{\theta}_k - \frac{\mu}{2}\left[-2\bar{e}_k\nabla_{\hat{\theta}_k}\left(\psi_k^T\theta - \psi_k^T\hat{\theta}_k + \xi_k\right)\right] \\ &= \hat{\theta}_k + \mu\psi_k\bar{e}_k. \end{aligned} \tag{3.10}$$

Theorem 1
Let ψ_k^i be the i^{th} component of ψ_k and define $\bar{\alpha}_i^2 := (\psi_k^i)^2 + (\psi_k^{i+n})^2$ for all values of i from 1 to n. Suppose the parameter adaptation algorithm in (3.10) has a variable step size μ_k that is bounded by

$$0 < \mu_{\min} < \mu_k < \mu_{\max} := \min_{i\in\{1,\cdots,n\}}\left[\frac{4}{\bar{\alpha}_i^2}\right] \tag{3.11}$$

for all values of $k \geq 0$ and some $\mu_{\min}$. Then, the estimated parameter vector updated by (3.10) converge to the real parameter vector θ in mean value with an exponential rate.

Proof
We use the discrete–time averaging theory for mixed time–scale systems Bai et al. (1988). *To follow the same notation as* Bai et al. (1988), *let $x(k) := \theta - \mathbf{E}\left[\hat{\theta}_k\right]$ be a state vector described by a difference equation of the form*

$$x(k+1) = x(k) + \epsilon f\left(k, x(k), \epsilon\right).$$

The averaging theory relates the solution of this difference equation to the solution of the so–called "averaged" system

$$x_{av}(k+1) = x_{av}(k) + \epsilon f_{av}\left(x_{av}(k)\right).$$

where

$$f_{av}(x) = \lim_{T\to\infty}\frac{1}{T}\sum_{k=k_0+1}^{k_0+T} f\left(k, x, 0\right).$$

and $0 < \epsilon \leq \epsilon_0$ *for some small value* ϵ_0*. Let* $\mu_k = \epsilon\tilde{\mu}_k$*. Based on the parameter adaptation algorithm given in* (3.10)*, we have*

$$\begin{aligned}
\epsilon f(k, x(k), \epsilon) &= x(k+1) - x(k) \\
&= \mathbf{E}\left[\hat{\theta}_k - \hat{\theta}_{k+1}\right] \\
&= -\mathbf{E}\left[\mu_k \psi_k \bar{e}_k\right] \\
&= -\mu_k \psi_k \mathbf{E}\left[\psi_k^T \left[\theta - \hat{\theta}_k\right] + \xi_k\right] \\
&= -\mu_k \psi_k \psi_k^T x(k)
\end{aligned}$$

where the last equality follows from the fact that ξ_k *is a zero mean random variable and* ψ_k *is a deterministic signal. The function* $f(k, x, \epsilon) = -2\tilde{\mu}_k \psi_k \psi_k^T x$ *satisfies assumptions* **A1-A2** *in* Bai et al. (1988)*. Define*

$$A_{av} = \lim_{T\to\infty} \frac{1}{T} \sum_{k=k_0+1}^{k_0+T} \tilde{\mu}_k \psi_k \psi_k^T.$$

One can show that the limit exists and it is bounded. This is because the term $\tilde{\mu}_k \psi_k \psi_k^T$ *is bounded, and as a result, the series* $S(T) := \frac{1}{T}\sum_{k=k_0+1}^{k_0+T} \tilde{\mu}_k \psi_k \psi_k^T$ *is Cauchy. Therefore,* A_{av} *is the converging point of Cauchy series* $S(T)$ *and assumption* **A3-A4** *of the averaging theory hold true. Moreover, the "averaged dynamics" can be simplified to*

$$x_{av}(k+1) = (I - \epsilon A_{av})\, x_{av}(k).$$

The matrix A_{av} *is* Positive Semi–Definite (PSD) *since it is the superposition of a set of PSD matrices. An upper–bound (in a positive definiteness framework) can be obtained*

$$\begin{aligned}
\epsilon A_{av} &= \lim_{T\to\infty} \frac{1}{T} \sum_{k=k_0+1}^{k_0+T} \mu_k \psi_k \psi_k^T \\
&\prec \mu_{\max} \left[\lim_{T\to\infty} \frac{1}{T} \sum_{k=k_0+1}^{k_0+T} \psi_k \psi_k^T\right] \\
&= \frac{\mu_{\max}}{2} \begin{bmatrix}
\bar{\alpha}_1^2 & \cdots & 0 & 0 & \cdots & 0 \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & \cdots & \bar{\alpha}_n^2 & 0 & \cdots & 0 \\
0 & \cdots & 0 & \bar{\alpha}_1^2 & \cdots & 0 \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & \cdots & 0 & 0 & \cdots & \bar{\alpha}_n^2
\end{bmatrix}.
\end{aligned} \tag{3.12}$$

With the same procedure we can show that if $0 < \mu_{\min} < \mu_k$ *the matrix* $I - \epsilon A_{av}$ *is positive definite. Therefore, we have* $0 \prec \epsilon A_{av} \prec 2I$ *which in accordance with* (3.11) *imply that*

$I - \epsilon A_{av}$ is a Schur matrix and $x_{av}(k)$ converges to zero exponentially. Since all assumptions ***A1-A4*** *are satisfied, by Theorem 2.2.1 in* Bai et al. (1988) *we have* $\lim_{k\to\infty} \mathbf{E}\left[\hat{\theta}_k\right] = \theta$. *Moreover, the convergence rate is exponential and determined by the eigenvalues of* $(I - \epsilon A_{av})$. □

***A–Posterior* PAA:**

The PAA proposed above generates the estimates of step k based on an *a–prior* auxiliary error that depends on the current step estimates. We can define an *a–posteriori* auxiliary error, say $\bar{e}_{k,p}$, such that $\bar{e}_{k,p}$ at time step k depends on the next step values of estimates $\hat{\theta}_{k+1}$

$$\bar{e}_{k,p} := \psi_k^T\theta - \psi_k^T\hat{\theta}_{k+1} + \xi_k.$$

By following a similar procedure as the *a–priori* case we have

$$\begin{aligned}\nabla_{\hat{\theta}_{k+1}}\bar{e}_{k,p}^2 &= 2\psi_k\bar{e}_{k,p}\\ \hat{\theta}_{k+1} &= \hat{\theta}_k + \mu\psi_k\bar{e}_{k,p}\end{aligned}$$

However, since this expression is not realizable directly, we have to relate it to the *a–priori* auxiliary error

$$\begin{aligned}\bar{e}_{k,p} &= \psi_k^T\theta - \psi_k^T\hat{\theta}_{k+1} + \xi_k\\ &= \left(\psi_k^T\theta - \psi_k^T\hat{\theta}_k + \xi_k\right) + \left(\psi_k^T\hat{\theta}_k - \psi_k^T\hat{\theta}_{k+1}\right)\\ &= \bar{e}_k - \mu\psi_k^T\psi_k\bar{e}_{k,p}\\ &= \frac{1}{1+\mu\psi_k^T\psi_k}\bar{e}_k.\end{aligned}$$

Using this last expression the PAA can be written in a causal form

$$\hat{\theta}_{k+1} = \hat{\theta}_k + \frac{\mu}{1+\mu\psi_k^T\psi_k}\psi_k\bar{e}_k. \tag{3.13}$$

Theorem 2
The adapted parameters in (3.13) *converge in mean value to θ for any positive constant or variable step size μ_k.*

Proof
The proof follows from the analysis in the proof of theorem 1. One needs to use $\psi_k^T\psi_k = \sum_{i=1}^n \bar{\alpha}_i^2$ *and*

$$\frac{\mu_k}{1+\mu_k\psi_k^T\psi_k} \le \frac{1}{\psi_k^T\psi_k} = \frac{1}{\sum_{i=1}^n \bar{\alpha}_i^2} \le \min_{i\in\{1,\cdots,n\}}\left[\frac{4}{\bar{\alpha}_i^2}\right]$$

to prove the claim. □

3.3.2 Convergence Rate Based on Weighting Parameters

Since the transfer function $R(z)$ is LTI, the filtered regressor in a steady state, ψ_k, can be simply calculated by knowing the frequency response of $R(z)$

$$\begin{aligned} \begin{bmatrix} \psi_k^i \\ \psi_k^{i+n} \end{bmatrix} &= m_i \alpha_i \begin{bmatrix} \sin(\omega_i k T + \delta_i) \\ \cos(\omega_i k T + \delta_i) \end{bmatrix} \\ m_i &:= |R\left(e^{j\omega_i T}\right)| \\ \delta_i &:= \angle R\left(e^{j\omega_i T}\right) \end{aligned} \tag{3.14}$$

for $i \in \{1, \cdots, n\}$. Here, $|x|$ and $\angle x$ denote the magnitude and phase of the complex number x. It is well known that the convergence rate of the LMS algorithm depends on the eigenvalue spread of the regressor correlation matrix Ungerboeck (1972). This was also shown in our proof for theorem 1. This fact suggests that the values of α_i's in (3.3) should be chosen such that the amplitude of all sinusoidal elements in (3.14) are equal – i.e. $\alpha_i = \frac{c}{m_i}$, where c is a constant scalar. This constant can be chosen to be one because any other value of c can be incorporated in the step size μ_k

$$\alpha_i = \frac{1}{m_i}. \tag{3.15}$$

3.3.3 Variable Adaptation Step Size

Although the coefficient vector on average converges to Wiener solution, the instantaneous deviation in the parameter vector, caused by the noise ξ_k, generates an *excess* mean squared error (MSE) appearing in the variance of $\bar{e}_k$. More important, this parameter oscillation prevents the *auxiliary error* signal from converging to the *error* signal, the equality that was shown in (3.8) under a steady state assumption.

We propose an adaptive law to adjust the step size based on an estimation of the total mean squared error. The key idea behind this scheme is that, as the estimated parameters get closer to the real ones, the step size becomes smaller and the parameters will be frozen in time when a certain desired performance (in terms of the mean squared error) is attained. This removes the excess error from the output and results in smaller steady state errors. However, in a practical situation, the system dynamics or disturbance $\bar{d}$ may be subjected to variations, and it is required that the step size activates the adaptation whenever the error becomes "large" due to these variations.

We use a moving average with a window width of h to estimate the *auxiliary error* power at time step k

$$V_k^h = V_{k-1}^h + \frac{1}{h}\left[(\bar{e}_k)^2 - (\bar{e}_{k-h})^2\right]. \tag{3.16}$$

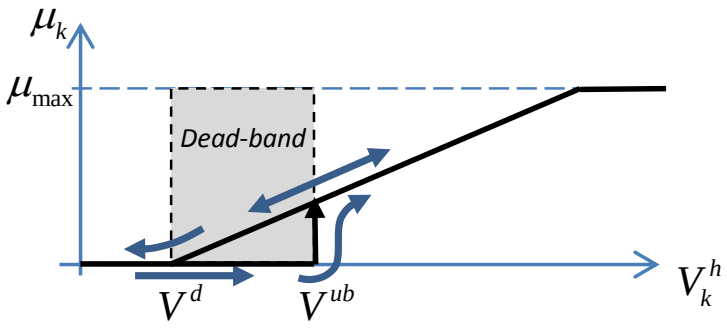


Figure 3.5: Adaptive variable step size with hysteresis behavior.

For a given *desired MSE* value, say V^d, we define the step size law as

$$\bar{\mu}_k = \rho\left(V_k^h - V^d\right) \tag{3.17}$$

$$\mu_k = \begin{cases} min(\bar{\mu}_k, \mu_{max}) & \text{if} \quad (\bar{\mu}_k > 0 \ \wedge \ \mu_{k-1} > 0) \vee (\bar{\mu}_k > \mu^{ub}) \\ 0 & \text{otherwise} \end{cases} \tag{3.18}$$

where the logical conjunction (*and*) and disjunction (*or*) are denoted by $\wedge$ and $\vee$ respectively. Constant ρ is a positive scalar gain, and the variable $\bar{\mu}_k$ determines how far the current error power is from the desired value V^d. Initial values of the parameters are not important and can be set to zero. Since V_k^h is not exactly equal to the auxiliary error variance, the system may show *chattering* behavior around the switch line (i.e. $V_k^h = V^d$) if $\bar{\mu}_k$ is used directly as the step size. To avoid this, we add a hysteresis behavior to the step size, which is defined by (3.18). The logic condition represented in (3.18) defines a *dead–band* $[V^d,\ V^{ub}]$ on the MSE surface, and represents a hysteresis behavior for the step size. That is, the adaptation is active as long as the estimated error power is above the *dead–band* ($V_k^h > V^{ub}$), and it is inactive whenever the error power falls behind the *dead–band* ($V_k^h < V^d$). Moreover, if the approximated error power enters the *dead–band* from the above, it stays active until it exits from the bottom (values smaller than V^d), and if the power error enters the *dead–band* from the bottom, it stays inactive as long it does not exceed the upper limit V^{ub}. To guarantee the convergence of the second moment of error, an upper limit μ_{max} is considered on the step size. It is well known that a sufficient condition for guaranteeing MSE convergence is to choose $\mu_{max} \leq 2/(3\psi_k^T\psi_k)$ for all values of k (see Feuer and Weinstein (1985) and note that the step size in that work is half of the step size used here). A schematic for the step size hysteresis behavior is shown in Fig. 3.5.

3.3.4 Scheduling Parameters

In the lack of a priori knowledge about the parameter values, the transient error may be large if many parameters are being updated simultaneously. This is mainly because the auxiliary error and actual error signals are not close as long as the estimated parameters variations are not small (c.f. (3.8)). This implies that despite the auxiliary error converges

to zero rapidly, the actual error is not necessarily small at transient period when there are multitude parameters being estimated.

To solve this issue, the adaptation of different parameters in transient can be scheduled in time. Let $\hat{\theta}_k^i$ represent the i-th element of $\hat{\theta}_k$. The parameter adaptation rule in (3.10) can be modified as

$$\begin{bmatrix} \hat{\theta}_{k+1}^i \\ \hat{\theta}_{k+1}^{i+n} \end{bmatrix} = \begin{bmatrix} \hat{\theta}_k^i \\ \hat{\theta}_k^{i+n} \end{bmatrix} + \gamma_k^i \mu_k \begin{bmatrix} \psi_k^i \\ \psi_k^{i+n} \end{bmatrix} \bar{e}_k, \quad 1 \leq i \leq n \tag{3.19}$$

where γ_k^i is a binary variable. It is one when the parameters corresponding to the i^{th} frequency should be updated and zero otherwise. Proving that this modification does not cause instability (i.e. divergence) is straightforward and similar to theorem 1. One needs to consider $\gamma_k^i \mu_k$ as a time–varying step size and follow the same procedure as in the proof of theorem 1.

From a practical point of view, the possibility of scheduling the adaptation is very appreciated. That is because in the case of multitude frequency disturbances the embedded processor may not be able to perform one step update of all parameters in the time interval between two consecutive samples of error signal. An example for choosing these parameters is given in section 4.2 (see (4.12)).

3.4 Variants of the Algorithm: Time–Domain and Time–Frequency–Domain Frameworks

We began our analysis in a spectral framework, where trigonometric functions were chosen to form an orthogonal basis for the space of real valued square integrable functions. The disturbance $\bar{d}_k$ in this space was decomposed to a a vector of unknown parameters θ and a known vector valued function (the regressor) ϕ_k that contained the sinusoidal and cosinusoidal functions (3.3). There are two alternatives to this approach, namely decomposition of the disturbance in time–domain and in time–frequency–domain. As a matter of fact, each of these three approaches considers a space that contains the disturbance signal and the basis of that space is represented by the regressor vector. This means that choosing each of these approaches is equivalent to choosing a special regressor vector to form the decomposition $\bar{d}_k = \theta^T \phi_k$. We consider three regressor vectors in this section and compare them to the one chosen in the spectral framework (3.3). The first two describe the disturbance in time–domain and the last one is in the form of a wavelet that represents the disturbance in a time–frequency domain.

3.4.1 Periodic Impulse Train (Time–Domain)

A "periodic impulse train" in discrete time is the summation of infinite (Kronecker) delta functions that form a periodic tempered distribution. Suppose the disturbance $\bar{d}_k$ is N–

periodic and let $\text{Ш}(k;t)$ be the associated periodic impulse train that has lag t

$$\text{Ш}(k;t) := \sum_{i=-\infty}^{\infty} \delta(k - iN - t).$$

The signal $\bar{d}$ can be written as a linear combination of periodic impulse trains

$$\bar{d}_k = \begin{bmatrix} \bar{d}_1 \\ \bar{d}_2 \\ \vdots \\ \bar{d}_N \end{bmatrix}^T \begin{bmatrix} \text{Ш}(k;1) \\ \text{Ш}(k;2) \\ \vdots \\ \text{Ш}(k;N) \end{bmatrix} = \theta^T \phi_k.$$

Note that the regressor vector ϕ_k in this framework creates the basis of a space such that the components on coordinate axes represent the values of signal in different time stamps. We can interpret this decomposition differently by thinking of θ as the coefficients of an N–tap FIR filter that transforms $\text{Ш}(k;1)$ to $\bar{d}_k$

$$\bar{d}_k = \left(\theta_1 + \theta_2 z^{-1} + \cdots + \theta_N z^{-N+1}\right) \left[\text{Ш}(k;1)\right].$$

Let T be the sampling frequency of the digital system under our study. The Fourier transformation (denoted by $\mathscr{F}(.)$) of the periodic impulse train contains all and only the harmonics of the periodic impulse train frequency

$$\left|\mathscr{F}\left[\text{Ш}(k,t)\right]\right| := \frac{1}{N} \sum_{i=0}^{\lceil N/2 \rceil - 1} \delta(\omega - \frac{2\pi}{T}\frac{i}{N}). \tag{3.20}$$

For instance, if the digital system is running at 40KHz and the impulse train is 10–periodic, the spectrum of impulse train has $\{0, 4, 8, 12, \cdots, 16\}$KHz contents (ignoring the Nyquist and aliased frequencies).

The sparse structure of periodic impulse train

$$\phi_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \phi_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \cdots, \quad \phi_N = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}, \quad \phi_N + 1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \ldots$$

makes it more computationally efficient than trigonometric regressor (3.3) when it is deployed in the proposed adaptive repetitive controller. This is because: (A1) unlike the trigonometric regressor vector, updating this periodic impulse regressor from step k to $k+1$ requires negligible effort.

(A2) the control signal synthesis step in (3.5) which is repeated below

$$u_{A,k} = \hat{\theta}_k^T \phi_k$$

is simply equivalent to choosing the $i(k)^{\text{th}}$ component of $\hat{\theta}_k$ where $i(1) = 1$ and

$$i(k+1) := \begin{cases} i(k)+1 & \text{if } i(k)+1 \leq N \\ 1 & \text{otherwise.} \end{cases}$$

(A3) the filtered regressor ψ_k is periodic in steady state. Hence, it can be calculated offline and then be implemented in the adaptive controller as a circular buffer.

Although this choice of regressor vector is very efficient in terms of computation, it has a few fundamental shortcomings that can make it unsuitable in some applications:
(D1) Based on (3.20) the period of the disturbance, N, is a function of the smallest common multiple of the sampling frequency $1/T$ and the frequency contents of the disturbance. For instance, suppose that the disturbance is a simple sinusoid at 19KHz and the sampling frequency of the system is 40KHz. The period of this disturbance in a discrete time domain is 40 steps (which is equivalent to 19 periods of the disturbance in continuous time). As a result, the number of required parameters can be very large when the system sampling frequency is not devisable by the frequency of disturbance components. Moreover, (3.20) says that each periodic impulse train can only capture a fundamental frequency $2\pi/T/N$ and its higher harmonics. If there are different frequency contents in the disturbance that are not harmonics of the same fundamental frequency the problem is even more tedious.
(D2) It is not a suitable approach for multi–input (e.g. dual–stage) systems since it is not possible to split (e.g. different frequency) contents of the disturbance between different input channels of the system. For instance, we will show in chapter 5 that using the trigonometric basis functions for RRO following in an HDD let us compensate the low frequency contents through the input of VCM and high frequency parts through the MA input channel. This type of separation cannot be performed when impulse train is used.
(D3) The trigonometric regressor (3.3) defines an orthogonal basis for the disturbance space. Therefore, it is possible to estimate different components of θ separately as in (3.19). The periodic train impulse regressor does not provide this property, meaning that all components of θ have to be estimated simultaneously. This may result in high transient error or slow convergence (depending on the choice of step size) when the period N is large. In other words, we cannot have the scheduling parameters γ_k^i in this framework.
(D4) The last deficiency is about the divergence of algorithm when the system dynamics is not known accurately. We will show in the next chapter that the convergence/divergence behavior of our algorithm only depends on the dynamics mismatches (between the actual and modeled dynamics) at compensation frequencies. Therefore, in a spectral framework, a selective frequency can be compensated if some robustness criteria (to be determined in theorem 3) are satisfied at that particular frequency. However, in the case of using periodic impulse train, it is needed to satisfy those criteria not only at the fundamental

frequency of the period, but also at all its higher harmonics (c.f. (3.20)). Note that in many practical applications the uncertainties increase by the frequency. Therefore, even if a model is accurate enough to satisfy the convergence criterion at the fundamental frequency, it may not satisfy it at higher harmonics. This can cause divergence in the case of using impulse train.

3.4.2 Periodic Impulse Train for the Filtered–Regressor (Time–Domain)

In the previous part the disturbance on **input** side was decomposed in a time–domain framework, and the computational cost of the algorithm was reduced considerably because of the sparse structure of regressor vector ϕ_k. A similar approach can be taken by decomposing the effect of the disturbance on the **output** side. That is, instead of decomposing the disturbance $\bar{d}_k$ in time domain, we can take the same approach on $\bar{e}_k$ and define a sparse ψ_k instead of ϕ_k. The advantage of this approach compared to the previous part is that the filtered regressor ψ_k appears more than ϕ_k in our algorithm. Therefore, the algorithm can be further simplified in terms of computation.

For a given sparse ψ_k it may not be possible to find ϕ_k such that $\psi_k = R[\phi_k]$. In particular, if $R(z)$ is not minimum phase, we will not be able to obtain $\phi_k = R^{-1}[\psi_k]$. We propose using an anti–causal *zero–phase inverse* of the closed loop system $R(z)$ when the system is not invertible Tomizuka (1987). Suppose that the closed loop system $R(z)$ is realized by

$$R(q^{-1}) = \frac{B(q^{-1})}{A(q^{-1})} = \frac{B^s(q^{-1})B^u(q^{-1})}{A(q^{-1})}$$

where q^{-1} denotes the *one step delay operator*. The polynomials $B^s(p)$ and $B^u(p)$ respectively have roots outside and inside (including on) the unit circle. Let n_s and n_u be the order of these two polynomials. We can say that $z^{n_s}B^s(z^{-1})$ and $z^{n_u}B^u(z^{-1})$ correspond to all minimum phase and non–minimum phase zeros of $R(z)$ respectively. Since the system has non–minimum phase zeros, it cannot be inverted perfectly. Nevertheless, an approximate inverse can be attained through the zero phase inverse technique Tomizuka (1987)

$$R^{\#}(z) := \frac{A(q^{-1})B^u(q^{-1})}{B^s(q^{-1})[B^u(1)]^2}.$$

Note that $B^u(q)$ is an anti–causal operator, meaning that its response at any time depends on the future values of input signal. This is a feasible operation since the future input to this system (filtered regressor) is known to us.

In order to construct a sparse filtered regressor, ψ_k, we first consider an ideal sparse function which is indeed a periodic impulse train

$$\psi^1_{k,ideal} := \text{Ш}(k, 1)$$

and then filter it by $R^{\#}(.)$

$$\phi^1_{k,ideal} := R^{\#}[\psi^1_{k,ideal}].$$

In steady state, the response is periodic and it determines the regressor vector

$$\phi^1_k = \lim_{i\to\infty} \phi^1_{k+iN,ideal} \quad 1 \le k \le N$$
$$\phi^T_k = \left[\phi^1_k, \quad \phi^1_{k-1}, \quad \phi^1_{k-2}, \quad \cdots, \quad \phi^1_{k-N}\right]$$

Now that we have ϕ_k, we can obtain $\psi_k = R\,[\phi_k]$ which is in principle close to the ideal filtered regressor that we considered initially. However, this vector valued function may have some very small values that make it unsparse. In this case, a threshold can be set and all components less than that are replaced by zeros to make the regressor sparse. We can think of this approximation as the effect of a small multiplicative uncertainty $\Delta(z)$

$$\psi_k = (R(z)\Delta(z))\,[\phi_k].$$

It will be shown in theorem 3 that the algorithm is very robust to this type of uncertainties. Moreover, the algorithm convergence after applying this approximation can be verified rigorously by the stability analysis provided in section 4.5.

In conclusion, a sparse filtered–regressor vector reduces the computational complexity of the algorithm more than a sparse regressor vector since the former one appears in (3.7) and (3.10), while the latter only shows up in (3.5). The advantages (A1) and (A3) stated in the previous approach are common between these two methods. As of the shortcomings, one can show that this method is subjected to all issues mentioned through disadvantages (D1) to (D4) in the previous section since both methods are inherently based on time–domain analysis.

3.4.3 Wavelet (Time–Frequency Domain)

The last approach belongs to a time–frequency domain framework that deploys wavelet–form regressor vectors. The key idea here is to take advantage of low computational complexity of time domain approaches and possibility of targeting selective frequencies provided by the spectral analysis. Similar to the spectral analysis at the beginning of this chapter, compensation at frequencies $\{\omega_1, \omega_2, \cdots, \omega_n\}$ is considered. We propose using a reference signal that superposes sinusoids at these frequencies:

$$\Gamma_k := \sum_{i=1}^{n} \tilde{\alpha}_i \sin(\omega_i k + \tilde{\delta}_i)$$

where $\tilde{\alpha}_i$ are positive and can be chosen by the method given in section 3.3.2 and $\tilde{\delta}_i$ can be arbitrary or selected such that Γ_k is equalized in time, meaning that the peak of Γ_k is kept small.

The regressor vector ϕ is a wavelet generated based on the reference signal and its translation in time

$$\phi_k = \begin{bmatrix} \Gamma_k \\ \Gamma_{k-1} \\ \vdots \\ \Gamma_{k-2n+1} \end{bmatrix}. \tag{3.21}$$

The following lemma shows that this regressor defines a basis for the space of disturbances that are summation of sinusoids at ω_i's.

Lemma 1
For any signal in the form of $\bar{d}_k = \sum_{i=1}^{n} \alpha_i \sin(\omega_i k + \delta_i)$ *with arbitrary* α_i*'s and* δ_i*'s there exists a unique* $\theta \in \Re^{2n}$ *that satisfies*

$$\bar{d}_k = \theta^T \phi_k$$

where ϕ_k *is defined in* (3.21).

Proof
Let $F(z)$ *be an (unknown) FIR filter defined by* θ

$$F(z) = \theta^1 + \theta^2 z^{-1} + \cdots + \theta^{2n} z^{-2n+1}.$$

We need to show that

$$\begin{aligned} \bar{d}_k &= \theta^1 \Gamma_k + \theta^2 \Gamma_{k-1} + \cdots + \theta^{2n} \Gamma_{k-2n+1} \\ &= F[\Gamma_k] \\ &= \sum_{i=1}^{n} F\left[\tilde{\alpha}_i \sin(\omega_i k + \tilde{\delta}_i)\right]. \end{aligned} \tag{3.22}$$

This equation in frequency domain implies

$$\begin{aligned} \alpha_i e^{j\delta_i} &= F(e^{j\omega_i}) \tilde{\alpha}_i e^{j\tilde{\delta}_i} \Rightarrow \\ \frac{\alpha_i}{\tilde{\alpha}_i} e^{j(\delta_i - \tilde{\delta}_i)} &= F(e^{j\omega_i}) \\ &= \theta^1 + \theta^2 e^{-j\omega_i} + \theta^3 e^{-2j\omega_i} + \cdots + \theta^{2n} z^{-j(2n-1)\omega_i} \end{aligned}$$

for all $1 \le i \le n$. *Indeed, this is a system of linear equations*

$$\begin{bmatrix} \frac{\alpha_1}{\tilde{\alpha}_1} \cos(\tilde{\delta}_1 - \delta_1) \\ \frac{\alpha_1}{\tilde{\alpha}_1} \sin(\tilde{\delta}_1 - \delta_1) \\ \vdots \\ \frac{\alpha_n}{\tilde{\alpha}_n} \cos(\tilde{\delta}_n - \delta_n) \\ \frac{\alpha_n}{\tilde{\alpha}_n} \sin(\tilde{\delta}_n - \delta_n) \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & \cos(\omega_1) & \cdots & \cos((2n-1)\omega_1) \\ 0 & \sin(\omega_1) & \cdots & \sin((2n-1)\omega_1) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \cos(\omega_n) & \cdots & \cos((2n-1)\omega_n) \\ 0 & \sin(\omega_n) & \cdots & \sin((2n-1)\omega_n) \end{bmatrix}}_{A} \begin{bmatrix} \theta^1 \\ \theta^2 \\ \vdots \\ \theta^{2n-1} \\ \theta^{2n} \end{bmatrix}$$

and matrix A is full rank for $\omega_i \neq 0$ for all i's and $\omega_i \neq \omega_j$ for $i \neq j$. Therefore, for any arbitrary values of α_i, $\tilde{\alpha}_i > 0$, δ_i and $\tilde{\delta}_i$ a unique θ exists. □

This method requires less time complexity than the spectral method with trigonometric regressor (3.4) since both ϕ_k and ψ_k do not require any major calculation to be updated from step k to $k+1$. This is because the reference signal Γ_k is periodic and its profile in one period can be calculated offline and stored in a table for the adaptive controller. Finding ϕ_k will then be a simple table–look–up task. Similar idea holds for ψ_k by calculating and storing $R[\Gamma_k]$. However, this approach is computationally more expensive than the two time–domain based methods since no further simplification can be made. On the other hand, the algorithm does not have the disadvantages (D1) to (D4) that were common among the previous two time–domain methods. That is, the number of estimated parameters is twice the number of compensation frequencies and it is possible to apply compensation only at (any) selective frequencies. In conclusion, this method has good properties of the frequency–domain approach and its computational complexity is higher than time–domain approaches and less than frequency–domain method. The time complexity and selective frequency attenuation capability associated to the four considered learning kernels are given in Table 3.1.

Table 3.1: Time complexity of *one step update* and frequency separation capability of proposed methods. N dentoes the number of estimated parameters.

Method	Time complexity	Selective frequencies
Trigonometric functions §3.3	O(5N)	Yes
Periodic impulse train §3.4.1	O(2N)	No
Sparse filtered regressor §3.4.2	O(N)	No
Wavelet §3.4.3	O(3N)	Yes

3.5 Summary and Conclusion

We briefly reviewed control methodologies for coping with periodic signals, and compared the feedback control and feedforward control methods. It was argued that for the class of problems under our study the latter methods are superior. We deployed a stochastic gradient descent method, to develop an adaptive feedforward control algorithm for compensating multitude narrow–band disturbances or trajectory tracking. Algorithm 1 summarizes the high–level outline of the proposed method. The method is also illustrated as a block diagram in Fig. 3.6.

Using the averaging theory, we derived a condition on the adaptation step size to guarantee the algorithm convergence and perfect compensation. We also proposed a novel adaptive step size and integrated it to the adaptive control algorithm to enhance the convergence rate and decrease the steady state error.

The analysis were initially carried out in a spectral framework where trigonometric functions form an orthogonal basis for the class of deterministic disturbances/trajectories under our study. Two alternatives to this approach, namely decomposition of the disturbance in time–domain and in time–frequency–domain, were also discussed. It was shown that time–domain approaches can be computationally more plausible than frequency–domain methods. However, the latter methods can be more robust to system dynamics uncertainty and provide the ability of targeting selective frequencies.

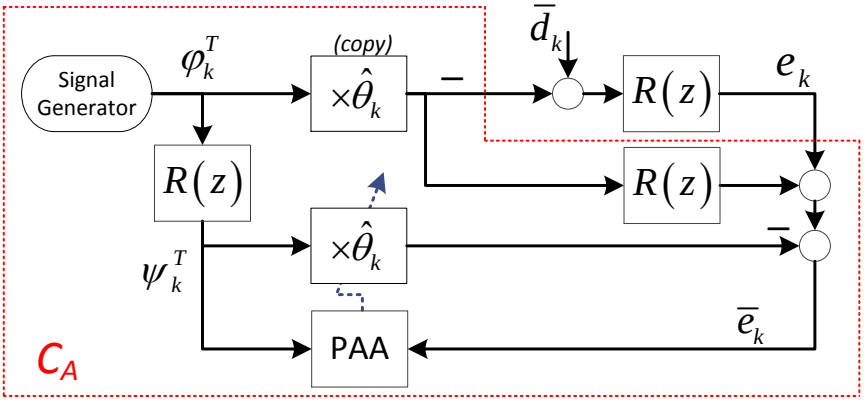


Figure 3.6: Adaptive feedforward repetitive controller block diagram.

Our work in this chapter did not include two important aspects: (1) the algorithm robustness to dynamic uncertainties when an exact model of the actual system is not available was not analyzed, (2) the algorithm was not evaluated experimentally. The next two chapters address these two topics.

Algorithm 1 Adaptive Feedforward Repetitive Control Algorithm

procedure AdaptiveFeedForward(n,V^d,V^{ub},ρ,$\mu^{\max}$,f_γ)
  Initialize( )
  while (1) **do**
    Apply $-u_{A,k}$ to the system ▷ e.g. update the DAC
    Read the error e_k ▷ e.g. read the ADC buffer
    $\bar{e}_k \leftarrow e_k + \tilde{e}_k$ ▷ update auxiliary error
    $\mu_k \leftarrow$ UpdateStepSize($\bar{e}_k$) ▷ update step size
    for $i = 1 : n$ **do** ▷ update estimates
      $\gamma_k^i \leftarrow f_\gamma(k, i)$ ▷ $f_\gamma(.,.)$ schedules adaptations of estimates
      $\hat{\theta}_{k+1}^i \leftarrow \hat{\theta}_k^i + \gamma_k^i \mu_k \psi_k^i \bar{e}_k$
      $\hat{\theta}_{k+1}^{i+n} \leftarrow \hat{\theta}_k^{i+n} + \gamma_k^i \mu_k \psi_k^{i+n} \bar{e}_k$
    end for
    $u_{A,k+1} \leftarrow \phi_{k+1}^T \hat{\theta}_{k+1}$ ▷ control for the next step
    $\tilde{e}_{k+1} \leftarrow R\left[u_{A,k+1}\right] - \psi_{k+1}^T \hat{\theta}_{k+1}$
    $k \leftarrow k + 1$
  end while
end procedure

procedure UpdateStepSize($\bar{e}_k$) ▷ update step size based on current aux. error
  $V_k^h = V_{k-1}^h + \frac{1}{h}\bar{e}_k^2 - \frac{1}{h}\bar{e}_{k-h}^2$ ▷ moving average over squared error
  $\bar{\mu}_k = \rho\left(V_k^h - V^d\right)$ ▷ scaled distance of current variance from desired variance
  if $(\bar{\mu}_k > 0$ && $\mu_{k-1} > 0)$ || $(V_k^h > V^{ub})$ **then** ▷ the adaptation was active in
    ▷ the last step and the variance is still unsatisfactory; or, the variance is above the dead–band
    $\mu_k = \min([\bar{\mu}_k, \mu_{\max}])$
  else
    $\mu_k = 0$
  end if
  return μ_k
end procedure

procedure Initialize( ) ▷ initialize parameters
  $k \leftarrow 0,\ u_{A,k} \leftarrow 0,\quad \tilde{e}_k \leftarrow 0$
  for $i = -h : -1$ **do**
    $V_i^h \leftarrow 0,\quad \bar{e}_i \leftarrow 0$
  end for
  for $i = 1 : 2n$ **do**
    $\hat{\theta}_k^i \leftarrow 0$
  end for
end procedure

Chapter 4

Indirect Adaptive Feedforward Repetitive Control for Systems with Unknown Dynamics

4.1 Uncertain System Dynamics

The parameter adaptation algorithm proposed in section 3.3.1 uses the closed loop system transfer function, $R(z)$, to construct the auxiliary error based on (3.7). In general, an exact dynamics of the system is not available in practical applications, especially at high frequencies. In addition to unmodeled dynamics, uncertainties can be caused by temperature variations and deterioration over time. For instance, temperature variation in an HDD causes voltage gain (i.e., output position versus voltage) alteration in the piezoelectric actuator Malang and Hutsell (2005). A classical approach to compensate for system dynamics alteration at different operation conditions is to consider a nominal model and provide a set of calibration coefficients to modify the nominal model accordingly Aphale et al. (2008). However, this type of offline calibration is applicable to correct plant variations only at low frequencies since dynamic variations at high frequencies are not similar among different instances of a system. In Fig. 4.1 an example of dynamic uncertainties and alterations over different temperatures and plants in an HDD are shown. The figure depicts the frequency response of the closed loop system from the micro–actuator inputs inside one HDD to the position error signal. The frequency response data is measured at different temperatures from 34°_{C} to 48°_{C} on 8 micro–actuators of a 4–platter hard disk drive. As can be seen from the figure, the phase response fluctuation at a high frequency, such as 18KHz, can even reach 180 degrees.

When the actual closed loop dynamics $R(z)$ is not available, the adaptive feedforward

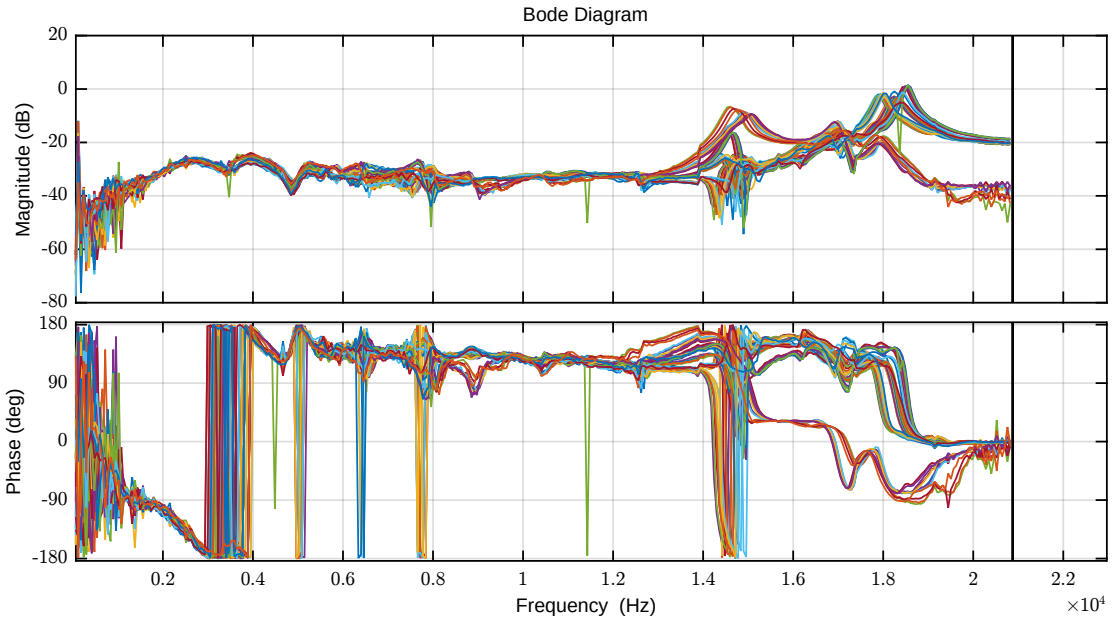


Figure 4.1: Secondary path responses from inputs of 8 micro–actuators to the position error signal in one HDD. Data is measured at temperatures varying from $34°C$ to $48°C$.

controller proposed in section 3.3 deploys an approximate model denoted by $\bar{R}(z)$

$$\begin{aligned} \bar{\psi}_k &:= \bar{R}[\phi_k] \\ \bar{e}_k &:= e_k + \bar{R}\left[u_{A,k}\right] - \bar{\psi}_k^T \hat{\theta}_k \\ \hat{\theta}_{k+1} &= \hat{\theta}_k + \gamma_k \mu_k \bar{\psi}_k \bar{e}_k. \end{aligned} \tag{4.1}$$

Similarly, the update rule of a–posteriori version of the algorithm given in (3.13) can be modified by replacing all occurrences of ψ_k by $\bar{\psi}_k$. The block diagram of the adaptive feedforward controller when a model is used instead of the actual system is shown in Fig. 4.2.

The robustness of the adaptive feedforward controller to the mismatch between the actual plant and its model is outlined in the following theorem.

Theorem 3 (Controller's robustness to unmodeled dynamics) *Let* $\bar{m}_i = |\bar{R}\left(e^{j\omega_i T}\right)|$ *and* $\bar{\delta}_i = \angle \bar{R}\left(e^{j\omega_i T}\right)$ *be the magnitude and phase response of* $\bar{R}(z)$ *at* ω_i. *Note that these variables are analogous to* m_i *and* δ_i *defined in* (3.14) *for* $R(z)$. *Assume that* α_i*'s in* ϕ *are chosen according to* $\alpha_i = 1/\bar{m}_i$. *The estimated parameters in the adaptive feedforward controller that deploys a model of closed loop system,* $\bar{R}(z)$, *converge to the actual parameters (in mean value) if*

$$0 < \mu_{\min} < \mu_{\max} \leq \min_{i \in \{1,\cdots,n\}} 4\cos(\delta_i - \bar{\delta}_i)\frac{\bar{m}_i}{m_i}. \tag{4.2}$$

for some $\mu_{\min}$ *that satisfies* $0 < \mu_{\min} < \mu_{\max}$.

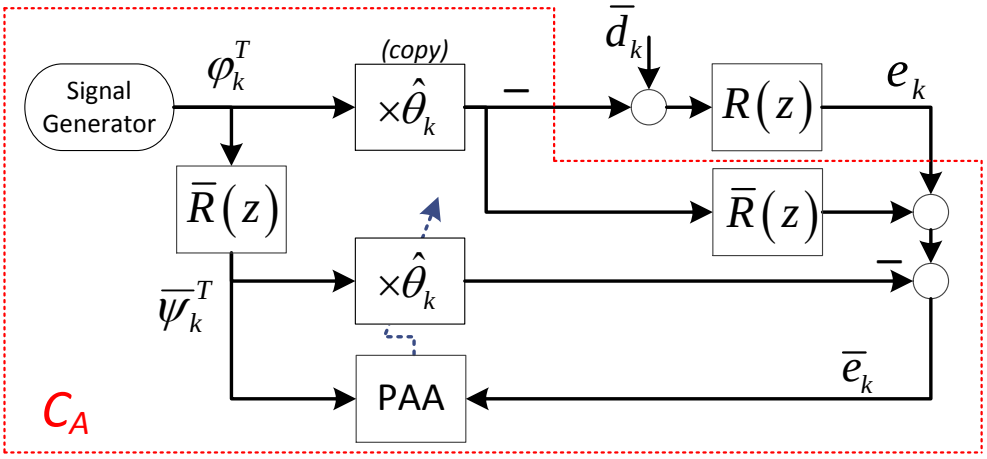


Figure 4.2: Adaptive feedforward repetitive controller block diagram for systems with uncertain dynamics.

Proof

The proof is again based on the results of theorem 2.2.1 in Bai et al. (1988). Let $x(k) := \theta - \mathbf{E}[\hat{\theta}_k]$ *be a state for the parameter adaptation algorithm. We have*

$$\begin{aligned} x(k+1) - x(k) &= -\mathbf{E}\left[\hat{\theta}_{k+1} - \hat{\theta}_k\right] \\ &= -\mathbf{E}\left[\mu_k \bar{\psi}_k \bar{e}_k\right] \\ &= -\mu_k \bar{\psi}_k \mathbf{E}\left\{R\left[\phi_k^T\theta - \phi_k^T\hat{\theta}_k\right] + \bar{R}\left[\phi_k^T\hat{\theta}_k\right] - \bar{R}\left[\phi_k^T\right]\hat{\theta}_k + \xi_k\right\} \\ &= -\mu_k \bar{\psi}_k \left\{R\left[\phi_k^T x(k)\right] + \bar{R}\left[\phi_k^T x(k)\right] - \bar{R}\left[\phi_k^T\right] x(k)\right\} \end{aligned} \tag{4.3}$$

For a given $\epsilon > 0$*, define* $\mu_k := \tilde{\mu}_k \epsilon$*. We have*

$$x(k+1) = x(k) + \epsilon f(k, x(k), \epsilon)$$

where

$$f(k, x(k), \epsilon) := -\tilde{\mu}_k \bar{\psi}_k \left\{R\left[\phi_k^T x(k)\right] + \bar{R}\left[\phi_k^T x(k)\right] - \bar{R}\left[\phi_k^T\right] x(k)\right\}.$$

One can show that this function satisfies all assumptions `A1-A4` *in Bai et al. (1988). The average system for this difference equation is defined as*

$$x_{av}(k+1) = x_{av}(k) + \epsilon f_{av}(x(k))$$

where

$$\begin{aligned} f_{av}(x) :&= \lim_{T\to\infty} \frac{1}{T} \sum_{k=k_0+1}^{k_0+T} f(x,k,0) \\ &= \lim_{T\to\infty} \frac{1}{T} \sum_{k=k_0+1}^{k_0+T} -\tilde{\mu}_k \bar{\psi}_k \left\{ R\left[\phi_k^T x\right] + \bar{R}\left[\phi_k^T x\right] - \bar{R}\left[\phi_k^T\right] x \right\} \\ &= \left(\lim_{T\to\infty} \frac{1}{T} \sum_{k=k_0+1}^{k_0+T} -\tilde{\mu}_k \bar{\psi}_k \psi_k^T \right) x \end{aligned}$$

We have

$$\begin{aligned} &\lim_{T\to\infty} \frac{1}{T} \sum_{k=k_0+1}^{k_0+T} \begin{bmatrix} \bar{\psi}_k^i \\ \bar{\psi}_k^{i+n} \end{bmatrix} \begin{bmatrix} \psi_k^i \\ \psi_k^{i+n} \end{bmatrix}^T \\ &= \lim_{T\to\infty} \frac{1}{T} \sum_{k=k_0+1}^{k_0+T} \frac{m_i}{\bar{m}_i} \begin{bmatrix} \sin(\omega_i kT+\delta_i)\sin(\omega_i kT+\bar{\delta}_i) & \sin(\omega_i kT+\delta_i)\cos(\omega_i kT+\bar{\delta}_i) \\ \cos(\omega_i kT+\delta_i)\sin(\omega_i kT+\bar{\delta}_i) & \cos(\omega_i kT+\delta_i)\cos(\omega_i kT+\bar{\delta}_i) \end{bmatrix}^T \end{aligned}$$

$$= \frac{1}{2}\frac{m_i}{\bar{m}_i} \begin{bmatrix} \cos(\delta_i-\bar{\delta}_i) & \sin(\delta_i-\bar{\delta}_i) \\ -\sin(\delta_i-\bar{\delta}_i) & \cos(\delta_i-\bar{\delta}_i) \end{bmatrix}$$

for all values of $i = 1, \cdots, n$. *This matrix has eigenvalues at*

$$\frac{1}{2}\frac{m_i}{\bar{m}_i}(\cos(\delta_i-\bar{\delta}_i) \pm j\sin(\delta_i-\bar{\delta}_i)).$$

Hence, the averaged system

$$x_{av}(k+1) = \left(I - \lim_{T\to\infty} \frac{1}{T} \sum_{k=k_0+1}^{k_0+T} \mu_k \bar{\psi}_k \psi_k^T \right) x_{av}(k)$$

has eigenvalues at $1 - \frac{\mu_{av} m_i}{2\bar{m}_i}\left(\cos(\delta_i-\bar{\delta}_i) \pm j\sin(\delta_i-\bar{\delta}_i)\right)$ *where* $\mu_{av} := \lim_{T\to\infty} \frac{1}{T} \sum_{k=k0+1}^{k_0+T} \mu_k$. *The averaged system is exponentially stable if*

$$\begin{aligned} 1 &> \left(1 - \frac{\mu_{av} m_i}{2\bar{m}_i}(\cos(\delta_i-\bar{\delta}_i)\right)^2 + \left(\frac{\mu_{av} m_i}{2\bar{m}_i}\sin(\delta_i-\bar{\delta}_i))\right)^2 \\ &= 1 + \left(\frac{\mu_{av} m_i}{2\bar{m}_i}\right)^2 - \frac{\mu_{av} m_i}{\bar{m}_i}(\cos(\delta_i-\bar{\delta}_i)) \end{aligned}$$

which is satisfied when

$$\mu_{\max} \leq \min_{i\in\{1,\cdots,n\}} 4\cos(\delta_i-\bar{\delta}_i)\frac{\bar{m}_i}{m_i}$$

□

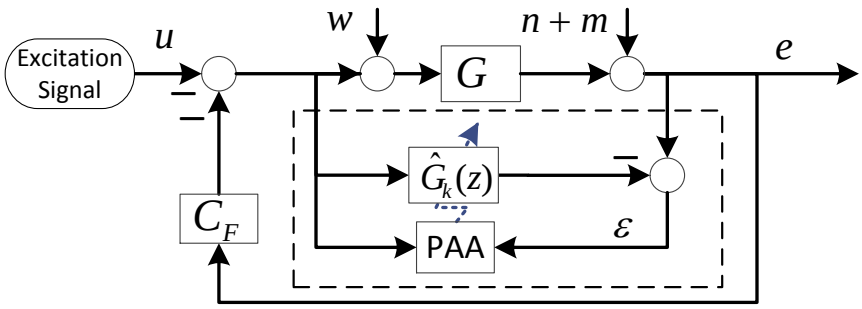


Figure 4.3: On–line secondary path identification architecture

In many practical applications an upper bound on the magnitude of uncertain systems can be determined. In such cases, a corollary of the above theorem is that when the phase mismatch between the actual system and its model is less than 90 degrees

$$\max_{i\in\{1,\cdots,n\}} |\angle R(e^{j\omega_i}) - \angle \bar{R}(e^{j\omega_i})| < \frac{\pi}{2}, \tag{4.4}$$

the maximum step size $\mu_{\max}$ can be chosen small enough such that the convergence criterion is satisfied.

4.2 Online Secondary Path Modeling

This section proposes an adaptive secondary path modeling architecture to stabilize the adaptive repetitive controller in case the phase mismatch between $R(z)$ and $\bar{R}(z)$ does not satisfy the aforementioned criterion. Unlike the previous section that we ignored the internal structure of the closed loop dynamics $R(z)$, we use our knowledge about the nominal feedback controller and the internal interconnections to identify the uncertain plant. This is an important distinction between this architecture and previous work Eriksson and Allie (1989); Akhtar et al. (2006), in which, modeling of the closed loop system $R\,(z)$ is studied. The order of the closed loop dynamics is equal to the summation of plant and controller orders when no pole–zero–cancellation occurs. Hence, the number of parameters required to identify the plant is less than the closed loop system in general.

We first consider a simple system identification case that is depicted in Fig. 4.3. An important note to make here is that the periodic disturbance $\bar{d}$ is not considered in this framework. Let $G(q^{-1})$ be a finite dimension transfer function of G represented by the *one step delay operator* q^{-1}

$$G\left(q^{-1}\right) = \frac{B^g\left(q^{-1}\right)}{A^g\left(q^{-1}\right)} = \frac{b_0^g + b_1^g q^{-1} + \cdots + b_{n_g}^g q^{-n_g}}{1 + a_1^g q^{-1} + \cdots + a_{n_g}^g q^{-n_g}}.$$

In the figure, $\hat{G}_k\,(z)$ is the estimated plant model at time step k and **PAA** denotes the parameter adaptation algorithm. A great deal of research effort has been focused on the

design of **PAA** based on the characteristics of input and output noises Regalia (1994); Ljung and Söderström (1983). The design of a parameter adaptation algorithm for system identification problem shown in Fig. 4.3 is beyond the scope of this work. Here, we assume that based on the prior knowledge about noises w, n and m, a proper recursive system identification algorithm is chosen such that it satisfies the following assumption.

Assumption `A.1`: *The model parameters estimated by the parameter adaptation algorithm in Fig. 4.3 converge asymptotically and the order for the model is large enough such that there exists a bounded integer $P < \infty$ and a positive real value $\bar{\mu}_{\max}$ such that the closed loop system*

$$\bar{R}_k(z) = \frac{\hat{G}_k(z)}{1 + C_F(z)\,\hat{G}_k(z)}. \tag{4.5}$$

satisfies

$$\bar{\mu}_{\max} < \min_{i=1,\cdots,n} \frac{\bar{m}_i}{m_i} \cos(\delta_i - \bar{\delta}_i) \qquad \forall k \geq P. \tag{4.6}$$

Note that we are only interested in acquiring a system model that fits to the actual system at compensation frequencies. This requirement is considerably less stringent than obtaining a model that describes the system at a broad frequency interval. In section 4.3, we propose a class of excitation signals that is tailored specifically for this purpose and has a very low computational cost since it takes advantage of calculations common with the adaptive controller.

Based on assumption `A.1` we advance to the problem of secondary path modeling for the adaptive feedforward controller. Figure 4.4 illustrates the proposed architecture that uses the same **PAA** as Fig. 4.3. As shown in the figure, the secondary path of the adaptive feedforward controller is now using the estimated model given in (4.5) and its parameter adaptation algorithm is now denoted by "LMS" to avoid ambiguity with the system identification PAA.

Lemma 2
Condition (4.6) is satisfied for the system shown in Fig. 4.4 if assumption `A.1` holds for the system shown in Fig. 4.3. In other words, the presence of adaptive feedforward controller and periodic disturbance $\bar{d}$ in the loop do not change the steady state behavior of the system identification mechanism.

Proof
Since the adaptive control is known, we can think of the summation of the feedback control, adaptive control, and the injected noise, as a single excitation signal that is known to us. Therefore, if we ignore filters $1-z^{-N}$ and signal $\bar{d}$, the estimated plant parameters in Fig. 4.4 converge if and only if they converge when there is no adaptive repetitive controller in the

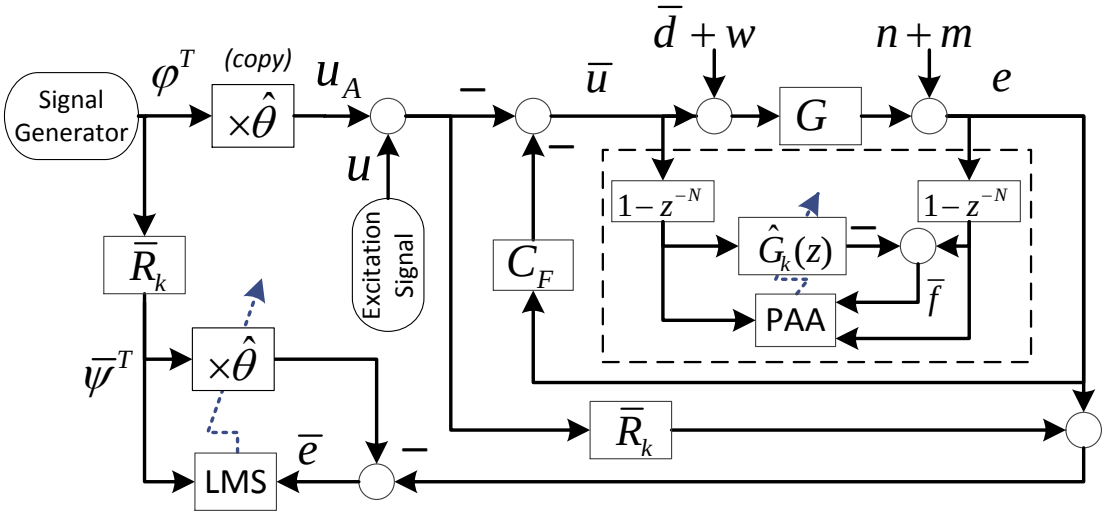


Figure 4.4: Adaptive repetitive controller with secondary path modeling (dashed box).

loop. However, the periodic disturbance $\bar{d}_k$ can cause biased parameters convergence, and it is required to be filtered out to validate this statement. The error signal in Fig. 4.4 is

$$e_k = \theta_g^T \phi_{g,k} + d_{g,k} + \xi_{g,k}$$

where

$$\theta_g^T := \left[a_1^g, \cdots, a_{n_g}^g, b_0^g, \cdots, b_{n_g}^g\right] \tag{4.7}$$

$$\phi_{g,k}^T := \left[e_{k-1}, \cdots, e_{k-n_g}, \bar{u}_k, \cdots, \bar{u}_{k-n_g}\right] \tag{4.8}$$

$$d_{g,k} := G\left[\bar{d}_k\right] \tag{4.9}$$

$$\xi_{g,k} := G\left[w_k\right] + n_k + m_k. \tag{4.10}$$

Note that $d_{g,k}$ is N–periodic in steady state since $\bar{d}_k$ is periodic and $G\left(q^{-1}\right)$ is an LTI system. Since we are interested in the steady state behavior, we ignore the transient state and assume that $d_{g,k}$ is periodic. This assumption leads us to filter both e_k and $\phi_{g,k}$ through $1-q^{-N}$ and define

$$\begin{aligned} \bar{f}_k :&= (e_k - e_{k-N}) - \hat{\theta}_{g,k}\left(\phi_{g,k} - \phi_{g,k-N}\right) \\ &= \tilde{\theta}_{g,k}^T \tilde{\phi}_{g,k} + \tilde{\xi}_{g,k} \end{aligned} \tag{4.11}$$

where $\tilde{\phi}_{g,k} = \phi_{g,k} - \phi_{g,k-N}$, $\tilde{\theta}_{g,k} = \theta_g - \hat{\theta}_{g,k}$ and $\tilde{\xi}_{g,k} = \xi_{g,k} - \xi_{g,k-N}$. Note that (4.11) is analogous to the estimation error dynamics in Fig. 4.3

$$\epsilon_k = \tilde{\theta}_{g,k}^T \phi_{g,k} + \xi_{g,k}.$$

In other words, the two filters $(1-z^{-N})$ transform the inputs of PAA such that it does not see the effect of $\bar{d}$. It is worth noting that this operation causes $\tilde{\xi}_g$ to have a higher variance than ξ_g. This may require a larger excitation signal to achieve the same accuracy level. □

Based on the separation property shown in the previous lemma and assumption A.1, by choosing $\mu_{\max} < 4\bar{\mu}_{\max}$ the convergence criterion (4.2) for the adaptive repetitive controller will be satisfied after a finite number of steps. On the other hand, the excitation signal u, is statistically independent of the regressor signal of the adaptive repetitive controller, and it cannot cause unbiased parameter estimation in the repetitive controller. This can be formally consolidated by revisiting the proof of theorem 3: 1)equation 4.3 for the current architecture becomes

$$\begin{aligned} x(k+1) - x(k) = -\mu_k \bar{\psi}_k \left\{ R\left[\phi_k^T x(k)\right] + \bar{R}\left[\phi_k^T x(k)\right] - \bar{R}\left[\phi_k^T\right] x(k) \right\} \\ - \mu_k \bar{\psi}_k \left\{ -R\left[u_k\right] + \bar{R}\left[u_k\right] \right\} \end{aligned}$$

2) the second term does not contribute in $f_{av}(x)$ since

$$\lim_{T\to\infty} \frac{1}{T} \sum_{k=k_0+1}^{k_0+T} \bar{\psi}_k \left(-R\left[u_k\right] + \bar{R}\left[u_k\right] \right) = 0$$

when $\bar{\psi}$ and u are statistically independent. Moreover, for the class of excitation signals that will be proposed in section 4.3 u is consisted of a set of sinusoidals that differ from the contents of $\bar{\psi}$ in frequency. Therefore, the above equality holds for this type of deterministic excitation signals too.

An important point to make is that, the adaptive control path does not converge as long as the updated transfer function $\bar{R}(z)$ has more than 90 degrees phase error relative to the actual transfer function $R(z)$. As a result when the initial parameters of $\hat{G}_0$ are not accurate, we expect that the adaptive control parameters diverge quickly. Although the adaptive controller becomes stable eventually — once the plant parameters get close enough (in terms of phase error) — this behavior is not desirable in many applications since the transient error may be very large, and accordingly the adaptive control requires a long time to recover. We thus suggest using an *initialization* period prior to the simultaneous adaptation, in which the adaptive controller is inactive till the secondary path modeling parameters converge.

On the other hand, in some practical applications the plant dynamics might be fairly known at a frequency range, say Ω. If such information is available, the frequency components of disturbance can be categorized into two sets $I_1 = \{i|\omega_i \in \Omega\}$ and $I_2 = \{i|\omega_i \notin \Omega\}$. The adaptation of parameters $(\hat{\theta}^i,\ \hat{\theta}^{i+n}), i \in I_1$ and plant parameters can be done simultaneously since the plant model is exact at that region. This can be done by choosing

$$\gamma_k^i = \begin{cases} 0 & i \in I_2 \text{ and } k < P \\ 1 & \text{otherwise} \end{cases} \tag{4.12}$$

where P denotes the length of initialization period required for secondary path modeling.

4.3 Exogenous Excitation Signal

It is well known that in an adaptive system identification algorithm the convergence towards zero of the prediction error $\epsilon(k)$ (c.f. Fig. 4.3) does not necessarily imply that the estimated parameters converge towards the true parameters Landau et al. (2006). In order to identify the actual parameters, it is necessary to apply a "frequency rich" excitation signal u. The standard solution in practice is usually provided by the use of *pseudo–random binary sequences* that have enough persistence of excitation for identification of any number of parameters Bitmead (1984). This approach is beneficial when the identification should be performed in a wide frequency range.

We propose a novel method that requires less computation and it is more effective for our problem since we need to identify the system dynamics only around the narrow band disturbances. The fact that the adaptive algorithms can operate with extremely weak excitation signals is a very much appreciated quality in practical situations. Our key idea is that we focus the excitation signal around the narrow band disturbances so that a low order identified system dynamics fits the complex high order system only around the disturbances. This is done by an effective use of sinusoidals that are distributed symmetrically around compensation frequencies (ω_i)

$$u_k = \frac{\alpha_k^u}{2} \sum_{i=1}^{n} \sin\left((\omega_i + \delta_u)kT\right) + \sin\left((\omega_i - \delta_u)kT\right) \tag{4.13}$$

which implies that the whole excitation energy is focused sharply around where identification should be performed. Here, δ_u is a small frequency shift such as a few percent of the smallest ω_i, and α_k^u is a fixed or variable positive gain that should be defined based on the noise level, desired convergence rate and unwanted transient error. An adaptive version of this gain is discussed in section 6.3.2.

Surprisingly, although this excitation signal contains many sinusoidals, it can be constructed with negligible computation when combined with the proposed adaptive feedforward controller. This is because the expression in (4.13) can be effectively evaluated by taking advantage of the regressor ϕ_k

$$\begin{aligned}
u_k &= \frac{\alpha_k^u}{2} \sum_{i=1}^{n} \sin\left((\omega_i + \delta_u)kT\right) + \sin\left((\omega_i - \delta_u)kT\right) \\
&= \frac{\alpha_k^u}{2} \sum_{i=1}^{n} \sin(\omega_i kT)\cos(\delta_u kT) + \cos(\omega_i kT)\sin(\delta_u kT) \\
&\qquad\qquad + \sin(\omega_i kT)\cos(\delta_u kT) - \cos(\omega_i kT)\sin(\delta_u kT) \\
&= \alpha_k^u \cos(\delta_u kT) \sum_{i=1}^{n} \sin(\omega_i kT).
\end{aligned}$$

Note that all trigonometric functions in front of the sum operator have been already evaluated for ϕ_k. Thus, even though the excitations signal is consisted of $n+1$ trigonometric functions,

only one function evaluation is needed and the rest is in hand once the regressor ϕ_k is calculated.

4.4 Band–Pass Filters for SNR Enhancement in System Identification

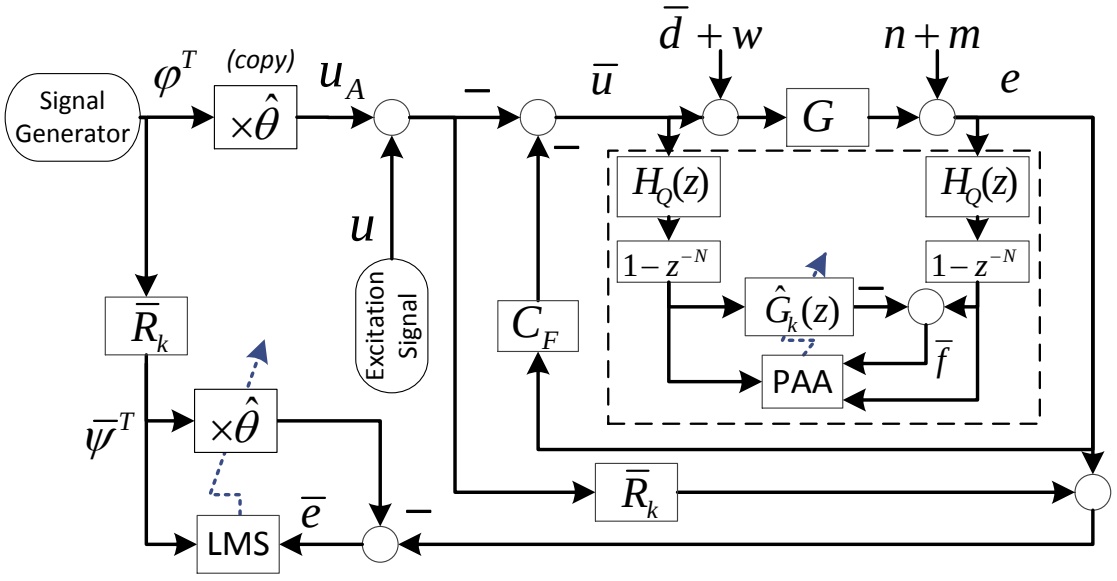


Figure 4.5: Adaptive repetitive controller with secondary path modeling and band–pass filters (dashed box)

As was mentioned in section 4.2 our expectation from the system identification PAA is less stringent than a general identification problem since we only need to satisfy criterion (4.6) at compensation frequencies. The excitation signal proposed in section 4.3 attempts to facilitate the identification by focusing energy around the compensation frequencies. Another method that can further lead to a more accurate identification with a low excitation signal is to use a pair of band–pass filters in order to decrease the effect of noises (w,n,m) through attenuating the spectrum of input/output signals at other frequencies. This idea is illustrated in Fig. 4.5 where the two filters are shown by $H_Q(z)$. For instance, if the adaptive compensator in an HDD is set to compensate harmonics 120 to 150, we would like to minimize the effect of all noises at other frequencies. Therefore, the passband of filter should be adjusted to [120, 150] (in terms of harmonic).

The bandpass filters can be designed offline if the set of compensation frequencies are known a–priori. The binary scheduling parameters, γ_k^i's, that were introduced in 3.3.4 provide a flexible framework for scheduling the adaptation of different frequencies in a practical application. When these parameters change in time, a mean is needed to adjust the bandpass

filters accordingly. An example of such a case was mentioned in section 4.2. Changing the set of compensation frequencies is also beneficial when the computational power is limited, or when the model order is not large enough to satisfy condition (4.6) and the frequency range should be decreased. We will be returning to this matter at various junctures, especially in section 6.4.2 where we discuss an automatic mechanism for determining the compensation frequencies. As a result, we pursue an online filter design approach.

Typical digital low–pass IIR filters such as Butterworth, Chebyshev and Elliptic filters, or FIR filters can be designed directly with relatively inexpensive algorithms thanks to the closed form solutions (as in Oppenheim et al. (1989)) for their pole and zero locations. However, direct design of digital band–pass or band–stop filters is more tedious especially in narrow–band cases as in our application. This makes it impossible to design a band–pass filter on a simple embedded system in real time[1]. *Digital Frequency Transformation* is an alternative to direct methods for designing band–pass (also band–stop and high–pass) filters that was initially proposed in a simple form for real filters in Constantinides (1970). This methodology was later extended to more general cases of real Nowrouzian and Constantinides (1990), multi–band Feyh et al. (1986) and real/complex multi–band filters Krukowski et al. (1995). This type of transformation are often performed using a fixed low–pass prototype filter and an all–pass frequency transformation. The drawback of indirect method is that the resulting filter may have a larger dimensionality (order) compared to the filters that can be designed using direct methods with optimization. However, the extra required computation caused by only having a few more parameters in the filter is negligible compared to the tedious task of designing filters optimally.

The general idea of the frequency transformation is to take an existing prototype filter and produce another filter from it that retains some of the characteristics of the prototype, in the frequency domain. Transformation functions achieve this by replacing each delaying element, z^{-1}, of the prototype filter with an allpass filter carefully designed to have a prescribed phase characteristic for achieving the modifications requested by the designer.

The basic form of transformation from a given prototype filter $H_0(z)$ to a desired filter $H_Q(z)$ is commonly performed by using a mapping $\bar{z} = Q(z)$ Oppenheim et al. (1989)

$$\begin{aligned} H_Q(z) &= H_0\left(\bar{z}\right) \\ &= H_0\left(Q\left(z\right)\right) \end{aligned} \tag{4.14}$$

where, generally, $Q(z)$ is a first or second order all–pass filter. In our particular application, we are interested in transforming a low–pass filter to a band–pass filter with desired cutoff edges, say w_1 and w_2. Therefore, we can think of this mapping as a transformation of two features, namely the cutoff frequency of low pass filter $\pm w_0$ to w_1 and w_2. Usually a second–order mapping in the form of

$$\bar{z} = Q(z) = \pm \frac{\frac{1}{\alpha_2} z^2 + \frac{\alpha_1}{\alpha_2} z + 1}{z^2 + \frac{\alpha_1}{\alpha_2} z + \frac{1}{\alpha_2}}.$$

[1]Here, we refer to systems with <500MHz processors and real–time applications with $<25\mu s$ sampling times.

is used since the coefficients α_1 and α_2 give enough degrees of freedom to migrate any two frequency domain features of the original transfer function $H_0(z)$ to any other (meaningful) frequency locations while preserving the filter stability. Negative sign is used when a feature from zero frequency should be translated, a case that is usually called "DC–mobility", and conversely a positive sign is used when a feature at the Nyquist frequency should be relocated.

Theorem 4
Given any asymptotically stable low–pass filter $H_0(z)$ with normalized cut–off frequency w_0,

$$H_Q(z) = H_0\left(-\frac{\frac{1}{\alpha_2}z^2 + \frac{\alpha_1}{\alpha_2}z + 1}{z^2 + \frac{\alpha_1}{\alpha_2}z + \frac{1}{\alpha_2}}\right) \tag{4.15}$$

is an asymptotically stable bandpass filter with (normalized) passband w_1 to w_2 if: (1) $0 \leq w_1 < w_2 < \pi$ and (2) the two parameters of mapping are set by

$$\begin{aligned} \alpha_1 &= \frac{2ab}{1-b} & \alpha_2 &= \frac{1+b}{b-1} \\ a &:= \frac{\cos\left(\frac{w_2+w_1}{2}\right)}{\cos\left(\frac{w_2-w_1}{2}\right)} & b &:= \frac{\tan\left(\frac{w_0}{2}\right)}{\tan\left(\frac{w_2-w_1}{2}\right)} \end{aligned} \tag{4.16}$$

Proof
See Nowrouzian and Constantinides (1990) for the proof.

Equation (4.15) along with (4.16) show that the coefficients of the desired band–pass filter $H_Q(z)$ are polynomials in a and b which is the most important advantage of using frequency transformation for implementation. More explicitly, for a given fixed even order $2n$, the desired filter $H_Q(z)$ has coefficients that are polynomials of (at most) order $2n$ (c.f. (4.17-4.18)) with coefficients that are functions of the prototype filter coefficients. Therefore, a fixed prototype lowpass filter can be designed first, and then, for each coefficient of $H_Q(z)$ an explicit polynomial expression can be obtained. Note that this whole procedure can be performed offline. Once it comes to designing the bandpass filter for a given passband (w_1, w_2) in an online fashion, it is only needed to calculate a and b given in (4.16) and then evaluate $2n$ polynomials of $2n^{\text{th}}$ order. An example for this procedure is given in the following.

Elliptic low pass filters are well known for their fast transition between passband and stopband. Since the filters on input and output signals are identical, the distortion caused by the phase delay of this pair of filters does not have any negative effect in the identification. In our experiments, we used a third order Elliptic lowpass prototype filter. The prototype cutoff frequency is not important and we set it to harmonic 100 of the HDD spinning frequency (i.e. 1.82 rad/s)

$$H_0(z) = \frac{0.2184z^3 + 0.5891z^2 + 0.5891z + 0.2184}{z^3 + 0.1832z^2 + 0.5611z - 0.1291}.$$

The magnitude frequency response of this filter is shown in Fig. 4.6.

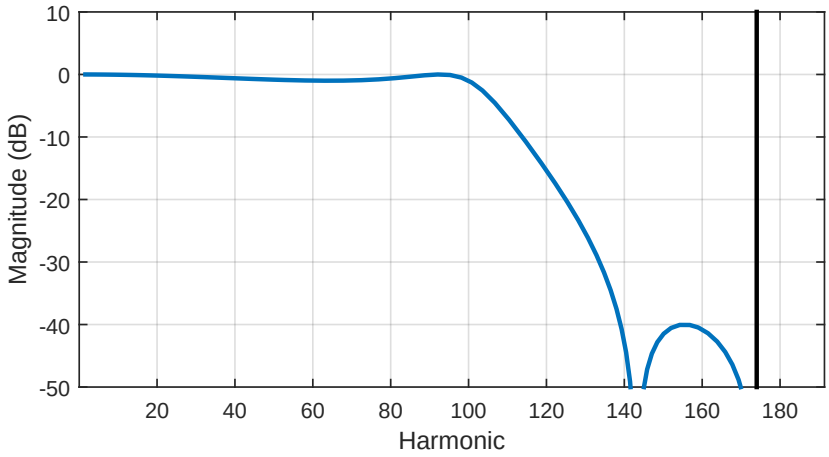


Figure 4.6: Frequency response of the Elliptic lowpass filter used as the prototype filter in our implementation.

Using the mapping given in (4.15)-(4.16), the bandpass filter is of the form

$$H_Q(z) = \frac{b_6 z^6 + b_5 z^5 + b_4 z^4 + b_3 z^3 + b_2 z^2 + b_1 z^1 + b_0}{z^6 + a_5 z^5 + a_4 z^4 + a_3 z^3 + a_2 z^2 + a_1 z^1 + a_0} \tag{4.17}$$

where

$$\begin{aligned}
b_6 &= 0.088b^2 + 1.1 & a_6 &= 1.0b^3 + 1.2b^2 + 2.0b + 1.1 \\
b_5 &= -0.35ab^2 & a_5 &= -6.0ab^3 - 5.0ab^2 - 4.0ab \\
b_4 &= 0.35a^2b^2 + 0.088b^2 - 3.2 & a_4 &= 12.0a^2b^3 + 5.0a^2b^2 + 3.0b^3 + 1.2b^2 - 2.0b - 3.2 \\
b_3 &= 0 & a_3 &= -8.0a^3b^3 - 12.0ab^3 + 8.0ab \\
b_2 &= -b_4 & a_2 &= 12.0a^2b^3 - 5.0a^2b^2 + 3.0b^3 - 1.2b^2 - 2.0b + 3.2 \\
b_1 &= -b_5 & a_1 &= -6.0ab^3 + 5.0ab^2 - 4.0ab \\
b_0 &= -b_6 & a_0 &= 1.0b^3 - 1.2b^2 + 2.0b - 1.1 \quad .
\end{aligned} \tag{4.18}$$

Suppose this filter is desired to be transformed to a band–pass filter with pass–band $[120,\ 140]$ in terms of harmonics. The parameters a and b can be easily evaluated for these w_1 and w_2 (after normalization) based on (4.16) and then all parameters of the band–pass filter can be calculated based on (4.18). The magnitude response of the filter designed by this method is depicted in Fig. 4.7.

4.5 Linear Periodically Time–Varying Realization of the Adaptive Controller

In section 4.1 we provided a method of robustness analysis for the case of using trigonometric regressors in the adaptive feedforward controller. Similar approaches can be taken to

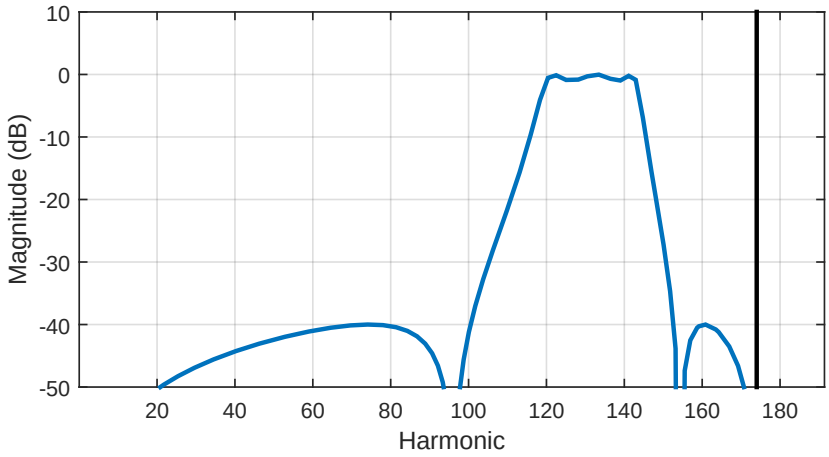


Figure 4.7: Frequency response of a bandpass filter with passband from harmonic 120 to 140. Coefficient of this filter are calculated using (4.18).

analyzing the behavior of adaptive controller when any of the three alternatives proposed in section 3.4 is deployed. In this section we study a fundamentally different methodology for robustness analysis of the adaptive repetitive controller which is "exact", in the sense that it does not rely on any assumptions such as slow adaptation in averaging theory.

Robustness to the mismatches between the actual closed loop system, $R(z)$, and the model $\bar{R}(z)$ is analyzed by outlining an exact linear periodically time varying (LPTV) realization of the closed loop system shown in Fig. 3.2. An important note to make is that our analysis is valid only when the controller can be realized by a periodic dynamics. This means that since the regressor in all cases is periodic, it is enough to have a constant or periodic step size to satisfy this condition. As a result, this machinery can be used to find the maximum constant/periodic step size for an arbitrary (but periodic in time) regressor vector ϕ_k without needing to have a regressor with a special structure as the ones in the four discussed methods.

Suppose that the actual and nominal closed loop transfer functions are related by

$$R(z) = \bar{R}(z) + \Delta(z)\,\bar{R}(z)$$

where $\Delta(z)$ denotes the modeling mismatch. Let $\tilde{\theta}_k := \theta - \hat{\theta}_k$. The *auxiliary error* in (4.1) is

$$\begin{aligned}\bar{e}_k &= e_k + \bar{R}[u_{A,k}] - \bar{\psi}_k^T\hat{\theta}_k \\ &= \left(\bar{R} + \Delta\bar{R}\right)\left[\phi_k^T\tilde{\theta}_k\right] + \xi_k + \bar{R}\left[\phi_k^T\hat{\theta}_k\right] - \bar{\psi}_k^T\hat{\theta}_k \\ &= \bar{\psi}_k^T\tilde{\theta}_k + \Delta\underbrace{\bar{R}\left[\phi_k^T\tilde{\theta}_k\right]}_{\tilde{y}_k} + \xi_k\,.\end{aligned}$$

Suppose that the state space realization of $\bar{R}(z)$ is

$$\bar{R} \equiv \left[\begin{array}{c|c} \bar{A} & \bar{B} \\ \hline \bar{C} & \bar{D} \end{array}\right]$$

which implies that $\bar{R}(z) = \bar{C}\left(zI - \bar{A}\right)^{-1}\bar{B} + \bar{D}$. The newly defined variables $\tilde{y}_k$ can be realized by

$$\begin{aligned} x_{k+1} &= \bar{A}x_k + \bar{B}\left(\phi_k^T\tilde{\theta}_k\right) \\ \tilde{y}_k &= \bar{C}x_k + \bar{D}\left(\phi_k^T\tilde{\theta}_k\right) \end{aligned} \tag{4.19}$$

Let $\delta_{\tilde{y}_k} := \Delta[\tilde{y}_k]$; then, the update equation for the estimates in (4.1) with the above equations imply

$$\begin{aligned} \tilde{\theta}_{k+1} &= \tilde{\theta}_k - \mu_k\bar{\psi}_k\bar{e}_k \\ &= \left[I - \mu_k\bar{\psi}_k\bar{\psi}_k^T \quad 0\right]\begin{bmatrix}\tilde{\theta}_k \\ x_k\end{bmatrix} + \left[-\mu_k\bar{\psi}_k \quad -\mu_k\bar{\psi}_k\right]\begin{bmatrix}\delta_{\tilde{y}_k} \\ \xi_k\end{bmatrix}. \end{aligned} \tag{4.20}$$

and the error signal is described by

$$\begin{aligned} e_k &= \left(\bar{R} + \Delta\bar{R}\right)\left[\phi_k^T\tilde{\theta}_k\right] + \xi_k \\ &= \tilde{y}_k + \delta_{\tilde{y}_k} + \xi_k \end{aligned} \tag{4.21}$$

and the one step evolution equations given by (4.19), (4.20) and (4.21) can be put all together to form a linear periodically time varying system, say G_1 , with state space realization

$$\begin{aligned} X_{k+1} &= \tilde{A}_kX_k + \tilde{B}_kU_k \\ Y_k &= \tilde{C}_kX_k + \tilde{D}_kU_k \end{aligned} \tag{4.22}$$

where the states, input and output are

$$X_k := \begin{bmatrix}\tilde{\theta}_k \\ x_k\end{bmatrix}, \quad U_k := \begin{bmatrix}\delta_{\tilde{y}_k} \\ \xi_k\end{bmatrix}, \quad Y_k := \begin{bmatrix}\tilde{y}_k \\ e_k\end{bmatrix} \tag{4.23}$$

and the augmented matrices are

$$G_1 \equiv \left[\begin{array}{c|c} \tilde{A}_k & \tilde{B}_k \\ \hline \tilde{C}_k & \tilde{D}_k \end{array}\right] = \left[\begin{array}{cc|cc} I - \mu_k\bar{\psi}_k\bar{\psi}_k^T & 0 & -\mu_k\bar{\psi}_k & -\mu_k\bar{\psi}_k \\ \bar{B}\phi_k^T & \bar{A} & 0 & 0 \\ \hline \bar{D}\phi_k^T & \bar{C} & 0 & 0 \\ \bar{D}\phi_k^T & \bar{C} & I & I \end{array}\right].$$

Note that the first input of the system is $\delta_{\tilde{y}_k} = \Delta[\tilde{y}_k]$. As a result, the system can be closed by connecting the first input to the first output through $\Delta(z)$ as is shown in Fig. 4.8. When

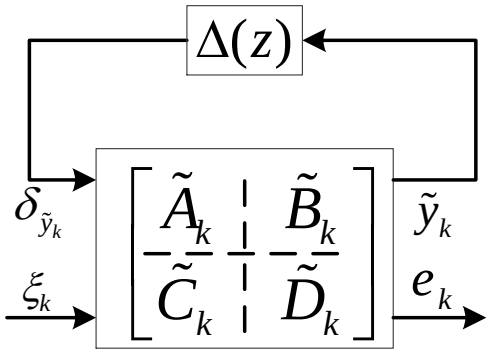


Figure 4.8: LFT of the periodic closed loop system and uncertainty

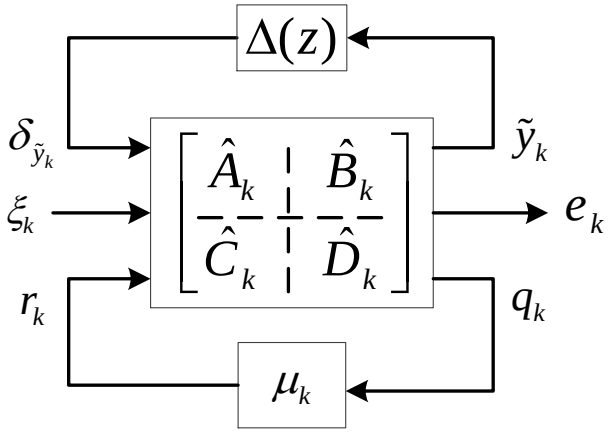


Figure 4.9: LFT of the periodic closed loop system and uncertainty

the step size μ_k is constant or periodic, the robustness analysis of the adaptive algorithm can be conducted by deploying robustness analysis of periodic systems (e.g. discussed rigorously in Bittanti and Colaneri (2009)) for the linear fractional transofrmation shown in Fig. 4.8.

By employing this formalism we can also optimize the step size rather than just restricting ourselves to robustness analysis of the adaptive control algorithm. The dynamics of $\tilde{\theta}$ – realized by the first row of above state space realization for G_1 – inspires the formation of a (lower) linear fractional transformation of μ_k and G_1. Let G_2 be an LPTV system realized by

$$G_2 \equiv \left[\begin{array}{c|c} \hat{A}_k & \hat{B}_k \\ \hline \hat{C}_k & \hat{D}_k \end{array}\right] = \left[\begin{array}{cc|ccc} I & 0 & 0 & 0 & \bar{\psi}_k \\ \bar{B}\phi_k^T & \bar{A} & 0 & 0 & 0 \\ \hline \bar{D}\phi_k^T & \bar{C} & 0 & 0 & 0 \\ \bar{D}\phi_k^T & \bar{C} & I & I & 0 \\ -\bar{\psi}_k^T & 0 & -I & -I & 0 \end{array}\right] . \tag{4.24}$$

The first two inputs and outputs of this system are the same as in (4.23), and the last input

and output are defined such that the LFT shown in Fig. 4.9 is equivalent to Fig. 4.8. That is, $q_k = -\bar{e}_k$, and $r_k = -\mu_k \bar{e}_k$. Based on Fig. 4.9, selecting the step size μ_k can be formalized as a (robust) control design for the (uncertain) LPTV system shown in the figure.

It is worth noting that the application of this periodic representation is not limited to robustness analysis or determining the bounds of a scalar step size. As a matter of fact, this framework enables us to take advantage of analysis and apparatuses developed for linear periodically time varying systems Bittanti and Colaneri (2009) to study/design different variants of our adaptive repetitive controller. For instance, one can go beyond using a static step size for the lower LFT in Fig. 4.9 and close the loop with a dynamic step size. We will not investigate these options here and leave it to future work to determine whether it is possible to synthesize a robust controller by deploying $\mathcal{H}_\infty$ control design methodologies for linear periodic systems Voulgaris et al. (1994) , or whether the $\mathcal{H}_2$ norm of the system can be minimized by using $\mathcal{H}_2$ optimal control synthesis for LPTV systems Voulgaris et al. (1994).

4.6 Summary and Conclusion

Our proposed adaptive algorithm requires a model for the system dynamics which may not be exact, especially at high frequencies, in many practical applications. The robustness of the adaptive feedforward controller to dynamic mismatch between the actual plant and the model deployed in the algorithm was analyzed in this chapter and it was shown that unlike many existing methods, the mismatches are only important at excitation frequencies.

An online system identification architecture was proposed to provide an accurate model of the system dynamics in case a model is not available or accurate. It was shown that under a set of assumptions, the proposed scheme is able to obtain a model that satisfies convergence criterion outlined in the aforementioned robustness analysis. As a result, the identification and compensation mechanisms together form an "indirect" adaptive controller for the class of problems under our study.

The proposed identification scheme requires an exogenous excitation signal. In general, this type of extra excitations are not desired from a practical point of view. We proposed a special excitation that is extremely low power, effective in our problem since its energy is focused around frequencies that are important to us, and easy to generate in DSP implementation. Besides, we suggested using adaptive band–pass filters on the inputs to the identification unit in order to further reduce the required excitation power. A design method for synthesizing these adaptive filters through *frequency transformation* of a prototype filter was proposed. It was shown that explicit "parametric" solutions for the filter coefficients can be obtained. The "parameters" of these relations depend on the pass–band of the filter and can be evaluated easily in real–time.

We also outlined a fundamentally different methodology for robustness analysis of the adaptive repetitive controller by representing the closed loop dynamics of the system as a robust feedback interconnection of an LPTV system with the step size and dynamic un-

certainty. We showed that this realization is not only useful for robustness analysis and determining the bounds of a scalar step size, but also, it enables us to take advantage of analysis and apparatuses developed for LPTV systems to study/design different variants of our adaptive repetitive controller.

Chapter 5

Repeatable Runout Following in HDD Using Indirect Adaptive Control

In this chapter we show how the proposed adaptive control algorithm can be efficiently integrated with the servo controller of an HDD for following / compensating repeatable runout, a problem that was introduced and formalized earlier in chapter 2. Computer simulation in MATLAB and implementation on a digital signal processor (DSP) unit are performed to compensate for RRO that has narrow–band contents at the HDD spinning frequency ($7200RPM$) and its 173 higher harmonics. In other words, n in (3.2) is 173 and

$$\omega_i(rad/s) = i \times 120(Hz) \times 2\pi \qquad i = 1, \cdots, 174.$$

The sampling frequency of the system is 41.760KHz, which is equal to $120Hz \times 348$. This implies that the highest harmonic (at $174 \times 120Hz$) is equal to the controller Nyquist frequency. Another implication is that the disturbance $\bar{d}$ has a period of 348–step in discrete time. More details about this system was earlier provided in chapter 2.

This control task is challenging since it requires estimating a very large number (348) of parameters which is order(s) of magnitude greater than other results reported in the literature. These frequencies span from 120Hz to extremely large frequencies (above 20KHz) where the plant dynamics uncertainties are large and feedback controller amplifies disturbances. Note that, feedback repetitive controllers that compensate a periodic disturbance with a large period (e.g. $N = 348$) can be found in the literature. However, to the best of our knowledge, practical viability of controllers that can compensate a very large number of narrow–band disturbances with selective frequencies have not bee reported yet.

Our contributions in this chapter can be summarized as:

- The adaptive repetitive controller proposed in the previous chapters was discussed for a single–input single–output system. We will extend the methodology to multi–input single–output systems and discuss the benefits that can be gained by following this method to design an adaptive controller for a dual–stage HDD (a system with two actuator inputs and one position error output signal).

- A procedure for modeling system dynamics and contaminating noise dynamics in the dual–stage hard disk drive under our study is described and it is shown how accurately these models can mimic the behavior of the actual system.
- Comprehensive computer simulation study is performed to illustrate the effectiveness of the proposed methods in following multiple sinusoidal trajectories with selective frequencies in the aforementioned HDD simulator. The results are reported for a conventional and a BPMR–like HDD system (i.e. system dynamics belong to a conventional HDD and the noise dynamics belong to a BPMR HDD).
- The algorithm is implemented on a digital signal processor (DSP) and different practical aspects are discussed. Experimental results for RRO compensation on a 3.5" conventional HDD are reported.
- The number of disturbances (trajectories) that are considered in our simulations and experiments is between 50 to 100, which is order(s) of magnitude larger than other results in the literature.

5.1 Adaptive Feedforward Repetitive Control of Dual–Stage Systems

The adaptive feedforward controller developed in the last two chapters was discussed for a single–input single–output (SISO) linear time–invariant system. However, as we mentioned at various junctures throughout the analysis, our method in spectral and wavelet frameworks is not limited to SISO systems and can be extended to multi–input single–output (MISO) systems. More explicitly, one can go over the analysis and replace the SISO systems $R(z)$ and $\bar{R}(z)$ by MISO systems with m input channels

$$\begin{aligned} R(z) &\mapsto \begin{bmatrix} R^1(z) & R^2(z) & \cdots & R^m(z) \end{bmatrix} \\ \bar{R}(z) &\mapsto \begin{bmatrix} \bar{R}^1(z) & \bar{R}^2(z) & \cdots & \bar{R}^m(z) \end{bmatrix} \end{aligned}$$

and instead of transforming the output disturbance r_k to one input disturbance $\bar{d}_k$, transform it to m disturbances $\{\bar{d}_k^1, \cdots, \bar{d}_k^m\}$ on the m individual input channels

$$r_k = R^1\left[\bar{d}_k^1\right] + R^2\left[\bar{d}_k^2\right] + \cdots + R^1\left[\bar{d}_k^m\right].$$

Here, $\bar{d}^i$'s contain non–overlapping frequencies and can be decomposed to known and unknown parts similar to the SISO case

$$\bar{d}_k^1 := (\theta^1)^T \phi_k^1 \quad \bar{d}_k^2 := (\theta^2)^T \phi_k^2 \quad \cdots \quad \bar{d}_k^m := (\theta^m)^T \phi_k^m.$$

Using this decomposition, the control objective is estimating each θ^i to create the corresponding feedforward control u_A^i that should be injected to input channel i. The parameter

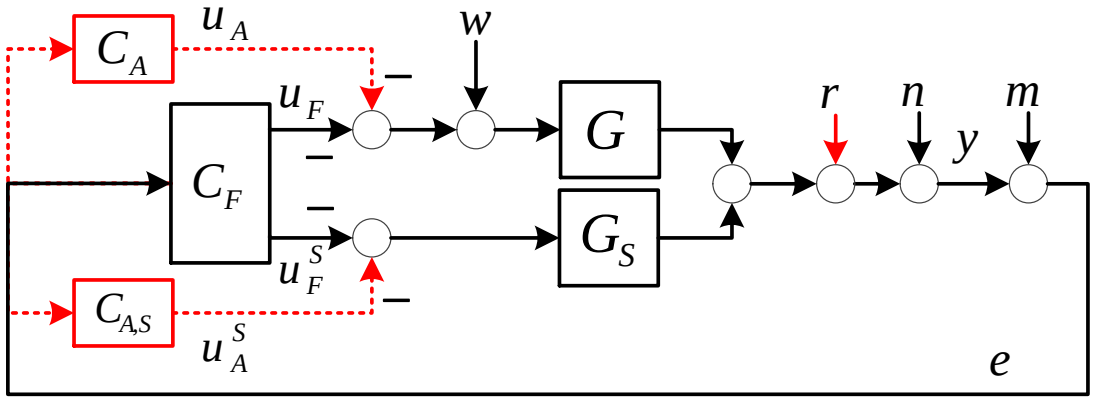


Figure 5.1: Dual–stage system with two adaptive feedforward controllers.

adaptation algorithm for this purpose will be analogous to (4.1)

$$\begin{aligned} \bar{\psi}_k^i &:= \bar{R}^i[\phi_k^i] \\ \bar{e}_k^i &:= e_k + \bar{R}^i\left[u_{A,k}^i\right] - (\bar{\psi}_k^i)^T\hat{\theta}_k^i \\ \hat{\theta}_{k+1}^i &= \hat{\theta}_k^i + \gamma_k^i\mu_k^i\bar{\psi}_k^i\bar{e}_k^i. \end{aligned} \tag{5.1}$$

for all channels $1 \leq i \leq m$. Accordingly, the adaptive control signal for input channel i is

$$u_{A,k}^i = (\hat{\theta}_k^i)^T\phi_k^i. \tag{5.2}$$

Equations (5.1) and (5.2) together define one *adaptive controller module* that receives the error signal e_k and injects control to one input channel in order to compensate the disturbance at certain frequencies.

This idea is illustrated for a dual–stage (dual–input single–output) HDD servo system in Fig. 5.1. The two adaptive controller modules, C_A and $C_{A,S}$ compensate for two disjoint frequency sets of the Repeatable Runout r. Recall from chapter 2 that G and G_S denote, respectively, the voice–coil motor (VCM) and the "S"econdary actuator (e.g. a mili– or micro–actuator (MA)).

The architecture in Fig. 5.1 may raise the question of why the compensation signal is not injected to the input of controller C_F which will require only one adaptive controller module. There are two reasons that prevent us from pursuing that approach: (1) The VCM and MA, and in general different actuators in a MISO system, have different operational constraints. As for a dual–stage HDD servo, the control effort for the two actuators have different saturation limits. Moreover, the MA stroke is limited. Since the two actuators have different magnitude responses in frequency domain, the control effort required for compensating a sinusoid at a particular frequency by each of them is not equal to the other one. For instance, the VCM gain drastically drops at high frequencies and there might not exists enough control effort available to the VCM to compensate a high frequency disturbance.

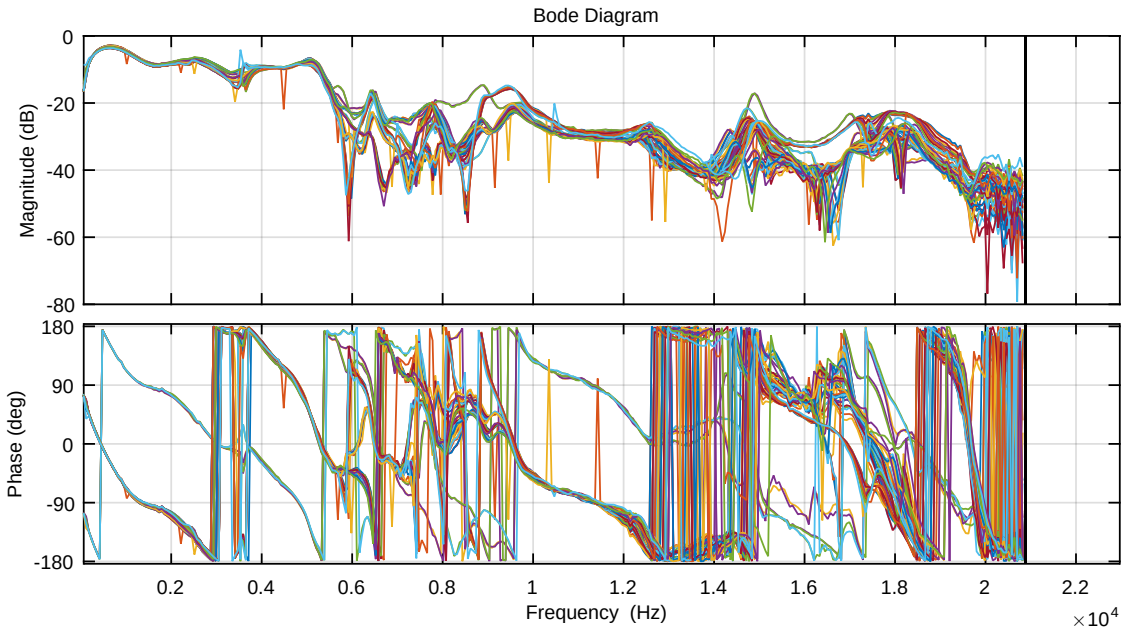


Figure 5.2: Secondary path responses from inputs of VCM to the position error signal in one HDD. Data is measured during servoign on 8 different heads and at temperatures varying from 34°_{C} to 48°_{C}.

The architecture in Fig. 5.1 let us satisfy these constraints on each of the actuators by allocating different frequency contents to each of them based on their characteristics. This is not possible when the control is injected to the input of C_F because it will be distributed between the two channels based on the dynamics of C_F and we cannot directly determine the contribution of each actuator. (2) The uncertainty level on the two actuators is not the same at different frequencies. The MA dynamics has high uncertainty at very low and high frequencies, whereas the VCM dynamics at low frequency is exact, but after a few Kilo Hertz it becomes very uncertain. The closed loop system frequency response from the MA input and VCM input to the position error signal (PES) are shown in Fig. 4.1 and Fig. 5.2 respectively. We can take advantage of our prior knowledge about the actuators and use the secondary path identification mechanism on each actuator only at frequencies that we are aware of large uncertainties in that particular actuator.

5.2 Modeling

This section describes our modeling methodology to construct a realistic simulation environment. The following subsections provide details regarding modeling dynamic systems involved in Fig. 5.1, noise models and the results of different case studies carried out in this framework.

5.2.1 Modeling Dynamic Systems

The system dynamics and disturbances in Fig. 5.1 are modeled based on an accurate frequency response of the servo system and power spectrum of the error signal e. Frequency response of the system from the adaptive control injection point to the position error signal (PES) is obtained by multi–frequency Fourier transformation method and a high order LTI transfer function is fitted to it. The frequency response from the MA input to the PES

$$R_1(z) := \frac{G(z)}{1 + \begin{bmatrix} G(z) & G_S(z) \end{bmatrix} C_F(z)}$$

obtained by this type of measurements is plotted in Fig. 5.3 (blue line). The figure also shows the response of an LTI model that is fitted by MATLAB `tfest()` function to the measured frequency response data (FRD). The `tfest()` settings are set such that the transfer function coefficients are initialized by `arx` method and then fine tuning is performed by a nonlinear least squares with automatically chosen line search method. The number of poles and zeros are increased manually till at least 90% fitting accuracy in frequency domain is achieved. For this particular response, a 17^{th} order model can achieve 93% accuracy. The model's frequency response is shown by the red line in Fig. 5.3 and the transfer function coefficients are given in Table B.1 in the appendix. The same procedure is used to obtain an accurate dynamic model from the VCM input to the PES, in closed loop

$$R_2(z) := \frac{G_S(z)}{1 + \begin{bmatrix} G(z) & G_S(z) \end{bmatrix} C_F(z)} \,.$$

The actual and model frequency responses are shown in Fig. 5.4. In this case, since the FRD model is more complicated than the MA case, a 50^{th} order system is required to achieve the above requirement. Table B.2 in the appendix presents the transfer function coefficients of this model.

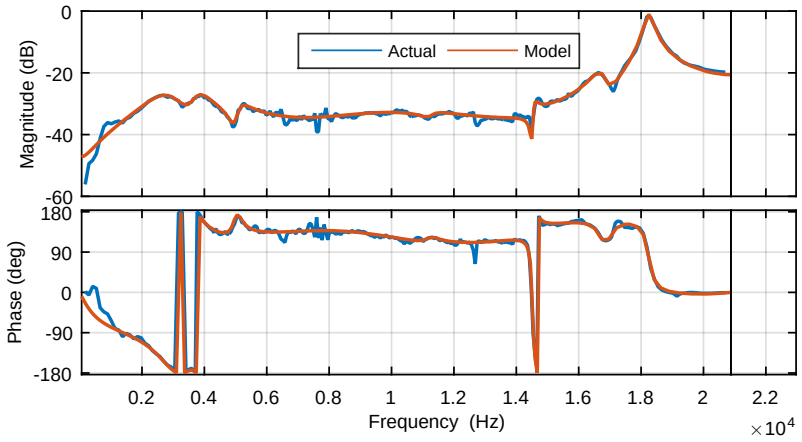


Figure 5.3: Frequency response of the closed loop dynamics from the MA injection point to the PES. The actual response is obtained from multi–frequency swept sine measurements and a 17^{th}–order LTI "model" is fit to it.

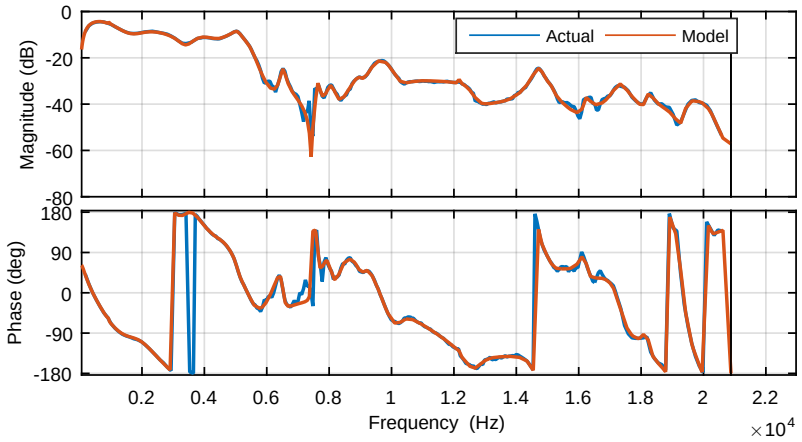


Figure 5.4: Frequency response of the closed loop dynamics from the VCM injection point to the PES. The "actual" response is obtained from multi–frequency swept sine measurements and a 50^{th}–order LTI "model" is fit to it.

5.2.2 Modeling Noise Dynamics

A long–time series of the error signal that contained approximately 100'000 measurements of the error when no external control is applied is collected from the HDD of our experimental setup – i.e. $e(k)$ is measured when no adaptive control is injected to the system. Let $S(z)$

be the sensitivity function on the output side

$$S(z) := \frac{1}{1 + \begin{bmatrix} G(z) & G_S(z) \end{bmatrix} C_F(z)}.$$

Based on Fig. 5.1, the error signal is

$$e(k) = \underbrace{\left(S\left[G[w_k]\right] + S[m_k + n_k]\right)}_{\xi_k} + \underbrace{S[r_k]}_{\bar{r}_k}$$

where $\bar{r}(k)$ is due to the RRO and ξ_k is due to all other noises. Figure 5.5a shows the error spectrum (red line) versus the frequency in terms of harmonic number. Note that the tall bars located at harmonics correspond to the RRO. As a result, the RRO profile, which is in fact the disturbance $\bar{r}(k)$, can be extracted from the error signal by Fourier transformation performed at the harmonics, or equivalently by a periodic averaging in time domain which is the approach taken here. The residue $e_r := e - \bar{r}$ is due to the NRRO that has a spectrum as shown in Fig. 5.5a (blue line). This signal is used in a time domain identification algorithm to obtain an AR model

$$e_r(k) = \frac{C(q^{-1})}{A(q^{-1})}\bar{w}(k).$$

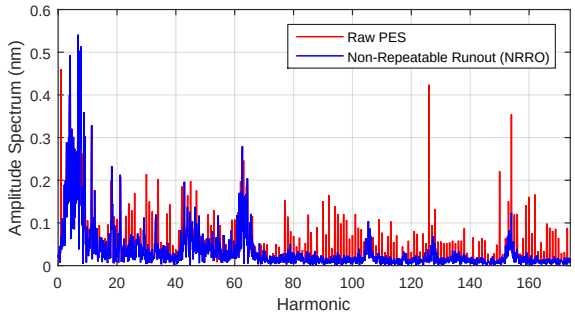
For this time domain fitting purpose we used an iterative search algorithm Ljung (1977a) that minimizes a robustified quadratic prediction error criterion. The cutoff value for the robustification is based on a threshold estimation option and on the estimated standard deviation of the residuals from the initial parameter estimate. The output spectrum of this estimated model when the input is white noise is compared with the original NRRO spectrum in Fig. 5.5b. Moreover, the DC gain of the model is set such that the $\mathcal{H}_2$ norm of the model matches the NRRO variance.

5.2.3 BPMR HDD Simulator

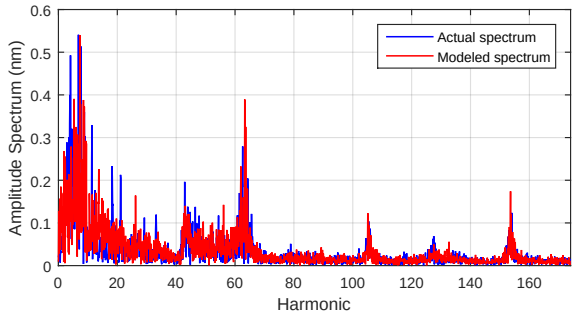
We use the same dynamic models and broad band noises to simulate a BPMR HDD system because the mechanical components and environmental noises are common among conventional and BPMR hard disk drives. However, the RRO is different since it is partially caused by media which is the inherent distinction between the conventional magnetic recording and BPMR. We use the RRO data from a prototype BPMR hard disk drive that is provided by our industry partner to create a realistic RRO profile for our simulations.

5.2.4 Control Design for Simulation Study

As mentioned above, the RRO profile is the summation of 174 sinusoidal disturbances that are at the spindle frequency and its higher harmonics. We split the compensation of this large number of narrow–band disturbances between the VCM and MA based on two factors: (1) the limits of available control (in terms of voltage/current) to the two actuators (2) a



(a) PES spectrum



(b) PES spectrum after removing RRO

Figure 5.5: PES spectrum versus frequency in terms of integer multiples (harmonics) of spinning frequency. (a-red): spectrum of raw PES. (a-blue): spectrum of NRRO that is by the definition the residue of PES after subtracting the RRO. (b-blue): similar to (a-blue). (b-red): spectrum of the output of a fitted dynamic model when its input is white noise. (All measurements are performed in closed loop.)

very rough knowledge about the amplitude of harmonics. These two pieces of information and the frequency response of the two loops shown in Fig. 5.3 and 5.4 imply that the low frequency contents should be allocated to the VCM and the left over should be compensated by the MA. In our case, controller C_A in Fig. 5.1 compensates the first 80 harmonics and the other controller, $C_{A,S}$, compensates harmonics 81 to 174. It should be remarked that these ranges are not unique and can be modified slightly. The two adaptive controllers, C_A and $C_{A,S}$, are then constructed based on the algorithm given in 5.1 for $m = 2$ with the regressors

$$\phi_k^1 := \begin{bmatrix} \alpha_1 \sin(\omega_0 kT) \\ \vdots \\ \alpha_{n_1} \sin(n_1\omega_0 kT) \\ \alpha_1 \cos(\omega_0 kT) \\ \vdots \\ \alpha_{n_1} \cos(n_1\omega_0 kT) \end{bmatrix}, \qquad \phi_k^1 := \begin{bmatrix} \alpha_{n_1+1} \sin((n_1+1)\omega_0 kT) \\ \vdots \\ \alpha_{n_2} \sin(n_2\omega_0 kT) \\ \alpha_{n_1+1} \cos((n_1+1)\omega_0 kT) \\ \vdots \\ \alpha_{n_2} \cos(n_2\omega_0 kT) \end{bmatrix}$$

where ω_0 denotes the fundamental frequency, $n_1 = 80$, and $n_2 = 174$.

The minimum achievable 3σ (3 times the standard deviation) error that corresponds to the case of perfect RRO tracking is $4.69nm$ in our setup. We choose the desired performance level that is required in the step size formulation (3.17) according to this number. The value of V^d chosen here corresponds to a 3σ value of $6.22nm$. Recall this means that the adaptive controller's step size is positive as long as the error 3σ is (approximately) above this value.

5.2.5 Simulation Results

The two adaptive controllers C_A and $C_{A,S}$ are constructed based on the previous section and then applied to the dual–stage system shown in Fig. 5.1. The RRO profile considered

at the beginning of our time simulation belongs to a "mid–diameter" track. We change this profile after 100 revolutions of the disk, which is equal to approximately 0.8 seconds. The second profile belongs to an "outer–diameter" track which is slightly different than the first profile.

The time–series of position error signal (PES) is shown in Fig. 5.6. The figure shows that the maximum amplitude of error converges in less than 0.2 seconds. After we change the RRO profile (after 0.8s), the controller converges in less than 0.05 seconds. The difference between these two convergence times is because of the similarity between the two profiles. This means that the adaptive controller requires less number of iterations to learn compensating the second profile when it already knows how to compensate the first one. Note that this figure only shows the maximum amplitude of error and it is difficult to analyze the standard deviation behavior based on it. Instead, we use Fig. 5.7 to illustrate an "approximation" of

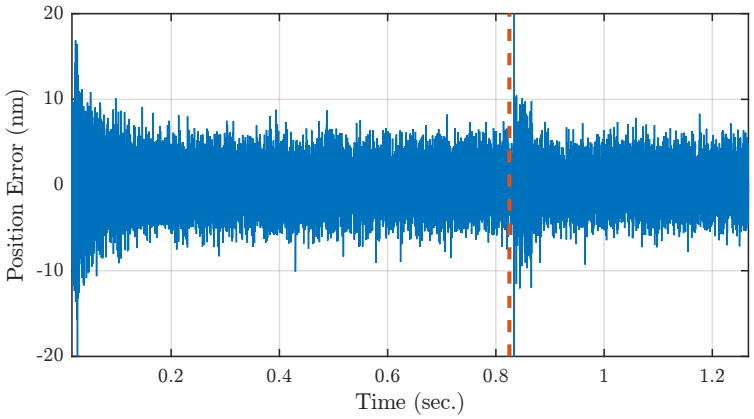


Figure 5.6: Error in tracking repeatable runout (RRO). The RRO profile changes after approximately 0.8 seconds.

the instantaneous 3σ value of error. More explicitly, we ran a moving average on the squared error and calculated its square root to approximate the instantaneous standard deviation of error. The value plotted in this figure is 3 times this quantity. As can be seen in the figure, it takes approximately the same time for the 3σ value to converge.

Recall that our proposed step size formulation (3.16, 3.17, 3.18) adjusts the step size based on an approximation of auxiliary error variance. The variable V^h in (3.16) provides a moving average on the squared auxiliary error and keeps the step size positive as long as this value is above V^d, which was defined as the *desired variance level*. Figures 5.8a and 5.8b show the 3σ values corresponding to V^d and V_k^h in C_A and $C_{A,S}$ respectively. The black dotted line in each figure shows the desired performance $3\sqrt{V^d}$ (i.e. in terms of 3σ) that was chosen in the previous section. The other horizontal line (magenta dashed line) in each figure corresponds to the upper bound of the dead–band defined in (3.18). Note that the two plots become very similar quickly. This is because the auxiliary errors of both controllers converge to the actual error, and as a results, the two values represent roughly the same

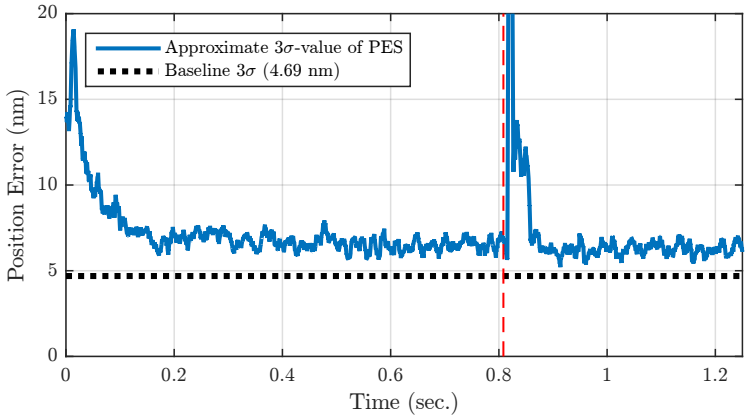
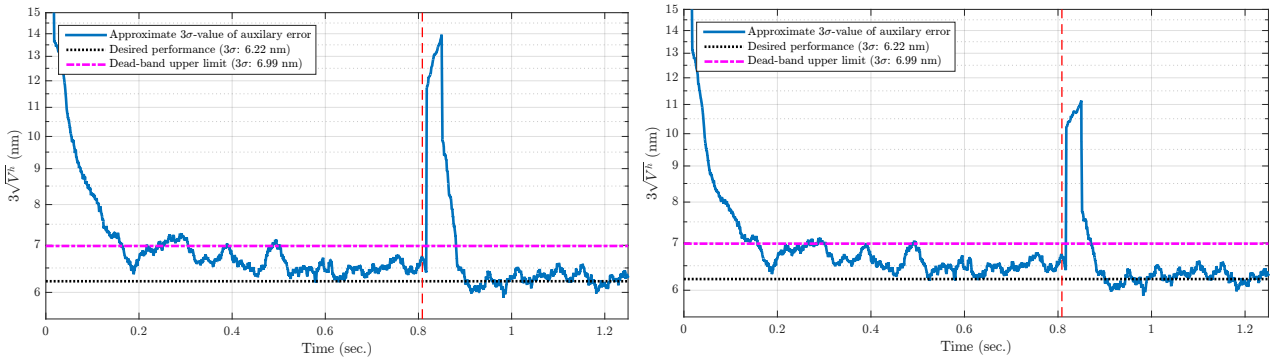


Figure 5.7: Approximate 3σ value of tracking error. The horizontal dashed line shows the minimum achievable 3σ. This value is due to all other noises when RRO is perfectly compensated.

quantity.



(a) Approximate 3σ of aux. error in C_A (b) Approximate 3σ of aux. error in $C_{A,S}$

Figure 5.8: Approximate 3σ value of auxiliary error (i.e. $3\sqrt{V^h}$ in (3.16)) that is used for determining the step size.

Our proposed step size mechanism keeps updating the adaptive controller parameters as long as the estimated 3σ value of error is above the dead–band. The adaptation continues when the performance enters the dead–band from above (from an unsatisfactory state) unless it reaches the desired value (black line). During this process, the step size is adjusted based on the distance between the current and desired performance. This behavior can be seen in Fig. 5.9 and 5.10 which respectively show the step size in C_A and $C_{A,S}$. The figures show that the step size decreases as the estimates approach the actual parameters. This is a desirable behavior in a stochastic environment because the contaminating noises make the estimates fluctuate around the actual parameters. The amplitude of these fluctuations depends on the adaptive algorithm step size and a small step size provides a better estimate.

From comparing these two figures with Fig. 5.7 it can be seen that the large step size in the beginning brings the estimates close to the actual parameters quickly, and then, the step size adjusts to this new condition and starts fine tunning the parameters. Finally, the adaptation stops, meaning that the step size becomes zero, when the desired performance is achieved.

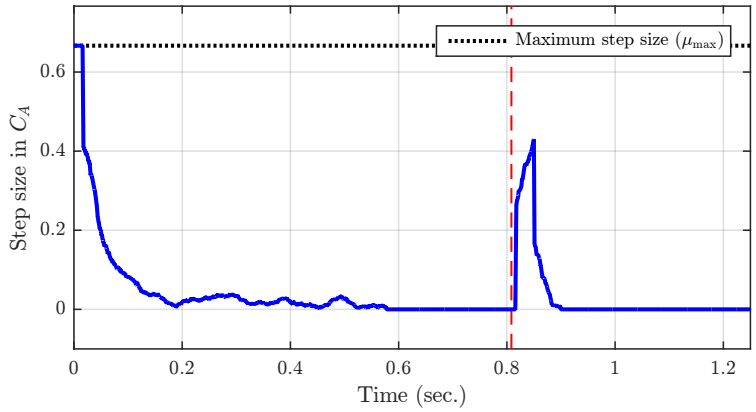


Figure 5.9: Variable Step size (μ_k in (3.18)) in the parameter adaptation algorithm of C_A.

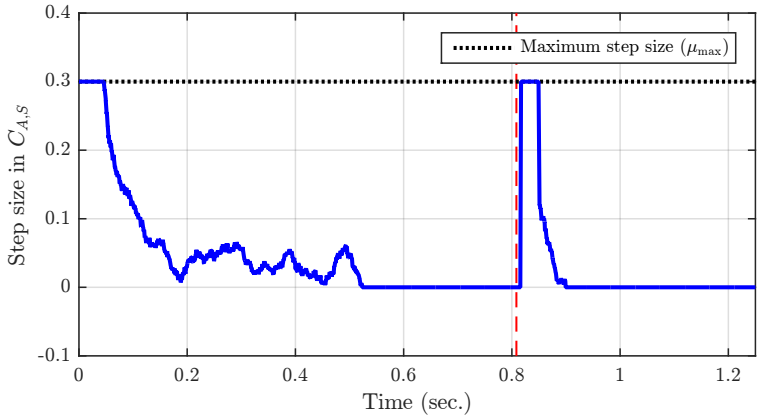


Figure 5.10: Variable step size (μ_k in (3.18)) in the parameter adaptation algorithm of $C_{A,S}$.

Since the system is stochastic, the approximate 3σ value of error is not exactly equal to the desired value even after convergence. This can be seen in Fig. 5.8, where $3\sqrt{V^h}$ fluctuates around $3\sqrt{V^d}$. However, the dead–band we considered in the variable step size scheme adds a hysteresis behavior that avoids the step size to chatter in this case. Nevertheless, the algorithm is still self–tunning, meaning that it starts the adaptation again whenever the disturbance changes. This can be seen in both figures (5.9) and (5.10) because both step sizes automatically become positive when we change the RRO profile (after 0.8 seconds). Moreover, it can be observed in the evolution of estimates that is shown in Fig. 5.11 for

C_A, and in Fig. 5.12 for $C_{A,S}$. The figures illustrate that the estimates take off from zero initial values and approach the convergence values quickly. However, as they get closer the adaptation becomes slower to provide more precise estimates.

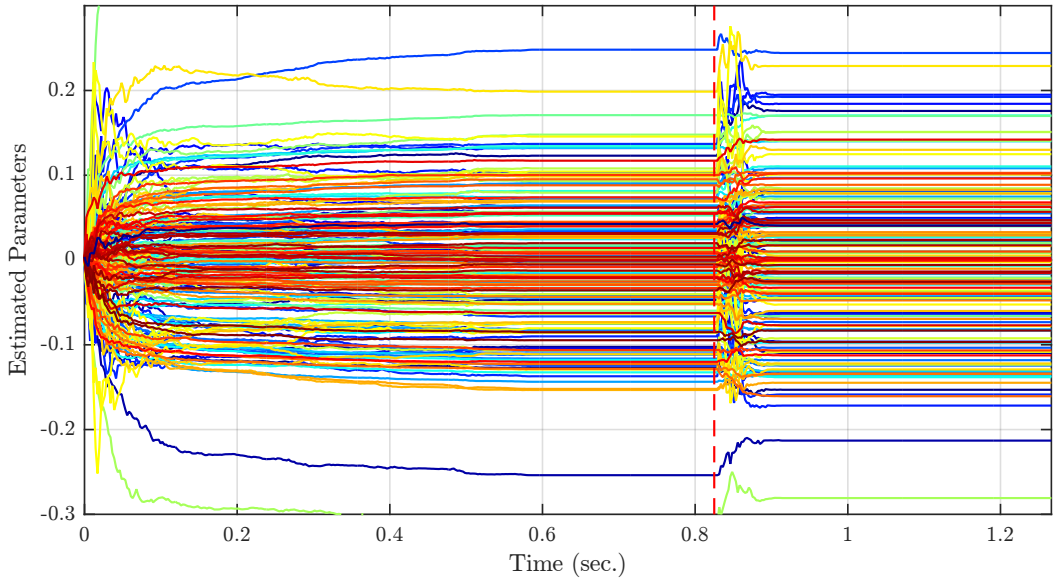


Figure 5.11: Evolution of parameters estimated by C_A. There are 160 parameters in the figure that correspond to the amplitudes of sine and cosine functions at harmonics 1 to 80.

In order to verify whether the convergence values are close to the actual parameters or not, the error spectrum before and after plugging the adaptive controller is shown in Fig. 5.13. Very small left over can be observed in the plot. This is because we chose "our desired" performance, V^d, slightly larger than the baseline error variance.

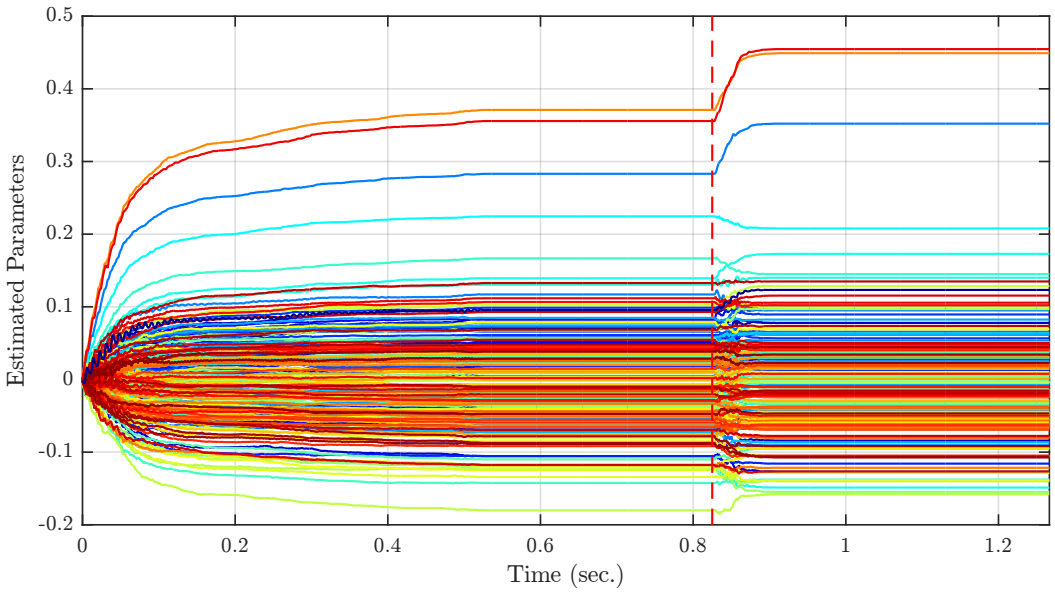


Figure 5.12: Evolution of parameters estimated by $C_{A,S}$. There are 188 parameters in the figure that correspond to the amplitudes of sine and cosine functions at harmonics 81 to 174.

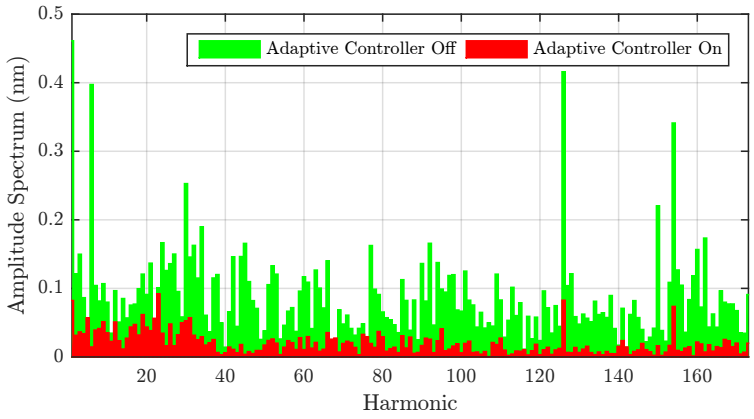


Figure 5.13: Tracking error spectrum before and after plugging the adaptive controllers to the dual–stage HDD (simulation).

5.3 Implementation Results

The proposed control and modeling algorithms are tested on a conventional 3.5″ HDD to validate the presented analytical and simulation results. Since the controller sampling frequency (41.760KHz) is relatively large compared to the clock rate of our digital signal processor (435MHz), it is not practical to estimate all 348 parameters concurrently. However, thanks to the modular design of our adaptive control scheme, and the scheduling parameters γ^i introduced in 3.3.4, we are able to easily split the computation by constructing multiple nonconcurrent adaptive controller modules and plugging them to the closed loop system. We use three modules, denoted by C_A, $C_{A,S,1}$ and $C_{A,S,2}$, and plug the first one to the VCM input and the other two to the MA input – i.e. $C_{A,S,1}$ and $C_{A,S,2}$ are plugged in the same way as $C_{A,S}$ in Fig. 5.1.

The 174 frequency contents of RRO are split between these three modules as follows. The "low frequency" contents, which include harmonics 1 to 55, are allocated to C_A since, as we mentioned in the simulation study, the closed loop system from the VCM input to the PES has a larger gain and smaller uncertainty. Note that the interval we call "low–frequency" includes contents up to 6.6KHz which is considered as a "high" frequency in related work. The "mid frequency" harmonics, including harmonics 56 to 115, are designated to $C_{A,S,1}$ and the remaining harmonics, namely 116 to 174, are left to $C_{A,S,2}$. The scheduling parameters in these modules are set such that compensation starts with $C_{A,S,1}$ and continues for at most 30 revolutions of the disk (0.25 sec). It is followed by running $C_{A,S,2}$ for at most 50 revolutions (0.4167 sec) and, subsequently, C_A for another 50 revolutions. These specifications are listed in Table 5.1.

Recall that the proposed adaptive control algorithm requires a dynamic model from the control injection point to the error signal such that its phase at compensation frequencies does not deviate more than 90 degrees (in absolute value) from the phase of actual system (c.f. criterion (4.4)). We use the online secondary path modeling scheme proposed in chapter 4.2 to attain two types of dynamic models, namely from the VCM and from the MA inputs to the PES. As for the dynamic model from the VCM input to the PES, the exogenous excitation signal proposed in section 4.3 in accordance with the use of band–pass filters suggested in section 4.4 result in an identification process that is focused at harmonics 1 to 55 – i.e. the same range as the compensation. We use a 4^{th} order adaptive IIR transfer

Table 5.1: Adaptive controller modules and frequency partitions considered in DSP implementation.

Controller name	Injection point	Compensation harmonics	Number of parameters	Active time
C_A	VCM	1 – 55	110	0.67–1.08 sec
$C_{A,S,1}$	MA	56 – 115	120	0.00–0.25 sec
$C_{A,S,2}$	MA	116 – 174	118	0.25–0.67 sec

function with relative degree 1. The model identified after 200 revolutions of the disk (1.67 seconds) is compared to the actual system dynamics – which is measured by an accurate swept sine method – in Fig. 5.14 (left). The maximum absolute phase mismatch between the model and actual system during the evolution of model parameters is illustrated in the same figure (left). The left plot shows how effectively the special excitation signal and band–pass filter could focus the identification process in the desired region, and the right plot illustrates that the convergence criterion (4.4) is satisfied after approximately 1.1 seconds.

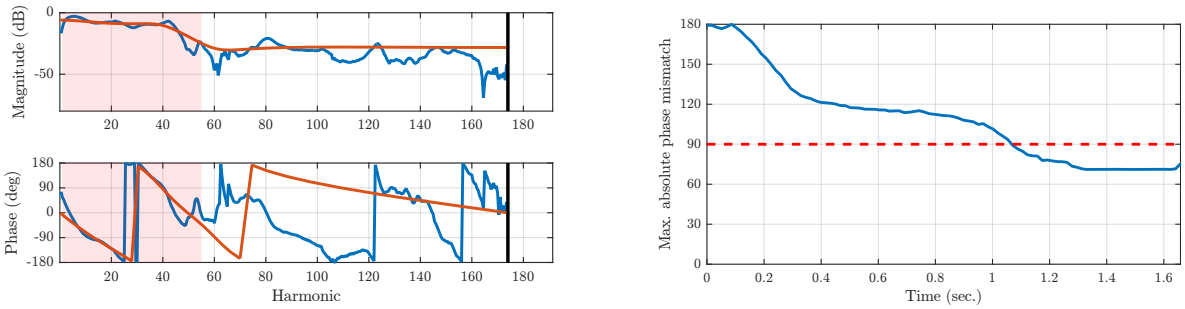


Figure 5.14: Left: Frequency response of the identified model (red) and the actual closed loop dynamics (blue) from VCM input to the PES. Excitation energy is focused to the shaded area (harmonics 1 to 55). Right: Maximum absolute phase mismatch (in degrees) between the model and actual dynamics only in the identification frequency range (shaded area in the left plot). Values below the 90-degree horizontal line result in a converging adaptive controller.

The same secondary path modeling scheme is applied to attain a model from the MA input to the PES. However, it can be seen from Fig. 5.4 that due to the sharp phase shift around 14.5KHz, a high order model is required to satisfy the phase matching criterion (4.4). The modularity of our control design let us use two simple low order models rather than one high order system. Accordingly, we separate harmonics 120 to 124 – that are located around the sharp phase change – and designate one identification module to them, and allocate the remaining harmonics from 56 to 174 to another identification module. It is shown in Fig. 5.15 and 5.16 that both modules were able to identify the system within a 90–degree phase mismatch window at target frequency ranges.

The identified models are then integrated with the three adaptive control modules listed in Table 5.1. The feedforward control profile is constructed by superposing the control sequences generated by these controllers. In particular, we store the periodic control signal generated by the first controller and use it as a feedforward control during executing the second one. Similarly, the superposition of the first and second control sequences are used as a feedforward control when the third controller is learning to compensate its specified frequency range. Note that, in principle, the order of applying these controllers should be unimportant if the system is linear time–invariant. However, our experiments showed that due to system dynamics non–linearities the best results can be achieved by prioritizing the

modules according to the aforementioned order.

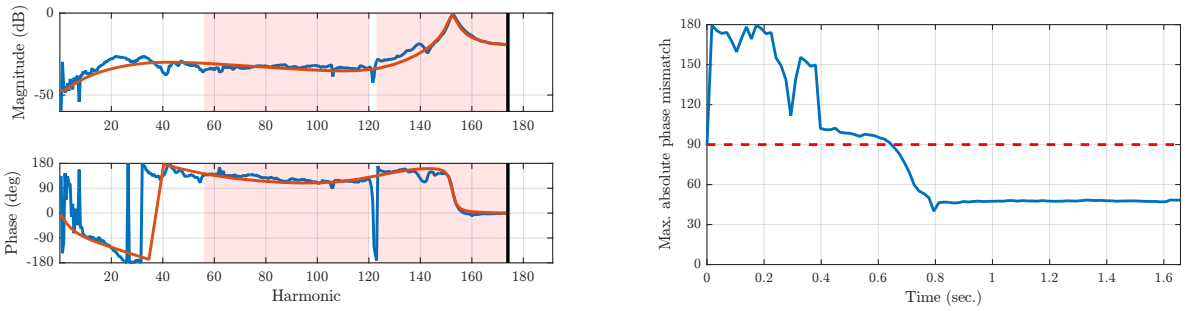


Figure 5.15: Left: Frequency response of the identified model (red) and the actual closed loop dynamics (blue) from MA input to the PES. Identification is performed only at the shaded area (harmonics 56 to 120, and 124 to 173). Right: Maximum absolute phase mismatch (in degrees) between the model and actual dynamics only in the identification frequency range (shaded area in the left plot).

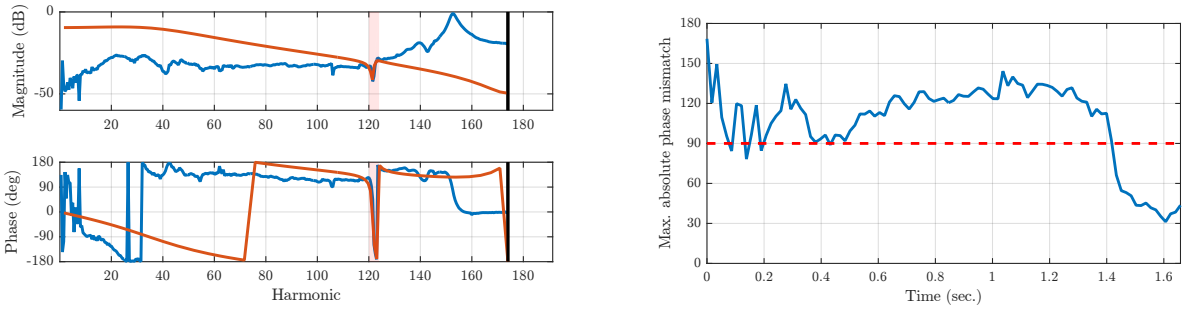


Figure 5.16: Left: Frequency response of the identified model (red) and the actual closed loop dynamics (blue) from MA input to the PES. Identification is performed only at the shaded area (harmonics 120 to 124). Right: Maximum absolute phase mismatch (in degrees) between the model and actual dynamics only in the identification frequency range (shaded area in the left plot).

The evolution of estimated parameters is shown in Fig. 5.17. The vertical dashed lines split the time to three intervals in which (from left to right) the three modules compensate "mid", "high" and "low" frequency harmonics according to Table 5.1. As can be seen from the figure, all parameters converge in the designated interval. Similar to the simulation study, we check the closeness of convergence values to the actual parameters by looking at the spectrum of tracking error which is plotted in Fig. 5.18. It can be seen from comparing this figure with the results illustrated in Fig. 5.13 that the experiments and simulations are in very close agreement.

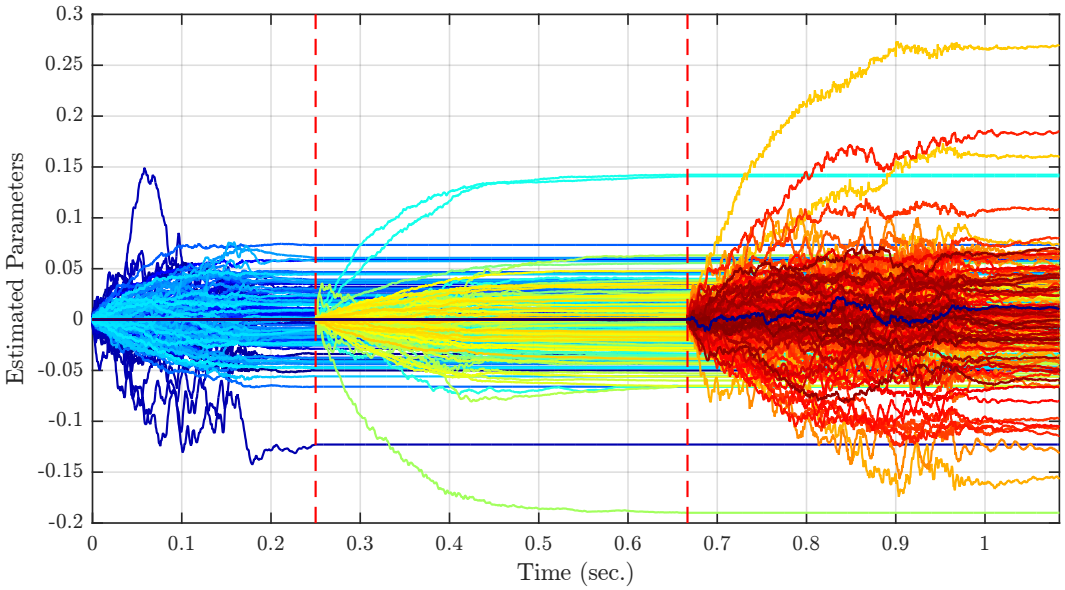


Figure 5.17: Evolution of 348 parameters estimated by three adaptive control modules implemented on a DSP and plugged to a 3.5" HDD.

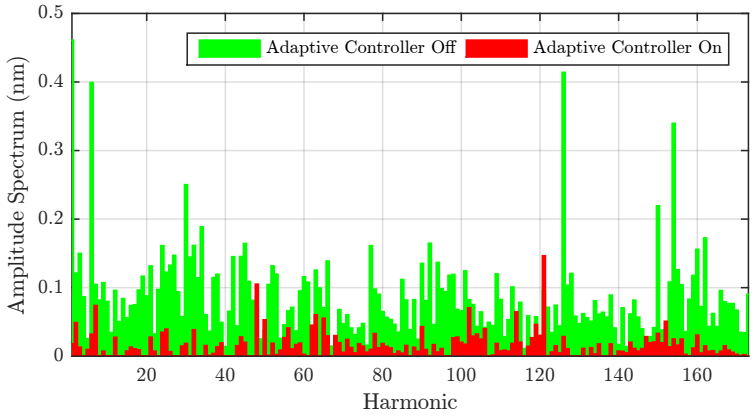


Figure 5.18: Tracking error spectrum before and after plugging the adaptive controllers to the dual–stage HDD (DSP implementation).

In this experiment, the 3σ value of position error signal before and after plugging the adaptive controllers to the system were 6.00 nm and 5.09 nm respectively. This implies that the adaptive controllers were able to improve the performance in terms of the total 3σ by 18%. Our analysis shows that the portion of 3σ due to non–repeatable runout (broad–band noises) is 4.89 nm. Therefore, the RRO contribution before and after using the controllers is 1.11 nm and 0.20 nm, which implies that the controllers were able to reduce the RRO by 82%.

5.3.1 Regressor Implementation

One of the key challenges in implementing the proposed adaptive controller on an embedded system is evaluating multitude sinusoidal and cosinusoidal functions that are used in the regressor vectors ϕ_k and $\bar{\psi}_k$. For instance, since the RRO profile is consisted of 174 sinusoids, it is required to evaluate 696 trigonometric functions which can be a very computationally intensive task. In this subsection we show when the narrow–band disturbances are all higher harmonics of one fundamental frequency, which is the case for RRO tracking, the computation can be effectively decreased by a simple method that is proposed below. The key idea is that a sinusoid at the fundamental frequency is sampled and stored as a reference table in the memory and then all higher harmonics and cosine values with arbitrary phase lags are generated from this reference.

Let N be the period of a discrete sequence obtained by sampling (every T seconds) a sinusoid running at fundamental frequency ω_0. In an HDD, N is equal to the number of samples obtained in each revolution of the disk since the fundamental frequency is equal to the disk spinning frequency. Suppose $x \in \Re^N$ is an array (e.g. a buffer in the DSP) such that

$$\begin{aligned} x &:= \begin{bmatrix} x_0 & x_1 & \cdots & x_N \end{bmatrix} \\ x_j &= \sin(\omega_0 jT). \end{aligned}$$

A sinusoidal at the i^{th} harmonic of fundamental frequency with an initial phase advance δ_i can be approximated by

$$\sin(\omega_i kT + \delta_i) \approx x_{p_i(k)}$$

where

$$\begin{aligned} p_i(0) &= \lfloor \frac{\delta_i}{\omega_0 T} \rfloor \\ p_i(k+1) &= \begin{cases} p_i(k) + i & \text{if } p_i(k) + i < N \\ p_i(k) + i - N & \text{otherwise} \end{cases}. \end{aligned}$$

This means that evaluating a higher harmonic with an arbitrary phase shift is equivalent to evaluating the above simple recursive expressions to find the corresponding index that should be read off from the reference array x. Hence, a trigonometric function evaluation becomes a process that requires one "summation" and one "logical" evaluation. A corollary of this result is that any cosinusoidal can be approximated by $\cos(\omega_i kT + \delta_i) \approx x_{p_i(k)}$ where the initial p_i is now $p_i(0) = \lfloor \frac{\delta_i + \pi/2}{\omega_0 T} \rfloor$.

The proof of this results is very straightforward. We have

$$\begin{aligned} \sin(\omega_i kT + \delta_i) &= \sin(i\omega_0 kT + \delta_i) \\ &= \sin\left(\left[ik + \lfloor \frac{\delta_i}{\omega_0 T} \rfloor \right] \omega_0 T + \epsilon_i \right) \\ &\approx \sin\left(\left[ik + \lfloor \frac{\delta_i}{\omega_0 T} \rfloor \right] \omega_0 T \right) \end{aligned} \tag{5.3}$$

where $\epsilon_i = \delta_i - \lfloor \frac{\delta_i}{\omega_0 T} \rfloor \omega_0 T$ and the approximation is accurate if ϵ_i is small. The expression for ϵ_i implies that $0 \leq \epsilon_i \leq \frac{1}{\omega_0 T}$. As a result, the accuracy of above approximation depends on $1/(\omega_0 T)$. In our particular case, ω_0 is equal to $\frac{2\pi}{NT}$ which implies that $0 \leq \epsilon_i \leq \frac{2\pi}{N} = 1.03^\circ$ since $N = 348$. This means that the above approximation results in a sinusoid that at most has 1.03° phase mismatch. Note that this error is far below 90 degrees phase mismatch that our proposed control algorithm can tolerate. Since N is one period of the discrete sequence, we have

$$\sin\left(\left[ik + \lfloor \frac{\delta_i}{\omega_0 T} \rfloor\right] \omega_0 T\right) = \sin\left(\text{mod}\left\{\left[ik + \lfloor \frac{\delta_i}{\omega_0 T} \rfloor\right], N\right\} \omega_0 T\right)$$

where mod(.,.) denotes the modulo operator. One can show that

$$p_i(k) = \text{mod}\left\{\left[ik + \lfloor \frac{\delta_i}{\omega_0 T} \rfloor\right], N\right\}$$

and accordingly

$$\sin\left(\omega_i k T + \delta_i\right) \approx \sin\left(p_i(k)\omega_0 T\right) = x_{p_i(k)}.$$

Chapter 6

Direct Adaptive Control for Rejecting Multiple Sinusoidal Disturbances

6.1 Introduction

In this section we introduce a new direct adaptive control for rejecting multiple sinusoidal signals with known frequencies. We denote this method as a "direct" algorithm since the control parameters are directly updated by the parameter adaptation algorithm and a "control design" block that commonly exists in indirect methods is avoided. The key distinction between this algorithm and the methods elaborated in the previous sections is that the plant parameters estimation and control synthesis are not performed by two completely separate mechanisms. The advantages of this method over the proposed indirect adaptive control are (1) In the indirect adaptive control, a special monitoring mechanism is required to assure that the model provided by the system identification mechanism satisfies the convergence criterion of the adaptive compensator. The direct adaptive control proposed in this section integrates both parts and, as a result, does not rely on such a mechanism. (2) The proposed direct method can be applied to slowly time–varying systems because the system dynamics is identified simultaneously with the adaptive control law. This method uses an adaptive gain for the excitation signal that makes the algorithm self–tuning when the system dynamics is changing slowly, or when it has abrupt but infrequent variations. In order to make the indirect adaptive controller applicable to this type of environments, we either need to constantly excite the system or use a monitoring mechanism that activates the system identification part whenever the system dynamics varies.

The proposed method is based on inversion of the plant transfer function numerator frequency response evaluated at compensation frequencies. Since this inversion is carried out in frequency domain, the method is applicable to both minimum and non–minimum phase systems. The plant dynamics is considered as a parametric model, which reduces the number of estimated parameters considerably when the number of frequencies, n, is large. The key idea in this approach is that when the control injection point and measurement point are distinct, the error signal – e.g. the PES – contains information about both the

system dynamics and the disturbances. It can be shown that the error signal is a nonlinear – in fact bilinear – function of the two sets of parameters. Under some mild assumptions and proper use of signal filtering we will show, by theoretical analysis and experiments, how standard linear adaptive control and identification can be combined to deal with this type of bilinear system of equations.

The main contributions of our proposed algorithm can be summarized as follows.

- The number of adapted parameters in this method is significantly less than the state-of-the-art when the number of narrow–band contents is large. For instance, the number of parameters in Chandrasekar et al. (2006) and Pigg and Bodson (2006), Pigg and Bodson (2010) is $4n$, where n is the number of frequencies. The proposed method estimates only $2n + 2n_A$ parameters where n_A is the order of an IIR model for the plant. An important note to make is that $n \geq n_A$ always and in many practical applications $n \gg n_A$.
- The algorithm does not need collecting and processing batches of data as apposed to prior work (e.g. Chandrasekar et al. (2006)).
- The broad–band noise model is explicitly considered in the algorithm and its parameters are identified.
- The algorithm is developed completely in discrete time domain and its convergence with probability one to actual parameters is proven using Ljung's method Ljung (1977a) for the analysis of recursive algorithms.
- The effectiveness of the algorithm is demonstrated by comprehensive simulation and experiments on an HDD servo system (through DSP implementation of the algorithm) that is subjected to 173 sinusoidal disturbances. Moreover, the performance in situations of abrupt and gradual time varying disturbance and plant dynamics is tested.

Notation: In the previous chapters we used the notation x_k to represent variable (sequence) x at time step k. Since many of the variables that are used in this chapter need a superscript and subscript index, we present time step k as an argument inside a pair of parenthesis. For instance, instead of ϕ_k and $\hat{\theta}_k$ we will be using $\phi(k)$, $\hat{\theta}(k)$, etc.

6.2 Mathematical Preliminaries

In the sequel we use a slightly different realization of the block diagram shown in Fig. 1.2. Suppose that the closed loop dynamics is realized by $R(q^{-1}) = \frac{B(q^{-1})}{A(q^{-1})}$ and the disturbances are represented by $\xi(k) = \frac{C(q^{-1})}{A(q^{-1})}\bar{w}(k)$ and $\bar{r}(k) = \frac{B(q^{-1})}{A(q^{-1})}\bar{d}(k)$. Using this new notation, the error signal in discrete time domain can be represented as a function of the closed loop

dynamics, input signals and disturbances

$$e(k) = \frac{B(q^{-1})}{A(q^{-1})}\left(u(k) + u_A(k)\right) + \frac{C(q^{-1})}{A(q^{-1})}\bar{w}(k) + \bar{r}(k) \tag{6.1}$$

where $u(k)$ is an exogenous excitation signal and $\bar{w}(k)$ is an unmeasurable wide-sense stationary sequence of independent random vectors with finite moments. We assume that the nominal feedback controller is able to stabilize the open loop plant, which implies that $A(q^{-1})$ is exponentially stable, i.e. $A(p)$ has all roots (strictly) outside the unit circle.

Similarly to the previous chapters, let ω_i for $i = 1, \cdots, n$ denote the frequency of sinusoidal components in $\bar{r}$, and assume that the disturbance signal can be factorized as

$$\bar{r}(k) = \theta_{\bar{R}}^T \phi_R(k) \tag{6.2}$$

where

$$\begin{aligned} \theta_{\bar{R}}^T &:= \left[\bar{\alpha}_1, \bar{\beta}_1, \ldots, \bar{\alpha}_n, \bar{\beta}_n\right] \\ \phi_R^T(k) &:= \left[\sin(\omega_1 kT), \cos(\omega_1 kT), \ldots, \sin(\omega_n kT), \cos(\omega_n kT)\right] \end{aligned} \tag{6.3}$$

The following lemma will be used for steady state analysis.

Lemma 3
Consider $\bar{r}(k)$ as a general periodic signal and $L(q^{-1})$ as a discrete-time linear system. The steady state response $\tilde{r}(k) := L(q^{-1})[\bar{r}(k)]$ is periodic. Moreover, when $\bar{r}(k)$ is a linear combination of sinusoidal signals factorized similarly to (6.2), $\tilde{r}(k)$ (in steady state) consists of sinusoidals with the same frequencies

$$\begin{aligned} \tilde{r}(k) &= L(q^{-1})\left[\theta_{\bar{R}}^T \phi_R(k)\right] \\ &= \theta_{\bar{R}}^T L(q^{-1})\left[\phi_R(k)\right] \\ &= \theta_{\bar{R}}^T \phi_{R_L}(k) \end{aligned}$$

where

$$\begin{aligned} \phi_{R_L}^T(k) := [&m_{L_1}\sin(\omega_1 kT + \delta_{L_1}), m_{L_1}\cos(\omega_1 kT + \delta_{L_1}), \ldots \\ &m_{L_n}\sin(\omega_n kT + \delta_{L_n}), m_{L_n}\cos(\omega_n kT + \delta_{L_n})]\,. \end{aligned} \tag{6.4}$$

Here, m_{Li} and δ_{Li} are the magnitude and phase of $L(e^{-j\omega_i})$. Since

$$\begin{bmatrix} m_{Li}\sin(\omega_i kT + \delta_{Li}) \\ m_{Li}\cos(\omega_i kT + \delta_{Li}) \end{bmatrix} = \underbrace{\begin{bmatrix} m_{Li}\cos(\delta_{Li}) & m_{Li}\sin(\delta_{Li}) \\ -m_{Li}\sin(\delta_{Li}) & m_{Li}\cos(\delta_{Li}) \end{bmatrix}}_{D_{Li}} \begin{bmatrix} \sin(\omega_i kT) \\ \cos(\omega_i kT) \end{bmatrix} \tag{6.5}$$

$\phi_R(k)$ can be transformed to $\phi_{R_L}(k)$ by a linear transformation

$$\phi_{R_L}(k) = \underbrace{\begin{bmatrix} D_{L1} & 0 & \cdots & 0 \\ 0 & D_{L2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & D_{Ln} \end{bmatrix}}_{D_L} \phi_R(k). \tag{6.6}$$

Moreover, we can apply the transformation to the response coefficients rather than the sinusoid vector

$$\tilde{r}(k) = \left(D_L^T \theta_{\bar{R}}\right)^T \phi_R$$

Proof

Refer to Kamen and Heck (2000) for a general formula for the steady-state sinusoidal response of a linear time-invariant system. □

6.3 Proposed Direct Adaptive Control

Although a real dynamic system cannot be exactly described by finite order polynomials in (6.7), in most of applications A, B and C can be determined such that they give a finite vector difference equation that describes the recorded data as well as possible, i.e.

$$\begin{aligned} A(q^{-1}) &:= 1 + a_1 q^{-1} + a_2 q^{-2} + \cdots + a_{n_A} q^{-n_A} \\ B(q^{-1}) &:= b_1 q^{-1} + b_2 q^{-2} + \cdots + b_{n_B} q^{-n_B} \\ C(q^{-1}) &:= 1 + c_1 q^{-1} + c_2 q^{-2} + \cdots + c_{n_C} q^{-n_C}. \end{aligned}$$

Without loss of generality, we assume that $n_A = n_C$ and the relative degree of the primary path transfer function is 1 which implies that $n_A = n_B$. Let $A^*(q^{-1}) := 1 - A(q^{-1})$, then the error is given by

$$e(k) = A^*(q^{-1})e(k) + B(q^{-1})\left(u(k) + u_A(k)\right) + C(q^{-1})\bar{w}(k) + A(q^{-1})\bar{r}(k) \tag{6.7}$$

This equation can be represented purely in discrete time domain in a vector form

$$e(k) = \theta_A^T \phi_e(k) + \theta_B^T \left(\phi_u(k) + \phi_{u_A}(k)\right) + \theta_C^T \phi_{\bar{w}}(k) + \tilde{r}(k) + \bar{w}(k) \tag{6.8}$$

where

$$\begin{aligned} \theta_A^T &:= \begin{bmatrix} -a_1, & -a_2, & \cdots, & -a_{n_A} \end{bmatrix} \\ \theta_B^T &:= \begin{bmatrix} b_1, & b_2, & \cdots, & b_{n_A} \end{bmatrix} \\ \theta_C^T &:= \begin{bmatrix} c_1, & c_2, & \cdots, & c_{n_A} \end{bmatrix} \\ \phi_e^T(k) &:= \begin{bmatrix} e(k-1), & e(k-2), & \cdots, & e(k-n_A) \end{bmatrix} \\ \phi_u^T(k) &:= \begin{bmatrix} u(k-1), & u(k-2), & \cdots, & u(k-n_A) \end{bmatrix} \\ \phi_{u_A}^T(k) &:= \begin{bmatrix} u_A(k-1), & u_A(k-2), & \cdots, & u_A(k-n_A) \end{bmatrix} \\ \phi_{\bar{w}}^T(k) &:= \begin{bmatrix} \bar{w}(k-1), & \bar{w}(k-2), & \cdots, & \bar{w}(k-n_A) \end{bmatrix}. \end{aligned} \tag{6.9}$$

and $\tilde{r}(k) := A(q^{-1})\bar{r}(k)$. Note that two regressors, respectively denoted by $\phi_u(k)$ and $\phi_{u_A}(k)$, are considered for the excitation signal $u(k)$ and the adaptive control $u_A(k)$ separately, in spite of the fact that they could be combined into a single regressor. The rationale behind this consideration will be explained later after (6.17). Since disturbance $\bar{r}(k)$ is periodic and $A(q^{-1})$ is a stable filter – i.e. it operates as an FIR filter – the response $\tilde{r}(k)$ is also periodic by Lemma 3.

$$\begin{aligned}\tilde{r}(k) &= A(q^{-1})\left[\theta_{\bar{R}}^T\phi_R(k)\right]\\ &= \theta_{\bar{R}}^T A(q^{-1})\left[\phi_R(k)\right]\\ &= \theta_{\bar{R}}^T\phi_{R_A}(k)\end{aligned}$$

where

$$\phi_{R_A}(k) = \underbrace{\begin{bmatrix} D_{A1} & 0 & \cdots & 0\\ 0 & D_{A2} & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & D_{An}\end{bmatrix}}_{D_A}\phi_R(k). \tag{6.10}$$

Accordingly, $\tilde{r}(k)$ can be represented using the same regressor vector, $\phi_R(k)$

$$\begin{aligned}\tilde{r}(k) &= \theta_{\bar{R}}^T\phi_{R_A}(k)\\ &= \underbrace{\theta_{\bar{R}}^T D_A}_{\theta_R^T}\phi_R(k)\\ &= \theta_R^T\phi_R(k).\end{aligned}$$

Substituting this expression in (6.8) yields

$$e(k) = \theta_A^T\phi_e(k) + \theta_B^T\left(\phi_u(k) + \phi_{u_A}(k)\right) + \theta_C^T\phi_{\bar{w}}(k) + \theta_R^T\phi_R(k) + \bar{w}(k). \tag{6.11}$$

Equation (6.7) shows that an ideal control signal $u_A^*(k)$ should satisfy

$$B(q^{-1})u_A^*(k) + A(q^{-1})\bar{r}(k) = 0. \tag{6.12}$$

Again, since $B(q^{-1})$ and $A(q^{-1})$ are both LTI systems and $\bar{r}(k)$ contains only sinusoidal signals, the ideal control signal $u_A^*(k)$ has to have sinusoidal contents at the same frequencies. This motivates us to decompose the ideal control signal into

$$u_A^*(k) = \theta_D^T\phi_R(k).$$

By this representation of the control signal, our goal will be to estimate θ_D in an adaptive manner. We define the actual control signal as

$$u_A(k) = \hat{\theta}_D^T(k)\phi_R(k) \tag{6.13}$$

where $\hat{\theta}_D(k)$ is the vector of estimated parameters that should ideally converge to θ_D. As a result, the residual in (6.12) when θ_D is replaced by $\hat{\theta}_D$ is

$$B(q^{-1})u_A(k) + A(q^{-1})\bar{r}(k) = B(q^{-1})\left(\hat{\theta}_D^T(k)\phi_R(k)\right) + \theta_R^T\phi_R(k). \tag{6.14}$$

The following lemma, known as the *swapping lemma*, can be used to simplify this equation further.

Lemma 4 (Discrete time swapping lemma)
Let a stable and proper rational transfer function $H(z)$ have a minimal realization $H(z) = C^T(zI - A)^{-1}B + D$ and $\theta(k)$ and $\phi(k)$ be two vector signals. Then,

$$H(z)\left[\theta(k)^T\phi(k)\right] = \theta^T(k)\left\{H(z)\left[\phi(k)\right]\right\} + w(k)$$

where

$$\begin{aligned} w(k) :=& -H_C(z)\left[\left\{H_B(z)z\left[\phi(k)\right]\right\}^T\left\{(z-1)\left[\theta(k)\right]\right\}\right] \\ H_C(z) :=& C^T(zI - A)^{-1} \\ H_B(z) :=& (zI - A)^{-1}B \end{aligned}$$

Proof
Refer to the discrete–time swapping lemma in Tao (2003). □

We define a new parameter vector

$$\theta_M^T(k) := \hat{\theta}_D^T(k)D_B + \theta_R^T \tag{6.15}$$

where D_B is a matrix similar to D_A in (6.6), but its block diagonal terms are formed by the magnitude and phase of $B\left(e^{-j\omega_i}\right)$ rather than $A(e^{-j\omega_i})$. The vector $\theta_M(k)$ corresponds to the imperfection in control synthesis. Accordingly, it is called the *residual parameters* vector throughout this section. By substituting the result of Lemma 4 in (6.14) we have

$$\begin{aligned} B(q^{-1})u_A(k) + A(q^{-1})\bar{r}(k) &= \hat{\theta}_D^T(k)\left(B(q^{-1})\phi_R(k)\right) + \theta_R^T\phi_R(k) + w_t(k) \\ &= \hat{\theta}_D^T(k)D_B\phi_R(k) + \theta_R^T\phi_R(k) + w_t(k) \\ &= \left(\hat{\theta}_D^T(k)D_B + \theta_R^T\right)\phi_R(k) + w_t(k) \\ &= \theta_M^T(k)\phi_R(k) + w_t(k) \end{aligned} \tag{6.16}$$

where the term $\theta_M^T(k)\phi_R(k)$ corresponds to the residual error at the compensation frequencies. The term $w_t(k)$ represents the transient excitation caused by the variation of $\hat{\theta}_D(k)$ over time and its expression is given by the swapping lemma. As a result, the term

$$\theta_B^T\phi_{u_A}(k) + \theta_R^T\phi_R(k)$$

in (6.11) can be replaced by (6.16)

$$e(k) = \theta_A^T\phi_e(k) + \theta_B^T\phi_u(k) + \theta_C^T\phi_{\bar{w}}(k) + \theta_M^T(k)\phi_R(k) + w_t(k) + \bar{w}(k). \tag{6.17}$$

Remark 1
The reason behind choosing two separate regressors for $u(k)$ and $u_A(k)$, as remarked earlier, is that the recent substitution in the above equation is not feasible if the two regressors were combined into a single regressor.

6.3.1 Parameter Adaptation Algorithm

The error dynamics shows that the information obtained from measurements cannot be directly used to estimate $\hat{\theta}_D$ as long as the closed loop system and noise dynamics are unknown. We propose an adaptive algorithm in this section that accomplishes the estimation of the closed loop system and noise dynamics in conjunction with the control synthesis.

Let $\hat{\theta}_A$, $\hat{\theta}_B$, $\hat{\theta}_C$ and $\hat{\theta}_M$ be the estimated parameters analogous to (6.17). We denote the *a–priori* estimate of the error signal at time k based on the estimates at $k-1$ as $\hat{y}^\circ(k)$ and define it as

$$\hat{y}^\circ(k) = \hat{\theta}_A^T(k-1)\phi_e(k) + \hat{\theta}_B^T(k-1)\phi_u(k) + \hat{\theta}_C^T(k-1)\phi_\epsilon(k) + \hat{\theta}_M^T(k-1)\phi_R(k) \tag{6.18}$$

Similarly, we define the *a–posteriori* estimate of the error signal at time k based on the estimates at k

$$\hat{y}(k) = \hat{\theta}_A^T(k)\phi_e(k) + \hat{\theta}_B^T(k)\phi_u(k) + \hat{\theta}_C^T(k)\phi_\epsilon(k) + \hat{\theta}_M^T(k)\phi_R(k)$$

and accordingly, the *a–priori* and *a–posteriori* estimation errors are defined as

$$\begin{aligned} \epsilon^\circ(k) &:= e(k) - \hat{y}^\circ(k) \\ \epsilon(k) &:= e(k) - \hat{y}(k) \end{aligned} \tag{6.19}$$

The regressor vector ϕ_ϵ in (6.18) contains the past values of the *a–posteriori* error

$$\phi_\epsilon^T(k) := \left[\epsilon(k-1),\, \epsilon(k-2),\, \ldots,\, \epsilon(k-n_A)\right]. \tag{6.20}$$

Assume that the estimates at time $k=0$ are initialized by either zero or some "good" values when prior knowledge about the system dynamics or disturbance is available. We propose the following adaptation algorithm for updating the estimated parameters

$$\begin{bmatrix} \hat{\theta}_A(k) \\ \hat{\theta}_B(k) \\ \hat{\theta}_C(k) \\ \hat{\theta}_M(k) \end{bmatrix} = \begin{bmatrix} \hat{\theta}_A(k-1) \\ \hat{\theta}_B(k-1) \\ \hat{\theta}_C(k-1) \\ \hat{\theta}_M(k-1) \end{bmatrix} + \begin{bmatrix} \gamma_1(k)F^{-1}(k) & 0 \\ 0 & \gamma_2(k)f^{-1}(k)I \end{bmatrix} \begin{bmatrix} \phi_e(k) \\ \phi_u(k) \\ \phi_\epsilon(k) \\ \phi_R(k) \end{bmatrix} \epsilon(k). \tag{6.21}$$

$F(k)$ is a positive (semi)definite matrix with proper dimensions and $f(k)$ is a positive scalar. These gains, which are usually known as learning factor or step size, can be updated via either recursive least squares algorithms, least mean squares type methods or a combination

of them. We use recursive least squares for the plant and noise dynamics parameters since the number of coefficients is usually "small". On the other hand, for large n the recursive least squares algorithm requires major computations. Therefore, it is of interest to reduce the computations, possibly at the price of slower convergence, by replacing the recursive least squares update law by the stochastic gradient method. It is well known that the step size of adaptive algorithms in stochastic environments should converge to zero or to very small values to avoid the "excess error" caused by parameter variations due to noises. Therefore, we consider positive real value decreasing scalar sequences $\gamma_1(k)$ and $\gamma_2(k)$ conjointly with the step sizes. More explicitly, the update rules for F and f are

$$F(k) = F(k-1) + \gamma_1(k) \left(\begin{bmatrix} \phi_e(k) \\ \phi_u(k) \\ \phi_\epsilon(k) \end{bmatrix} \begin{bmatrix} \phi_e(k) \\ \phi_u(k) \\ \phi_\epsilon(k) \end{bmatrix}^T - F(k-1) \right)$$
$$f(k) = f(k-1) + \gamma_2(k) \left(\phi_R^T(k)\phi_R(k) - f(k-1) \right).$$

Remark 2
$\phi_u(k)$ should be persistently exciting of order $2n_A$ in order to guarantee that $F(k)$ is non–singular and (6.21) is not susceptible to numerical problems. A possible choice of excitation signal that satisfies this condition when $n \geq n_A/2$ is the exogenous excitation signal proposed in 4.3. It is clear that $f(k)$ is not susceptible to this issue since $\phi_R^T(k)\phi_R(k)$ is always strictly positive.

Remark 3
The update rule (6.21) is not suitable for implementation since the estimates at time k depends on $\epsilon(k)$ and on the other hand based on (6.19) $\epsilon(k)$ depends on the estimates at step k. However, since the update rule is known, the a–posteriori *estimation error can be related to the* a–priori *error which relies only on the estimates at $k-1$. The parameter adaptation algorithms that follows this causality is given in (6.22).*

$$\begin{aligned} \epsilon(k) :&= e(k) - \hat{y}(k) \\ &= \epsilon^\circ(k) - (\hat{y}(k) - \hat{y}^\circ(k)) \\ &= \epsilon^\circ(k) - \left(\left[\begin{array}{c} \phi_e(k) \\ \phi_u(k) \\ \hdashline \phi_\epsilon(k) \\ \hdashline \bar{\phi}_R(k) \end{array} \right]^T \left[\begin{array}{c:c} \gamma_1(k)F^{-1}(k) & 0 \\ \hdashline 0 & \gamma_2(k)f^{-1}(k)I \end{array} \right] \left[\begin{array}{c} \phi_e(k) \\ \phi_u(k) \\ \hdashline \phi_\epsilon(k) \\ \hdashline \bar{\phi}_R(k) \end{array} \right] \right) \epsilon(k) \Rightarrow \end{aligned} \tag{6.22}$$

$$\epsilon(k) = \frac{\epsilon^\circ(k)}{1 + \left[\begin{array}{c} \phi_e(k) \\ \phi_u(k) \\ \hdashline \phi_\epsilon(k) \\ \hdashline \bar{\phi}_R(k) \end{array} \right]^T \left[\begin{array}{c:c} \gamma_1(k)F^{-1}(k) & 0 \\ \hdashline 0 & \gamma_2(k)f^{-1}(k)I \end{array} \right] \left[\begin{array}{c} \phi_e(k) \\ \phi_u(k) \\ \hdashline \phi_\epsilon(k) \\ \hdashline \bar{\phi}_R(k) \end{array} \right]}$$

Suppose that the parameter vector $\theta_M(k)$ and response matrix D_B are known at time step k. Then, a possible update rule that satisfies (6.12) would be

$$\begin{aligned} 0 &= \hat{\theta}_D^T(k+1)D_B + \theta_R^T \\ &= \hat{\theta}_D^T(k+1)D_B + \theta_M^T(k) - \hat{\theta}_D^T(k)D_B \Rightarrow \\ \hat{\theta}_D^T(k+1) &= \hat{\theta}_D^T(k) - \theta_M^T(k)D_B^{-1}. \end{aligned} \tag{6.23}$$

Here, we have used the fact that D_B is a block diagonal combination of scaled rotation matrices, which implies that it is full rank and invertible. This is an infeasible update rule since neither $\theta_M(k)$, nor D_B is known. We replace these variables by their respective estimated values and use a small step size α in order to avoid large transient and excess error

$$\hat{\theta}_D^T(k+1) = \hat{\theta}_D^T(k) - \alpha\hat{\theta}_M^T(k)\hat{D}_B^{-1}(k).$$

Note that this update rule works as a first order system that has a pole at 1. In order to robustify this difference equation we alternatively propose using a *Ridge* solution for (6.23). More formally, we are interested in minimizing the instantaneous cost function

$$J_c(k) := \frac{1}{2}\|\hat{\theta}_D^T(k) + \theta_R^T D_B^{-1}\|_2^2 + \frac{1}{2}\lambda\|\hat{\theta}_D^T(k)\|_2^2$$

where λ is a (positive) weight for the penalization term. We use a gradient descent algorithm to recursively update $\hat{\theta}_D$. Let $\beta = 1 - \alpha\lambda$ and the gradient of $J_c(k)$ with respect to $\hat{\theta}_D^T(k)$ be denoted by

$$\frac{\partial J_c(k)}{\partial \hat{\theta}_D^T(k)} = \left(\hat{\theta}_D^T(k) + \theta_M^T(k)D_B^{-1} - \hat{\theta}_D^T(k)\right) + \lambda\hat{\theta}_D^T(k).$$

Since the actual values of θ_M and D_B are unknown, we use the estimates and define the gradient descent update rule for $\hat{\theta}_D$ as

$$\begin{aligned} \hat{\theta}_D^T(k+1) &= \hat{\theta}_D^T(k) - \alpha\frac{\partial J_c(k)}{\partial \hat{\theta}_D^T(k)} \\ &= \beta\hat{\theta}_D^T(k) - \alpha\hat{\theta}_M^T(k)\hat{D}_B^{-1}(k). \end{aligned} \tag{6.24}$$

This expression yields that a positive value of β that is less than 1 results in a bounded value of $\hat{\theta}_D$ in steady state as long as $\hat{\theta}_M^T\hat{D}_B^{-1}$ stays bounded. Moreover, assuming that $\hat{\theta}_M$ and $\hat{D}_B$ converge to the actual θ_M and D_B – which will be proved later – the steady state residue is

$$\begin{aligned} \lim_{k\to\infty}\theta_M(k) &= D_B^T \lim_{k\to\infty}\hat{\theta}_D(k) + \theta_R \\ &= \frac{-\alpha}{1-\beta}\lim_{k\to\infty}\theta_M(k) + \theta_R \\ &= \frac{1-\beta}{1-\beta+\alpha}\theta_R. \end{aligned}$$

This expression shows that there is a compromise between the steady state attenuation level and robustness, and in order to achieve both, the two gains should be chosen such that

$$0 < \alpha \ll \beta < 1.$$

Now that we have an update law for $\hat{\theta}_D(k)$, we have a complete algorithm for synthesizing the control signal (repeated from (6.13))

$$u_A(k) = \hat{\theta}_D^T(k)\phi_R(k).$$

Theorem 5
*The residual error θ_M converges to $\frac{1-\beta}{1-\beta+\alpha}\theta_R$ with probability 1, the only equilibrium point of the closed loop system is stable in the sense of Lyapunov and it corresponds to $\hat{\theta}_A = \theta_A$, $\hat{\theta}_B = \theta_B$, $\hat{\theta}_C = \theta_C$ and $\hat{\theta}_M = \frac{1-\beta}{1-\beta+\alpha}\theta_R$ **if** the following conditions are satisfied:*

1. *$u(k)$ is persistently exciting of at least order $2n_A$.*
2. *$\sum_{k=1}^{\infty} \gamma(k) = \infty$ and $\gamma(k) \to 0$ as $k \to \infty$.*
3. *The estimated $\hat{\theta}_A(k)$ belongs to*

$$\mathscr{D}_A := \left\{ \hat{\theta}_A : 1 + \hat{a}_1 q + \cdots + \hat{a}_{n_A} q^{n_A} = 0 \Rightarrow |q| > 1 \right\}.$$

   infinitely often with probability one.
4. *The estimated $\hat{\theta}_B(k)$ always belongs to*

$$\mathscr{D}_B := \left\{ \hat{\theta}_B : 0 < |\hat{b}_1 e^{-j\omega_h} + \cdots + \hat{b}_{n_A} e^{-jn_A\omega_h}| < \frac{\alpha}{1-\beta} |b_1 e^{-j\omega_h} + \cdots + b_{n_A} e^{-jn_A\omega_h}| \right\}$$

   for all $h \in \{1, \ldots, n\}$.
5. *$\frac{1}{C(q^{-1})}$ is strictly positive real (SPR).*
6. *Real $\left(\frac{B(e^{-j\omega_h})/\hat{B}_k(e^{-j\omega_h})}{C(e^{-j\omega_h})} \right) > 0$ for all $h \in \{1, \ldots, n\}$ infinitely often with probability one.*

Proof
Only the sketch of proof is provided here and the details are left to the appendix. We use the method of analysis of stochastic recursive algorithms developed by Ljung (1977a) to study the convergence and asymptotic behavior of the proposed adaptive algorithm with update and control rules given in (6.21), (6.24) and (6.13). First, the parameter adaptation is outlined as an expression of the from

$$\hat{\Theta}(k) = \hat{\Theta}(k-1) + \gamma(k) Q(k; \hat{\Theta}(k-1), \Phi(k))$$

where $\hat{\Theta}(k)$ is a sequence of column vectors that represents all adaptive parameters, and $\Phi(k)$ is a column vector that contains the known and unknown regressors. Then, the dynamics of regressor is formalized as a time–varying linear system

$$\Phi(k) = A_\Phi(\hat{\Theta}_{k-1})\Phi(k-1) + B_\Phi(\hat{\Theta}_{k-1})U(k).$$

These two sets of equations present the closed loop dynamics. The key idea is that a differential equation counterpart for this closed loop system can be defined under some regularity conditions. Three sets of regularity conditions are proposed in Ljung (1977a) that target the analysis of deterministic and stationary stochastic processes. The problem under our study cannot be exactly outlined in these frameworks since the input signal consists of stochastic and deterministic parts and as a matter of fact it is a cyclostationary stochastic process. However, "Assumptions C" in Ljung (1977a) are adopted and generalized to this case with very minor modifications. In the appendix we show that these regularity conditions are satisfied when the assumptions of theorem 5 hold true. This implies that the only convergence point of the system is the stable equilibrium of the differential equation counterpart. This equilibrium point corresponds to the actual values of plant and noise parameters. It will be shown that the estimated parameters will converge with probability one to this equilibrium point which results in $\theta_M = \frac{1-\beta}{1-\beta+\alpha}\theta_R$. □

Remark 4

Assumption 2 can be satisfied by a broad range of gain sequences $\gamma(k)$. For instance, both regularity conditions hold for $\gamma(k) = \frac{C}{k^\alpha}$ when $0 < C < \infty$ and $0 < \alpha \leq 1$.

Remark 5

Assumption 3 requires monitoring the poles of $\hat{A}(q^{-1})$ polynomial. This is a common issue in adaptive control and several methods have been proposed. For instance, the estimates can be projected to the interior of $\mathscr{D}_A$ whenever the poles fall out of (or on) the unit circle. Assumption 4 requires monitoring the magnitude of polynomial $B(e^{-j\omega_h})$ for all compensation frequencies. The left inequality guarantees that $\hat{D}_B$ is always invertible. The right inequality requires some very rough knowledge about the plant magnitude because the term $\alpha/(1-\beta)$ is large. Both inequalities can be satisfied by projecting the estimates into the interior of $\mathscr{D}_B$ whenever they do not belong to $\mathscr{D}_B$.

Assumption 5 depends on the noise dynamics and may not hold true in some applications. In this case, the update rule (6.21) can be modified by using an estimate of $C(q^{-1})$, say $\bar{C}(q^{-1})$. Define a new sequence ϵ_c as

$$\epsilon_c(k) = \bar{C}(q^{-1})\left[\epsilon(k)\right].$$

The new update rule is

$$\begin{bmatrix} \hat{\theta}_A(k) \\ \hat{\theta}_B(k) \\ \hat{\theta}_C(k) \\ \hat{\bar{\theta}}_M(k) \end{bmatrix} = \begin{bmatrix} \hat{\theta}_A(k-1) \\ \hat{\theta}_B(k-1) \\ \hat{\theta}_C(k-1) \\ \hat{\bar{\theta}}_M(k-1) \end{bmatrix} \begin{bmatrix} \gamma_1(k)F^{-1}(k) & 0 \\ 0 & \gamma_2(k)\bar{f}^{-1}(k)I \end{bmatrix} \begin{bmatrix} \phi_e(k) \\ \phi_u(k) \\ \phi_\epsilon(k) \\ \bar{\phi}_R(k) \end{bmatrix} \epsilon_c(k). \tag{6.25}$$

Theorem 6
The results of theorem 5 hold true when (6.21) is replaced with (6.25) and assumptions 5 and 6 are substituted by the following two assumptions

7. $\frac{\bar{C}(q^{-1})}{C(q^{-1})}$ *is strictly positive real.*

8. $Real\left(\frac{B^i(e^{-j\omega_h})/\hat{B}^i_k(e^{-j\omega_h})}{C(e^{-j\omega_h})/\bar{C}(e^{-j\omega_h})}\right) > 0$ *for all* $i \in \{1,\ldots,n_i\}$ *and* $h \in \{1,\ldots,n\}$ *infinitely often with probability one.*

Proof
Refer to the appendix. □

Assumption 6 (or assumption 8 when (6.25) is used) is in general difficult to verify since the $B(.)$ polynomial is unknown. However, based on theorem 5 the equilibrium point of the closed loop system satisfies this assumption when assumption 6 (assumption 8) is satisfied. It is shown in the proof that one of the parameters that determine the domain of attraction of the equilibrium point is the excitation sequence $u(k)$ intensity. In the sequel we propose a practical consideration to enlarge the domain of attraction by a frequency–shaped excitation sequence with an adaptive gain.

6.3.2 Excitation Signal Gain

Earlier, in section 4.3, we proposed a special exogenous excitation signal that is specifically tailored for the problem under our study. We propose a variable gain $\alpha^u(k)$ for this excitation (see (4.13) for its definition) that adjusts itself based on the error residual. More explicitly, this gain is based on the norm of $\hat{\theta}_M(k)$ because this vector is an estimate of the residual error. The residuals are large at the beginning and more excitation is beneficial to decrease model inaccuracy quickly. On the other hand, as the algorithm converges, $\hat{\theta}_M$ converges towards zero and the portion of error caused by the excitation signal is desired to be as small as possible. We propose a variable excitation gain based on exponential smoothing of the residual norm

$$\alpha^u(k) = (1-\beta_u)\alpha^u(k-1) + \alpha_{dc}\beta_u\|\hat{\theta}_M(k)\|_2. \tag{6.26}$$

Here, $0 < \beta_u \leq 1$ and α_{dc} are design parameters that determine the bandwidth and DC gain of the above first order system.

Theorem 7
Suppose that (6.24), (6.25) and (6.13) are used to define the adaptive controller, and (4.13) and (6.26) are deployed to structure the excitation signal. By choosing $\alpha_{dc} > 0$ *large enough,* θ_M *converges to* $\frac{1-\beta}{1-\beta+\alpha}\theta_R$ *with probability 1 if the following conditions are satisfied.*

9. *Assumptions 1, 2, 3, 4 and 7 hold true.*

10. $\|\hat{\theta}_M(k)\|_2 \leq \gamma_{\max}$ *for some positive and bounded* $\gamma_{\max}$.

Moreover, the only equilibrium point of the closed loop system is stable in the sense of Lyapunov and it corresponds to $\hat{\theta}_A = \theta_A$, $\hat{\theta}_B = \theta_B$, $\hat{\theta}_C = \theta_C$ *and* $\hat{\theta}_M = \frac{1-\beta}{1-\beta+\alpha}\theta_R$.

Proof
Refer to the appendix. □

The intuition behind theorem 7 is that by keeping $\hat{\theta}_M$ bounded and increasing the excitation power the domain of attraction of the equilibrium point defined in theorem 5 enlarges and by choosing α_{dc} large enough, assumption 8 becomes true.

6.4 Practical Aspects

6.4.1 $\hat{D}_B$ Inversion

Matrix $\hat{D}_B$ is an estimate of D_B that can be defined analogously to D_A in (6.6) based on the magnitude and phase of $\hat{B}(e^{-j\omega_i})$. Its inverse matrix that is used in (6.24) can be calculated easily by inverting the block diagonal terms which are scaled rotation matrices similar to D_{Ai} in (6.5). This operation involves inverting the estimated magnitudes that might be very small during the transient interval, especially when $\hat{\theta}_B$ is initialized by zeros. In that case, any small fluctuation of $\hat{\theta}_B$ can cause large transient error. One possible solution is to use small and slowly decreasing step sizes in the parameter adaptation algorithm, which is not desirable since its side effect is decreasing the convergence rate. We take an alternative approach, which is to use exponential smoothing to relax the transient errors. Let $\bar{m}_{Bi}(k)$ and $\bar{\delta}_{Bi}(k)$ be the magnitude and phase of $\hat{B}(e^{-j\omega_i})$ at step k. The smoothed values that form the scaled rotation matrices as in (6.5) are defined as

$$\begin{aligned} \hat{m}_{Bi}(k) &= (1-\alpha_B)\hat{m}_{Bi}(k-1) + \alpha_B \bar{m}_{Bi}(k) \\ \hat{\delta}_{Bi}(k) &= (1-\alpha_B)\hat{\delta}_{Bi}(k-1) + \alpha_B \bar{\delta}_{Bi}(k) \end{aligned} \tag{6.27}$$

where α_B is a positive scalar less than 1.

6.4.2 Spectrum Partitioning

Let Ω be the set of all compensation frequencies

$$\Omega := \{\omega_i | 1 \leq i \leq n\}.$$

As mentioned earlier, there are two main constraints that prevent us from compensating all of these frequencies by a single controller module when n is large, ω_i's are spread and the system frequency response is complicated: (1) the embedded processor may not have enough computation power to perform one step update of all parameters in one sampling interval (2) a high order model (i.e. large n_A) may be needed to capture the dynamics of system at a broad frequency range.

Our objective is to divide Ω to a collection of subsets, run the adaptive controller on these regions subsequently and superpose the feed–forward control signal learned in each region in order to synthesize a general control signal that attenuates all narrow–band disturbances at frequencies contained in Ω. More explicitly, we look for

- a collection, say $C : \{\Omega_j | 1 \leq j \leq m\}$, of mutually disjoint non–empty sets whose union is Ω.
- each subset contains consecutive elements
- a low order model can be fit to the actual system dynamics in each region
- the number of subsets, m, is kept as small as possible
- the running time needed to perform the algorithm does not exceed the sampling time of the system, which means that, the size of subsets should be bounded by a particular number that depends on the available computational power.

An admissible partition depends on the complexity of system dynamics and the controller processing unit. We propose a heuristic algorithm inspired by divisive hierarchical clustering Rokach and Maimon (2005) in conjunction with a monitoring mechanism to perform the partitioning in an online fashion. Let

$$C(t) := \{\Omega_j(t) | 1 \leq j \leq m(t)\}$$

be the collection at iteration t and assume that subsets $\Omega_j(t)$ are ordered such that the largest frequency in $\Omega_{j1}(t)$ is strictly less than the smallest frequency in $\Omega_{j2}(t)$ if and only if $j1 < j2$. We use $\Omega_{j1}(t) < \Omega_{j2}(t)$ notation for this type of set ordering. The collection is initialized with only one cluster

$$\Omega_0(0) = \Omega.$$

Let $M_i(k) := \|\hat{\theta}_{M,2i-1}(k) + j\hat{\theta}_{M,2i}(k)\|$, where $\hat{\theta}_{M,i}(k)$ denotes the i-th element of $\hat{\theta}_M(k)$ and j is the imaginary unit only in this particular expression. Indeed, $M_i(k)$ is the magnitude of residue corresponding to ω_i at time step k. At iteration t the adaptive algorithm is applied on region $\Omega_c(t)$ for N steps and the $\mathcal{L}_\infty$ norm of $M^T(k) := [M_1(k), \cdots, M_n(k)]$ is monitored. Two cases are considered:

I. $\|M(k)\|_\infty$ is increased for a particular number of consequent steps, say n_{fail}, before reaching step N: it indicates divergence of residues and the adaptive algorithm will be terminated. In this case of early termination, the current region $\Omega_c(t)$ will be divided to two regions $\Omega^1 < \Omega^2$. Ω^1 which is the first half of $\Omega_c(t)$ will be the region under verification in the next iteration and Ω^2 will be merged to the next region. That is

$$\begin{aligned} \Omega_{c+1}(t) &\leftarrow \Omega_{c+1}(t) \cup \Omega^2 \\ C(t+1) &\leftarrow (C(t) - \{\Omega_c(t)\}) \cup \{\Omega^1\}. \end{aligned} \tag{6.28}$$

Note that in this case c is not updated. Therefore, $\Omega_c(t+1) = \Omega^1$.

II. Otherwise: in this case no divergence is indicated. Therefore, $\Omega_c(t)$ is an admissible region and we need to check the next region in the next iteration

$$c \leftarrow c + 1.$$

The pseudo–code for this algorithm is given in Algorithm 2. `AdaptiveControl`$(\Omega_c(t))$ refers to applying the proposed adaptive control algorithm to region $\Omega_c(t)$. Function `Divide(.)` in the algorithm splits $\Omega_c(t)$ to Ω^1 and Ω^2 such that $\Omega^1 \neq \emptyset$, $\Omega^2 \neq \emptyset$, $\Omega^1 < \Omega^2$, $\Omega^1 \cap \Omega^2 = \emptyset$ and $\Omega^1 \cup \Omega^2 = \Omega_c(t)$. A natural solution can be obtained by simply halving $\Omega_c(t)$ into almost equal segments.

Algorithm 2 Spectrum Partitioning

procedure SPECTRUMPARTITIONING
  Initialize: $t \leftarrow 0,\ c \leftarrow 0,\ \Omega_0(0) = \Omega,\ C(0) = \{\Omega_0(0)\}$
  while $c <$ `length`$(C(t))$ **do** ▷ last region not verified yet
    $[Divergenc, u_n] \leftarrow$ `AdaptiveControl`$(\Omega_c(t), u_A)$ ▷ verify $\Omega_c(t)$
    if $Divergence = True$ **then** ▷ $\Omega_c(t)$ is too wide
      $[\Omega^1, \Omega^2] \leftarrow$ `Divide`$(\Omega_c(t))$ ▷ e.g. halve it
      $\Omega_{c+1}(t) \leftarrow \Omega_{c+1}(t) \cup \Omega^2$ ▷ merge Ω^2 to the next region
      $C(t+1) \leftarrow (C(t) - \{\Omega_c(t)\}) \cup \{\Omega^1\}$ ▷ replace $\Omega_c(t)$ with a smaller one
      $C(t+1) \leftarrow$ `Sort`$(C(t+1))$ ▷ $\Omega_i(t+1) < \Omega_j(t+1) \Leftrightarrow i < j$
    else ▷ $\Omega_c(t)$ is OK
      $u_A \leftarrow u_n$ ▷ update the feedforward control
      $c \leftarrow c + 1$ ▷ let's check the next region
    end if
    $t \leftarrow t + 1$ ▷ next iteration
  end while
end procedure

6.5 Comparison with Related Work

The harmonic steady–state (HSS) control algorithm is one of the simplest and most natural methods for rejection of sinusoidal disturbances. The method uses measurements of the steady–state response amplitude and phase to determine the required amplitude and phase of the control signal. This technique was developed independently within two research fields, namely *higher harmonic control* commonly used for helicopter vibration attenuation, and *convergent control* for active rotor balancing Lovera et al. (2003); Knospe et al. (1996); Friedmann and Millott (1995); Pearson and Goodall (1994).

HSS with Known Plant Frequency Response:

The simple form of HSS assumes that the plant frequency response is known and uses

$$u(k) = 0$$
$$u_A(k) = \hat{\theta}_D^T(k)\phi_R(k)$$

where $\hat{\theta}_D(k)$ is a vector of parameters that should be determined by the algorithm. The output signal when this control is applied is

$$e(k) = \frac{B(q^{-1})}{A(q^{-1})}\left(\hat{\theta}_D^T(k)\phi_R(k)\right) + \frac{C(q^{-1})}{A(q^{-1})}\bar{w}(k) + \theta_{\bar{R}}^T\phi_R(k). \tag{6.29}$$

The HSS control algorithm applies a fixed–rule control, i.e. fixed $\hat{\theta}_D$ over time batches and waits until the output approximately reaches steady state. Let j denote the j-th batch and any variable indexed by j be the corresponding value during that time interval. For instance,

$$e_{ss,j}(k) = \left[D_R^T\hat{\theta}_{D,j} + \theta_{\bar{R}}\right]^T\phi_R(k) + \frac{C(q^{-1})}{A(q^{-1})}\bar{w}(k) \tag{6.30}$$

refers to the steady state error in the j-th batch. The definition of D_R should be clear from (6.5) and (6.6). The term $\hat{\theta}_{D,j}^T D_R + \theta_{\bar{R}}^T$ is then extracted using a Fourier–like averaging

$$e_a(j) := D_R^T\hat{\theta}_{D,j} + \theta_{\bar{R}} = \underset{\text{batch } j}{\text{AVG}}\left[2\phi_R(k)e_{ss,j}(k)\right]. \tag{6.31}$$

Note that the averaging operation transforms the original problem to a very simple linear system that is defined in the iteration (batch) space – i.e. indexed by j rather than k. Therefore, we can think of the variable $e_a(j)$ as the state and $\hat{\theta}_{D,j}$ as the input of this system. The minimization of e_a in HSS control is then formulated as a simple quadratic regulation problem with the following cost function

$$J(e_a(j), \hat{\theta}_{D,j}) := e_a^T(j)Qe_a(j) + 2e_a^T(j)S\hat{\theta}_{D,j} + \hat{\theta}_{D,j}^T M\hat{\theta}_{D,j} \tag{6.32}$$

where $\begin{bmatrix} Q & S \\ S^T & M \end{bmatrix}$ is positive definite.

This algorithm is easy to understand, implement and apply to systems with known frequency responses. However, it suffers from slow convergence since it relies on averaging over batches of data. The next section presents a more involved version that does not require prior information about the plant frequency response since it can identify the response recursively.

Adaptive HSS Control:

The adaptive HSS control attempts to find the optimal control law for (6.32) when the system frequency response, i.e. D_R, is unknown. The disturbance parameters $\theta_{\bar{R}}$ can be

omitted from (6.31) by numerical differentiation

$$\begin{aligned}\tilde{e}(j) :&= e_a(j) - e_a(j-1)\\ &= D_R^T\left(\hat{\theta}_{D,j} - \hat{\theta}_{D,j-1}\right)\\ &= D_R^T\tilde{\theta}(j).\end{aligned} \tag{6.33}$$

Since both $\tilde{e}(j)$ and $\tilde{\theta}(j)$ are known, a recursive least squares algorithm – or in general any algorithm that can solve the above linear regression problem – can be used to obtain an estimate of D_R Chandrasekar et al. (2006).

Similar to the simple HSS algorithm, this method suffers from slow convergence since it uses averaging over batches of data. Moreover, as will be discussed in the following section, a further consideration is required to guarantee that the estimates converge to "good" values.

Adaptive HSS Control with Persistent Excitation:

Since $\tilde{\theta}(j)$ in (6.33) is generated by the adaptive controller, it may not be persistently exciting. Therefore, the estimated values do not necessarily converge to the real value of D_R and there is no guarantee that the adaptive control performs any disturbance attenuation. Chandrasekar et al. (2006) proposes adding an exogenous excitation to the adaptive control to guarantee enough persistence of excitation

$$\hat{\theta}_{D,j+1} = -G(j)\left(e_a(j) - \hat{D}_R^T(j)\hat{\theta}_{D,j}\right) + \tilde{\delta}(j). \tag{6.34}$$

The vector $\tilde{\delta}(j)$ is $2n$-periodic and at each iteration j contains only one non–zero value (similar to the impulse train).

Unlike the previous methods, this algorithm does not suffer from system uncertainties. However, the slow convergence is still a drawback.

Modified Adaptive HSS – Scheme 1:

In Pigg and Bodson (2006) an algorithm that enables simultaneous estimation of the plant frequency response and the control signal is proposed. The algorithm is sketched for continuous time systems, but one can show that an analogous method can be derived for discrete time systems. The key idea in this algorithm is that (6.31) can be restructured such that the unknown plant and disturbance parameters are included in one unknown vector. We abuse the notation and denote $e_a(t)$ as the output of the averaging mechanism at time t – previously $e_a(j)$ was used for the mean value over batch j. The first two elements of $e_a(t)$

are

$$\begin{bmatrix} e_{a,1}(t) \\ e_{a,2}(t) \end{bmatrix} \approx \begin{bmatrix} D_{R1,11} & -D_{R1,12} \\ D_{R1,12} & D_{R1,11} \end{bmatrix} \begin{bmatrix} \hat{\theta}_{D,t,1} \\ \hat{\theta}_{D,t,2} \end{bmatrix} + \begin{bmatrix} \theta_{\bar{R},1} \\ \theta_{\bar{R},2} \end{bmatrix}$$
$$= \underbrace{\begin{bmatrix} \hat{\theta}_{D,t,1} & -\hat{\theta}_{D,t,2} & 1 & 0 \\ \hat{\theta}_{D,t,2} & \hat{\theta}_{D,t,1} & 0 & 1 \end{bmatrix}}_{w_1^T(t)} \underbrace{\begin{bmatrix} D_{R1,11} \\ D_{R1,12} \\ \theta_{\bar{R},1} \\ \theta_{\bar{R},2} \end{bmatrix}}_{x_1}$$

where $e_{a,i}$ and $\theta_{\bar{R},i}$ and $\hat{\theta}_{D,t,i}$ denote the i-th element of e_a and $\theta_{\bar{R}}$ and $\hat{\theta}_{D,t}$ respectively. Similarly, $D_{R1,mk}$ refers to element (m,k) in D_{R1} (c.f. (6.6) for the definition of D_{R1}). The same decomposition can be performed for other pairs of e_a elements that jointly characterize one of the frequency contents

$$e_a(t) \approx \underbrace{\begin{bmatrix} w_1^T(t) & 0 & \cdots & 0 \\ 0 & w_2^T(t) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_n^T(t) \end{bmatrix}}_{W^T(t)} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix}}_{X} \tag{6.35}$$
$$= W^T(t)X.$$

Note that this expression is similar to a linear system of equations with the known regressor $W(t)$ and the unknown vector of parameters X. A continuous time least squares algorithm with forgetting factor is proposed in Pigg and Bodson (2006) to obtain the least–squares–error estimates, say $\hat{X}(t)$, in an online fashion. Based on (6.31), perfect rejection is achieved when $\hat{\theta}_D(t) = -D_R^{-T}\theta_{\bar{R}}$. Once the vector X is estimated, an estimate for D_R and $\theta_{\bar{R}}$ can be extracted from $\hat{X}(t)$. The control signal is then generated by replacing the unknown values by the estimates

$$\hat{\theta}_D(t) = -\hat{D}_R^{-T}(t)\hat{\theta}_R(t). \tag{6.36}$$

Slow adaptation is a crucial requirement for this algorithm to justify the steady–state approximations. An averaging mechanism, in general, can reduce the effect of noises. However, since the parameters are constantly updated, the effect of noises on the performance and convergence requires further study. Moreover, even under gradual adaptation and no noise in the system, the convergence of estimates to real values is not guaranteed since the regressor matrix $W(t)$ is not necessarily persistently exciting of order $4n$.

Modified Adaptive HSS – Scheme 2:

Another continuous time adaptive algorithm that obtains online estimates of the plant frequency response and of the disturbance parameters is proposed by Pigg and Bodson Pigg

and Bodson (2010). The most important aspect that distinguishes this algorithm from the previous methods is that it does not rely on an averaging mechanism like (6.31) that extracts the term $\hat{\theta}_{D,j}^T D_R + \theta_{\bar{R}}^T$ from (6.30). Instead, it uses the raw output, as in (6.30), directly to identify the parameter vector X in (6.35). The crucial approximation is that $e(k)$ in (6.29) is approximately equal to the steady state error in (6.30) and the noises are negligible

$$e(k) \approx e_{ss}(k) \approx \left[\hat{\theta}_{D,j}^T D_R + \theta_{\bar{R}}^T\right] \phi_R(k).$$

We abuse the notation by using the same symbols for continuous time and discrete time signals. The former ones are indexed by t that refers to time while the latter ones are indexed by k that stands for the time step. Similar to other methods, the error term can be split into known and unknown parts. From (6.35) and (6.29) we have

$$\begin{aligned} e(t) &\approx \phi_R^T(t) \left[D_R \hat{\theta}_{D,j} + \theta_{\bar{R}}\right] && (6.37)\\ &= \phi_R^T \left[W^T(t) X\right] && (6.38)\\ &= \left[\phi_R^T W^T(t)\right] X. && (6.39) \end{aligned}$$

The first term on the right hand side

$$W(t)\phi_R = \begin{bmatrix} \sin(\omega_1 t)\hat{\theta}_{D,t,1} + \cos(\omega_1 t)\hat{\theta}_{D,t,2} \\ -\sin(\omega_1 t)\hat{\theta}_{D,t,2} + \cos(\omega_1 t)\hat{\theta}_{D,t,1} \\ \sin(\omega_1 t) \\ \cos(\omega_1 t) \\ \vdots \\ \sin(\omega_n t)\hat{\theta}_{D,t,n-1} + \cos(\omega_n t)\hat{\theta}_{D,t,n} \\ -\sin(\omega_n t)\hat{\theta}_{D,t,n} + \cos(\omega_n t)\hat{\theta}_{D,t,n-1} \\ \sin(\omega_n t) \\ \cos(\omega_n t) \end{bmatrix} \in \Re^{4n}$$

is the regressor (known) matrix. Having the error $e(t)$ and the regressor $W(t)\phi_R$, the parameter vector $X \in \Re^{4n}$ can be estimated by a recursive algorithm, such as recursive least squares. The estimate of X at each step is then used to extract $\hat{D}_R$ and $\hat{\theta}_R$ and the control update is generated by (6.36).

Unlike all previous methods, e_a was not involved in this method. Therefore, no averaging mechanism as in (6.31) is needed and the algorithm convergence – as reported in Pigg and Bodson (2010) – is considerably faster than aforementioned HSS algorithms.

The dynamic behavior of the algorithm is analyzed in Pigg and Bodson (2010) using averaging theory and it is shown that the $4n$-dimensional averaged system has a $2n$-dimensional equilibrium surface, which can be divided into stable and unstable subsets. Trajectories generally converge to a stable point of the equilibrium surface, implying that the disturbance is asymptotically canceled even if the true parameters of the system are not exactly determined.

6.5.1 Distinctions Between the Algorithm Presented in this Dissertation and Other Methods

The method proposed in this dissertation is applicable in applications that include time–varying system parameters and disturbances, a property that does not apply to simple HSSs and the algorithm proposed by Wu and Bodson (2004). Moreover, the proposed method does not require batches of data as in Chandrasekar et al. (2006) since the control parameters are updated continuously in the discrete time domain. The number of parameters in Chandrasekar et al. (2006) and Pigg and Bodson (2006), Pigg and Bodson (2010) is $4n$, where n is the number of frequencies. Under the same situation (i.e. ignoring the other noises), the proposed method estimates only $2n+2n_A$ parameters where n_A is the order of an IIR model for the plant. An important note to make is that $n \geq n_A$ always and in many practical applications $n \gg n_A$. In the latter case, the number of parameters in our proposed method is considerably less than the other methods. Roughly speaking, this usually happens when n is larger than 10 since in most of practical applications, especially mechatronic systems, a 10^{th} order IIR can fit very complicated dynamics. In our HDD nanopositioning example n is 173. We will show by experiments and simulation that the total number of parameters for this case is approximately 400 whereas other methods would require 692 parameters. In Pigg and Bodson (2006) and Pigg and Bodson (2010) the algorithm is derived in continuous time and then applied to a discrete time system via discretization. The new adaptive controller proposed in this work is completely developed in discrete time domain. The approach taken in Pigg and Bodson (2010) attempts to estimate the disturbance parameters and then compensate them, whereas our method generates the control signal directly based on the residual error. It is shown in Pigg and Bodson (2010) that the plant and disturbance parameters converge to an equilibrium surface of $2n$-dimensional and the estimates do not necessarily converge to the actual values. Therefore, in case of time–varying disturbances all $4n$ values need to be adapted continuously whereas in the proposed algorithm only the $2n$ parameters of residual error will be adjusted. Similarly, in case of time varying systems the only parameters that will be adapted in our method are the ones corresponding to the plant model while in the other method still $4n$ parameters need to be estimated. The effect of other noises that contaminate the system is not studied in the prior cited articles.

Chapter 7

Direct Adaptive Control for Repeatable Runout Following in HDD

In this chapter we evaluate the effectiveness of the proposed direct adaptive control algorithm in following repeatable runout in a hard disk drive. The simulation and experimental arrangements are similar to the ones described in detail in chapter 5. We evaluate the algorithm with the same procedure as the indirect adaptive control algorithm in order to make the comparison between these two methods easy.

7.1 Computer Simulation Results

The closed loop dynamics from the VCM input to the PES, which was shown earlier in Fig. 5.4, has high magnitude response at frequencies less than 7KHz. Accordingly, we allocated the compensation of narrow–band disturbances at harmonics 1 to 58 to adaptive control modules that inject the control signal to the input of VCM. These modules are plugged similarly to C_A in Fig. 5.1. The remaining 115 disturbance components are left for the controller(s) that are plugged to the input of MA, which is similar to $C_{A,S}$ in Fig. 5.1.

The divisive hierarchical clustering method presented in Algorithm 2 is used for both actuators, starting from the VCM. The design parameters of the adaptive control algorithm, which are listed in Table 7.1, are common among all controllers applied to different frequency regions. The frequency allocation evolution by algorithm 2 is illustrated in Fig.7.1. As is shown in the figure, 32 iterations (shown on the vertical axis) were needed to partition the full spectrum to 11 regions. Each row in the figure indicates one iteration and the "current" region in that iteration, which was denoted by $\Omega_c(t)$ in the previous chapter, is shaded. Recall that the "current" region $\Omega_c(t)$ includes the harmonics that are being compensated at iteration t. For instance, based on the figure, in the first iteration the partitioner attempts to compensate the first 58 harmonics by the adaptive control module that is plugged to the input of VCM. As another example, at iterations 5, 6 and 7 the 3^{rd} region is being adjusted and then the algorithm checks region 4 at iteration 8. Based on the figure, the first

Table 7.1: Hyper Parameters of the adaptive control algorithm for both simulation and experiment.

n_A(6.9)	α(6.24)	β(6.24)	n_{fail} (6.28)
3	4E-5	1-(2E-7)	7
α_{dc}(6.26)	β_u(6.26)	α_β(6.27)	δ_u (4.13)
10	1E-3	1E-3	40Hz

Table 7.2: Final frequency partitions generated by algorithm 2 in our simulation study.

VCM			MA		
Region	Length	Harmonics	Region	Length	Harmonics
1	29	$1-29$	**5**	15	$59-73$
2	14	$30-43$	**6**	24	$74-97$
3	4	$44-47$	**7**	19	$98-116$
4	11	$48-58$	**8**	4	$117-120$
			9	2	$121-122$
			10	23	$123-145$
			11	28	$146-173$

8 iterations were used for partitioning the first 58 harmonics to 4 regions for the VCM loop. The following 24 iterations split the remaining 115 harmonics to 7 regions for the MA loop.

The final frequency segments are listed in Table 7.2. Comparing this table with the frequency responses of VCM and MA loops that are shown in Fig. 5.3 and 5.4 illustrates how effective the algorithm was to find the sharp and complex parts of the frequency responses and determine the region widths accordingly. For instance, Fig. 5.3 shows a very sharp phase shift around harmonic 120 which is at 14.4KHz, and Table 7.2 indicates a narrow region exactly at the same position that contains only 2 harmonics. It is an interesting result because the partitioning was completely automatic and no manual effort was required. The algorithm had no knowledge about the system dynamics but was able to iteratively partition the full spectrum between a set of "small" control modules that can compensate the disturbance at the allocated frequencies.

The estimated coefficients for $A(q^{-1})$ and $B(q^{-1})$ that construct $\hat{\theta}_A$ and $\hat{\theta}_B$ in region 1 are shown in Fig. 7.2. The figure shows that the estimated parameters converge to "some" values quickly. In order to evaluate the convergence point, we generated the transfer function $\frac{\hat{B}(q^{-1})}{\hat{A}(q^{-1})}$ that corresponds to these values. The frequency response of this transfer function is compared to the actual transfer function of the VCM loop – which is measured by an accurate

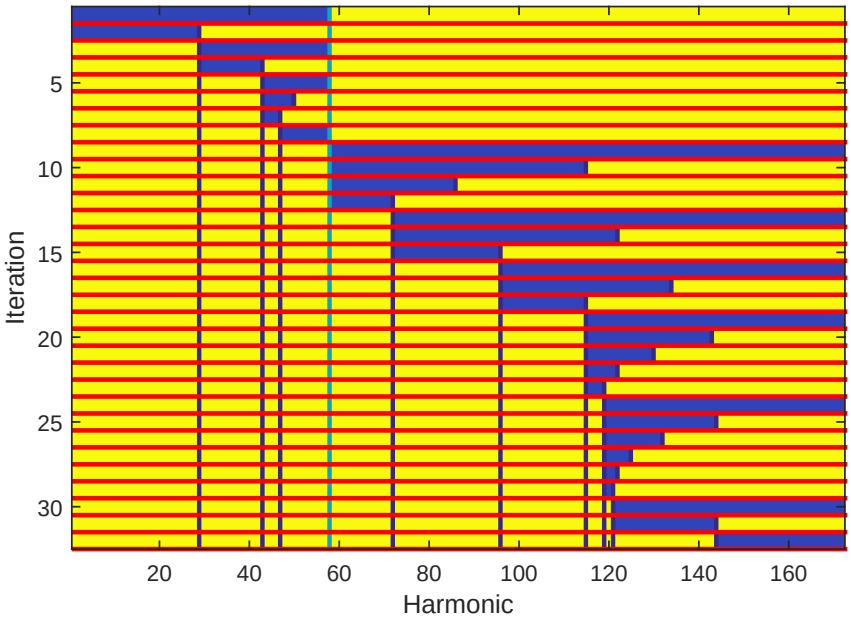


Figure 7.1: Frequency partitions generated in algorithm 2 iterations. The "current" region in each iteration t, which was denoted by $\Omega_c(t)$, is shaded. Harmonics "1 to 58" and "59 to 173" are partitioned for the VCM and MA loop respectively.

swept sine method – in Fig. 7.3. The shaded strip indicates region 1 where the adaptive controller was active. Recall that our objective is that each control module uses a fairly small number of parameters to find the system dynamics locally, and to this end, we use a special exogenous excitation signal with a pair of adaptive band–pass filters. As can be seen from the figure, these considerations were very effective and the estimated transfer function matches the actual dynamics at the compensation region very well. To consolidate this, the response mismatch between the two transfer functions is shown in Fig. 7.4. As can be seen in the figure, the phase mismatch between the two dynamics is within a 90-degree window in the compensation area.

The estimated residue parameters, $\hat{\theta}_M$, are depicted in Fig. 7.5. The plot shows that the residual disturbance in region 1 converges towards zero as the algorithm evolves. This is more clear in Fig. 7.6 that illustrates the convergence of $\alpha^u(k)$ in (6.26). Recall that $\alpha^u(k)$ is a smoothed version of $\|\hat{\theta}_M(k)\|_2$ that is used as the gain of excitation signal. Figure. 7.6 shows that the norm of residue vector (almost) exponentially converges to zero.

The previous results showed implicitly that the residual error is attenuated by the adaptive controller. This can be verified in frequency domain based on the spectrum of error too. The amplitude spectrum of the error before and after plugging the adaptive controller to the closed loop servo system are depicted in Fig. 7.7. For clearness, the figure only shows the amplitude of the error Fourier transformation at compensation frequencies. Comparing this

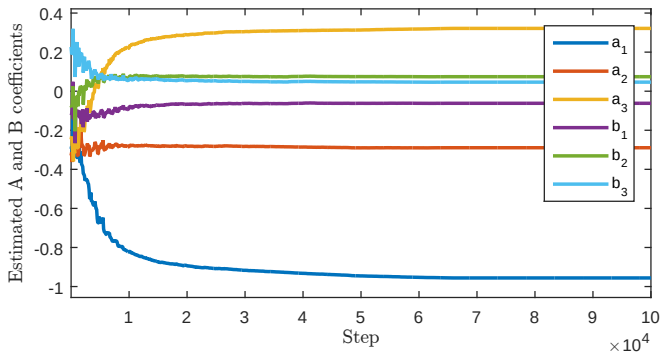


Figure 7.2: $-\hat{\theta}_A$ and $\hat{\theta}_B$ parameters in region 1 (harmonics 1 to 29 in simulation).

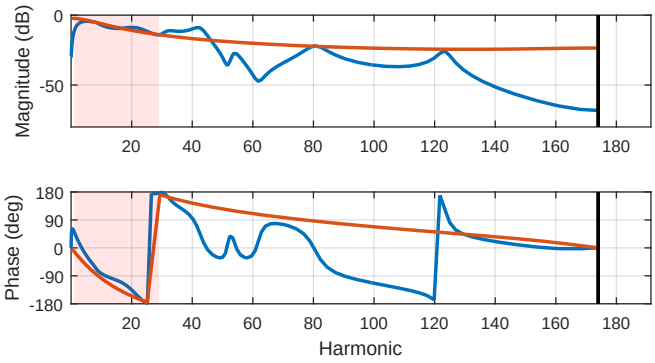


Figure 7.3: Frequency response comparison of the identified model and the actual VCM loop in region 1 which is indicated by a shaded strip (simulation study).

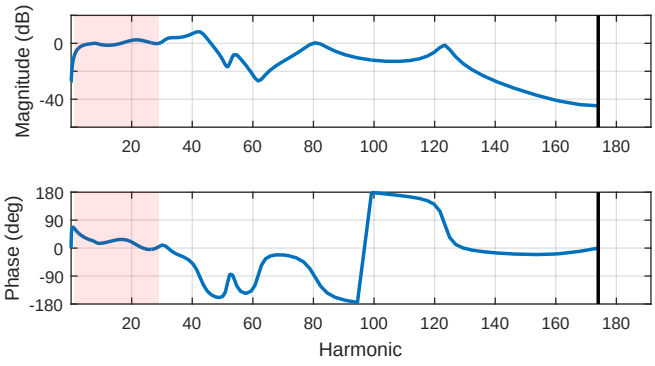


Figure 7.4: Frequency response mismatch between the identified model and the actual VCM loop in region 1 which is indicated by a shaded strip (simulation study).

figure with Fig. 5.5b that illustrates the non–repeatable runout spectrum verifies that the RRO is almost completely removed from this portion of spectrum.

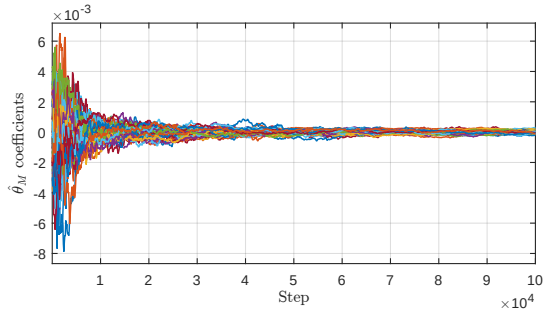


Figure 7.5: Estimated residue parameters, $\hat{\theta}_M$, in region 1 (simulation study).

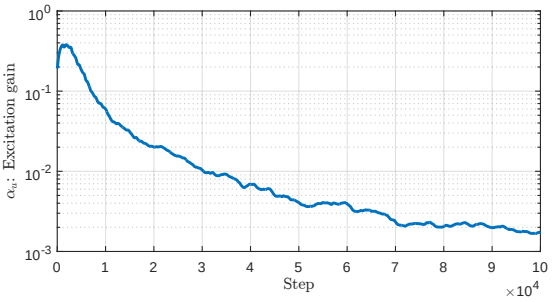


Figure 7.6: Excitation signal gain, $\alpha^u(k)$, in regions 1 (simulation study). Recall that $\alpha^u(k)$ is a smoothed version of $\|\hat{\theta}_M(k)\|_2$ as in (6.26) (simulation study).

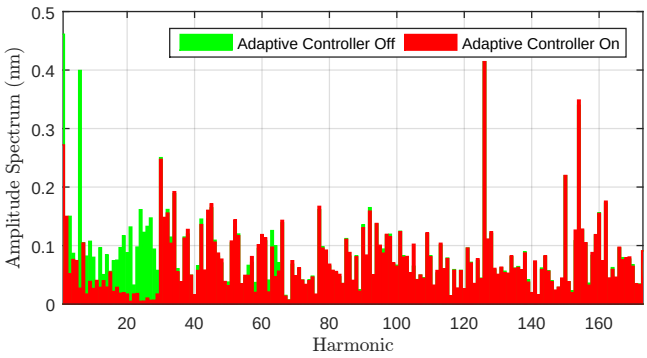


Figure 7.7: Comparison of the position error spectrum before and after plugging the adaptive controller to region 1. This figure shows the amplitude of Fourier transformation only at harmonics – i.e. other frequencies are removed (simulation study).

Note that all the results provided to this point belong to the first adaptive control "module" that is plugged to the input of VCM to compensate the first 15 harmonics. The control signal u_A learned by this controller is plotted in Fig. 7.8. This profile should be saved as one period of a periodic feedforward control sequence that can compensate the first 15 har-

monics. Analogous to this controller module, there are 10 other modules responsible for the remaining frequency partitions that are listed in Table 7.2. The controllers corresponding to these partitions can be run simultaneously or non-concurrently depending on the available computational power.

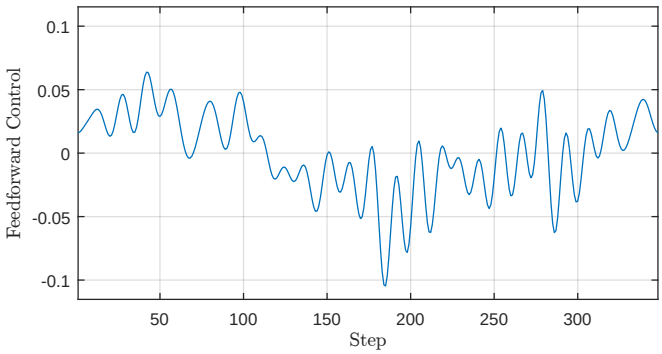


Figure 7.8: Feedforward control signal, $u_A(k)$, learned in region 1 (simulation study).

We obtained the same type of results for the other 10 adaptive controllers. Since the convergence behavior of all cases were similar and due to space limitation we only present the results for the last region (region 11). Figures 7.9 to 7.11 show the estimated parameters and the corresponding transfer function for the closed loop dynamics from the MA input to the PES. As can be seen from the figures, the mismatch in this case is even smaller than the previous results. Figures. 7.12 and 7.13 verify that the residual error in all compensation frequencies converge towards zero. Finally, Fig. 7.15 depicts the learned control signal that should be saved as one period of a periodic feedforward sequence for compensating the last 28 frequencies.

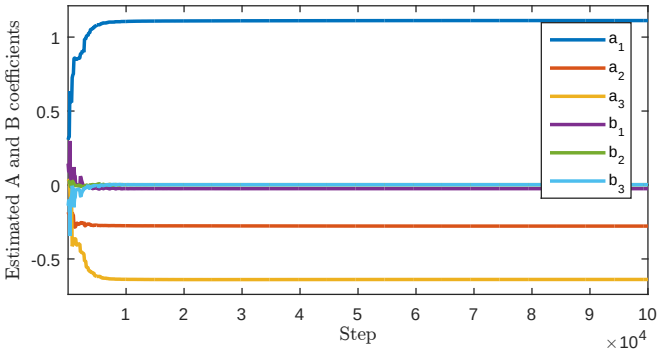


Figure 7.9: $-\hat{\theta}_A$ and $\hat{\theta}_B$ parameters in region 11 (harmonics 146 to 173 in simulation).

After learning the adaptive control signal in each region, we superpose it to a feedforward control table. Since the system has two inputs, two individual feedforward tables are considered for the input of MA and the VCM. Once the control sequences for all frequencies are

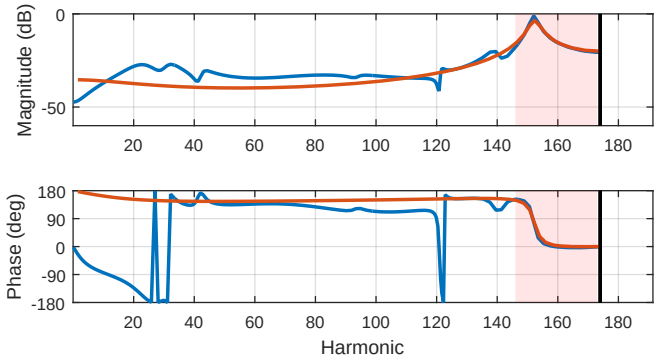


Figure 7.10: Frequency response comparison of the identified model and the actual MA loop in region 11 which is indicated by a shaded strip.

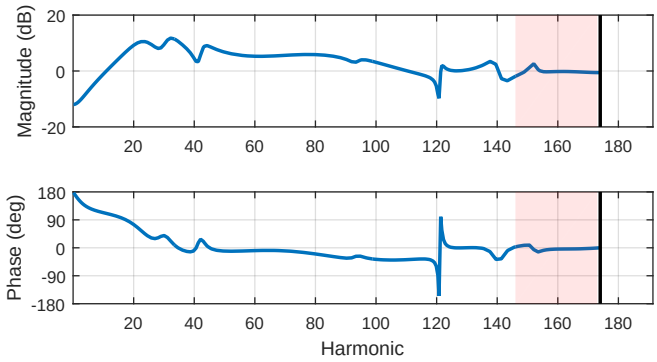


Figure 7.11: Frequency response mismatch between the identified model and the actual MA loop in region 11 which is indicated by a shaded strip (simulation study).

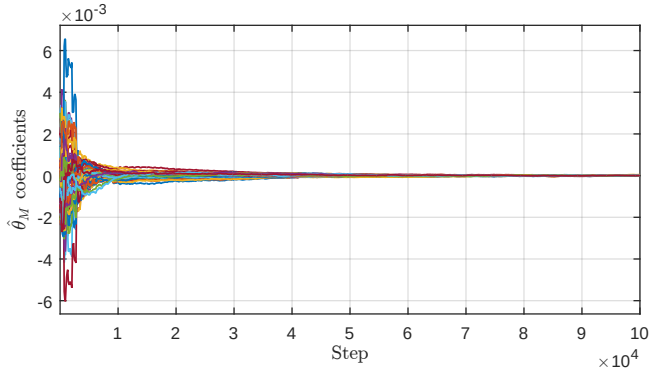


Figure 7.12: Estimated residue parameters, $\hat{\theta}_M$, in region 11 (simulation study).

learned, it is expected that full spectrum compensation can be achieved by injecting these two feedforward sequences to the dual–stage system. Figure 7.15 shows the amplitude of the

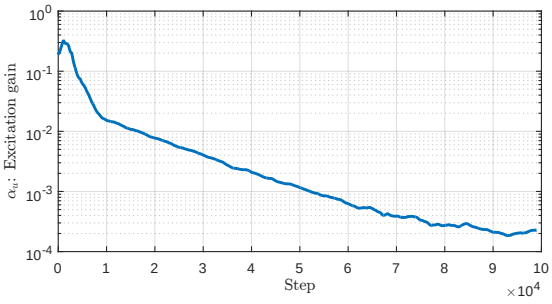


Figure 7.13: Excitation signal gain, $\alpha^u(k)$, in regions 11. Recall that $\alpha^u(k)$ is a smoothed version of $\|\hat{\theta}_M(k)\|_2$ as in (6.26) (simulation study).

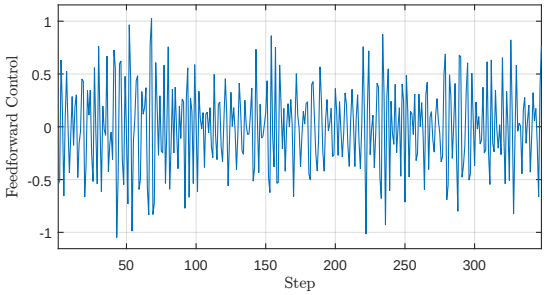


Figure 7.14: Feedforward control sequence, $u_A(k)$, learned in regions 11 (simulation study).

error Fourier transformation after applying these feedforward controls. As can be seen from the figure, all control modules were successful in attenuating the narrow–band disturbances. We will validate these results with experiments on an actual hard disk drive in the next section.

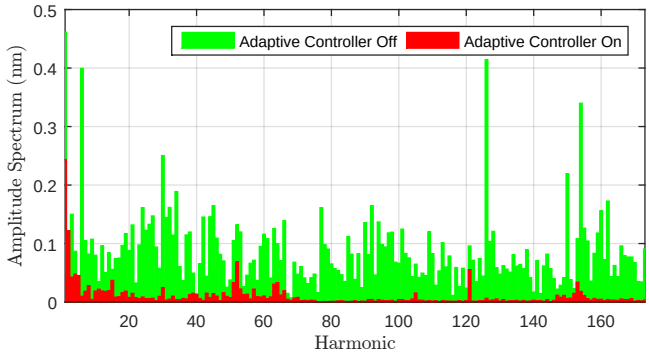


Figure 7.15: Position error amplitude spectrum. Green: uncompensated position error signal. Red: full spectrum narrow–band disturbance rejection after learning the feedforward control in all 11 regions (simulation study).

7.2 Experimental Results

7.2.1 Time–Invariant System Dynamics and Disturbance Profile

The direct adaptive control algorithm was implemented on a Digital Signal Processor, and the same set of parameters as the simulation (Table. 7.1) were used. In this section we report the results for the case that the system dynamics and disturbance are time–invariant. That is, the HDD dynamics and RRO profile do not change because we only track–follow on head 0 and track 3000 of the disk. The next section studies to the cases that the head and track changes cause abrupt variations in the closed loop dynamics and disturbance profile.

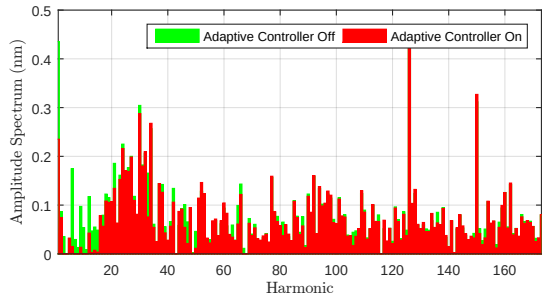
The frequency partitions obtained by the proposed divisive hierarchical clustering method are listed in Table 7.3. The intervals are slightly different than the simulation results that are listed in Table 7.2. This can be due to the noise model mismatch because whether a low order model fits properly to the actual dynamics or not depends on broad–band noises that cause biased estimation. However, the number of regions is the same in both simulation and experiment, showing that the same numbers of parameters have been estimated.

The Fourier transform amplitudes of error after the first and last region compensation are shown in Fig. 7.16. Again, our approach is to learn the control signal by running the adaptive control algorithm on each region when the control learned from previous regions is superposed and injected to the system as a feedforward control. Therefore, after learning the last region, which is 11 here, the control signal will contain contributions from all regions. As the figure shows, all harmonics are attenuated to significantly small amplitudes. The left over is slightly larger than the simulation results, which can be caused by non-stationarity of broad band noises in the real environment.

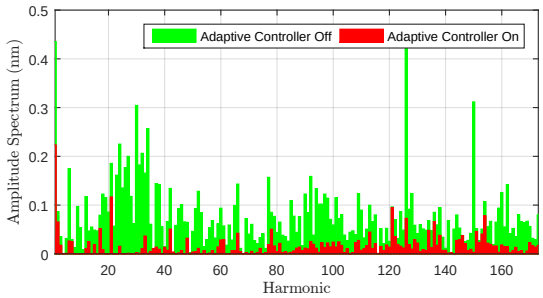
The convergence of error towards zero can also be verified through the estimates of residual. The parameter vector $\hat{\theta}_M$ estimated at regions 1 and 11 is shown in Fig. 7.17. The attenuation can also be evaluated by the gain of the excitation signal (6.26) which is related

Table 7.3: Final frequency partitions generated by algorithm 2 (DSP implementation).

VCM			MA		
Region	Length	Harmonics	Region	Length	Harmonics
1	15	$1-15$	**5**	20	$59-78$
2	18	$16-33$	**6**	21	$79-99$
3	15	$34-48$	**7**	21	$100-120$
4	10	$49-58$	**8**	2	$121-122$
			9	16	$123-138$
			10	20	$139-158$
			11	15	$159-173$

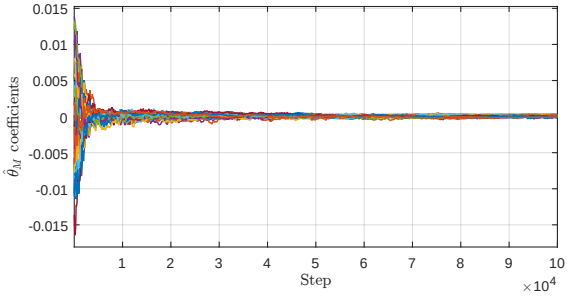


(a) Compensation of region 1 (1-15)

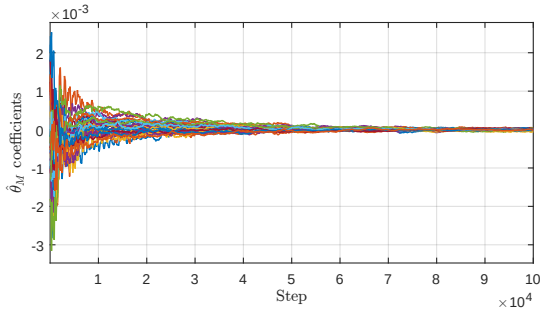


(b) Compensation of region 11 (159-173)

Figure 7.16: Comparison of the position error spectrum before and after plugging the adaptive controller (DSP implementation). Only harmonics are shown and the broad band disturbance is removed.



(a) Residual parameters in region 1



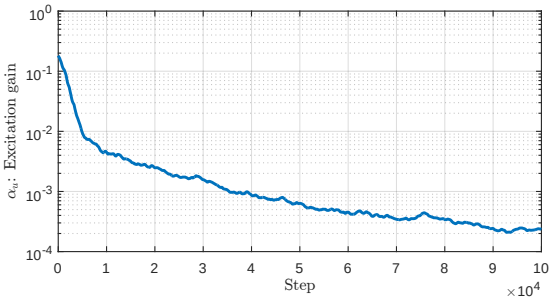
(b) Residual parameters in region 11

Figure 7.17: Estimated parameters for the residual error, $\hat{\theta}_M$, at regions 1 and 11.

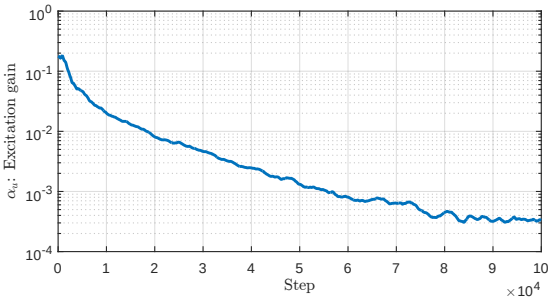
to the norm of residual. Similar to the simulation results, this gain converges toward zero almost exponentially as illustrated in Fig. 7.18. These parameters represent the convergence of residual error toward zero in time domain.

The plant parameters, $-\hat{\theta}_A$ and $\hat{\theta}_B$ for the same two regions are illustrated in Fig. 7.19. Since these two regions are different than the corresponding ones in the simulation, the plots differ from Fig. 7.2 and 7.9. The transfer functions associated with these set of parameters are compared to the actual system dynamics in frequency domain in Fig. 7.20. As the plots show, the local excitation signal has driven the estimates toward values that fit the actual dynamics locally around the compensation frequencies (shaded areas in the figure). The mismatches of these models are depicted in Fig. 7.21 which illustrates that the phase difference was in a $\pm$90-degree window at all compensation frequencies.

The total injection signal which contains both the excitation signal and control effort for the same two regions is shown in Fig. 7.22. The total injection signal is larger at the beginning since the residual error is large. As the control parameters converge to optimal values and the residual error gets smaller, the excitation signal decreases. This implies that

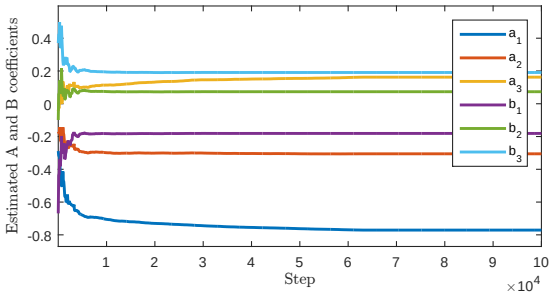


(a) Region 1

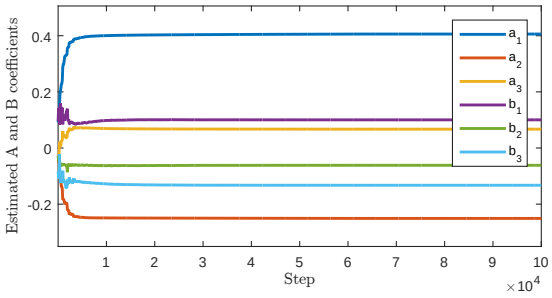


(b) Region 11

Figure 7.18: Excitation signal gain, $\alpha^u(k)$, in experiments by DSP implementation. Recall that $\alpha^u(k)$ is a smoothed version of $\|\hat{\theta}_M(k)\|_2$ as in (6.26).

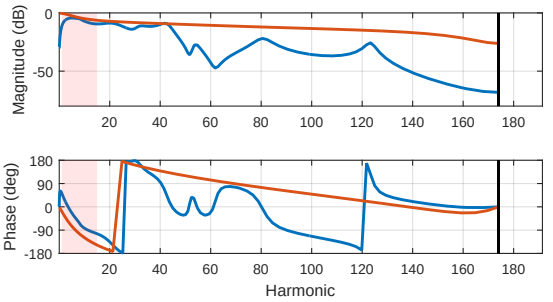


(a) Plant parameters in region 1

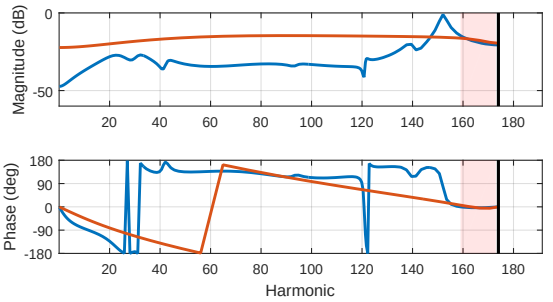


(b) Plant parameters in region 11

Figure 7.19: $-\hat{\theta}_A$ and $\hat{\theta}_B$ parameters estimated in DSP implementation.



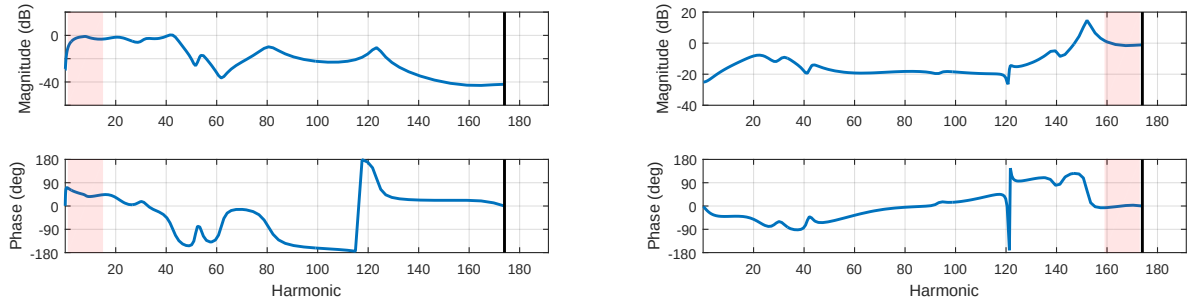
(a) Local model learned in region 1



(b) Local model learned in region 11

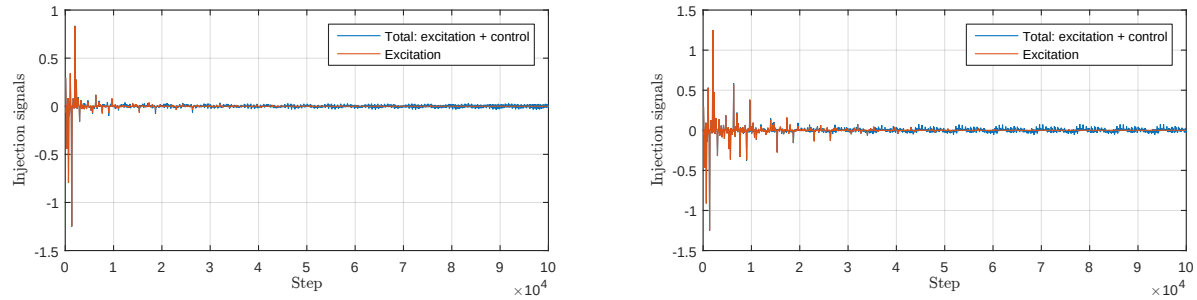
Figure 7.20: Frequency responses of the actual and estimated systems (DSP implementation). The shaded area shows the region of compensation.

the excess error caused by the excitation signal vanishes in steady state. Figure 7.23 shows the final feedforward control signals for the two actuators after running the algorithm on all regions.



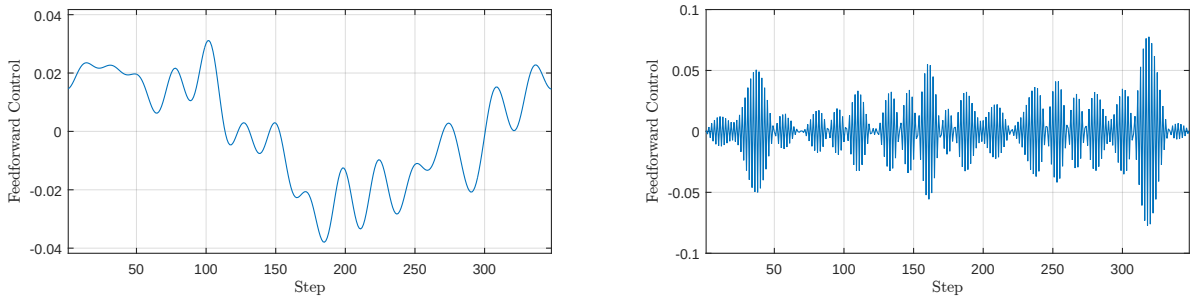
(a) Frequency response mismatch in region 1 (b) Frequency response mismatch in region 11

Figure 7.21: Frequency response mismatches for the plant models shown in Fig.7.20. The shaded area shows the region of compensation.



(a) Injection signal in region 1 (b) Injection signals in region 11

Figure 7.22: The control signal (blue) and total injection signal that contains both the excitation and control signals. The excitation signal amplitude is determined adaptively and it fades quickly as the residual error converges toward zero.



(a) Feedforward control for the VCM (b) Feedforward control for the MA

Figure 7.23: Feedforward control signals injected to the input of VCM and MA. The control profile at each injection point (i.e. VCM or MA) is generated by superposing the control signals obtained from regions allocated to that particular point.

7.2.2 Time–Varying System Dynamics and Abrupt Changes in Disturbance Profile

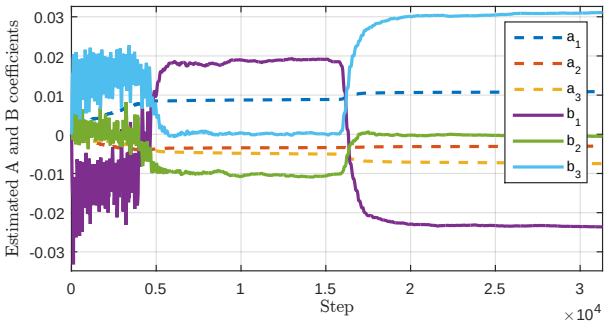
In this subsection we study the behavior of the proposed control algorithm in situations that the system dynamics or the disturbance profile changes. These cases occur when different magnetic heads are used for servoing. Imagine the case when head−0 reads the PES from disk surface−0 and feeds it back to the controllers. In this case, MA−0 actuates the suspension assembly that holds head−0. By changing the servoing head to head−1 the closed loop dynamics varies because: (1) The heads dynamics are different. (2) The mechanical assemblies that hold the heads are different. As a result, the transfer functions from both the VCM and the MA to the PES change. (3) The two heads read the PES from two sides of one disk which implies that the sign of PES is reversed. The third factor can be easily fixed by multiplying the PES of even heads by -1. However, we do not do this here because from a control point of view, this sign difference is equivalent to 180 degrees phase error and we would like to know whether our controller can cope with such a case.

In the following, we show the controller behavior in frequency regions 8 and 11 (see Table 7.3) and for the cases that we change the head from 0 to 1 and 2. Region 8 contains the frequencies where the MA dynamics has a very sharp phase change and it varies significantly from head to head and by temperature. On the other hand, the system dynamics in frequency interval 11 is very similar among different heads. However, note that we still have approximately 180 degrees phase mismatch between head−0 and head−1. Therefore, only changing the head from 0 to 2 at region 11 causes "small" variations in system dynamics. These 4 cases are summarized in the following.

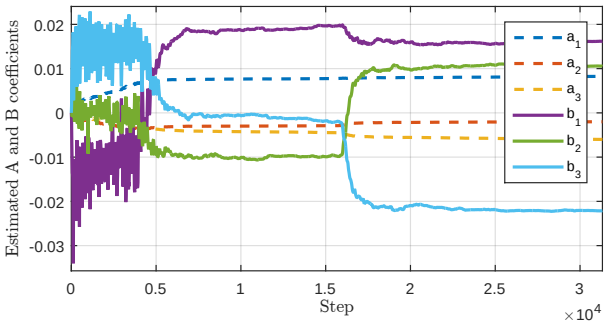
- **Case 1:** Changing from head−0 to head−1 in region 8. This causes large dynamics and large disturbance variations.
- **Case 2:** Changing from head−0 to head−2 in region 8. This causes large dynamics and small disturbance variations.
- **Case 3:** Changing from head−0 to head−1 in region 11. This causes large dynamics and large disturbance variations.
- **Case 4:** Changing from head−0 to head−2 in region 11. This causes small dynamics and small disturbance variations.

The estimated parameters in $\hat{\theta}_A$ and $\hat{\theta}_B$ that correspond to the system dynamics are shown in Fig. 7.24 for the above four cases. In all cases, we changed the servoing head after approximately 1.6×10^4 steps. As can be seen from the plots, in all situations the algorithm is able to learn the parameters for the new dynamics and no divergence behavior can be observed. The question of whether these new parameters are "meaningful" or not can be answered by looking at the error residues.

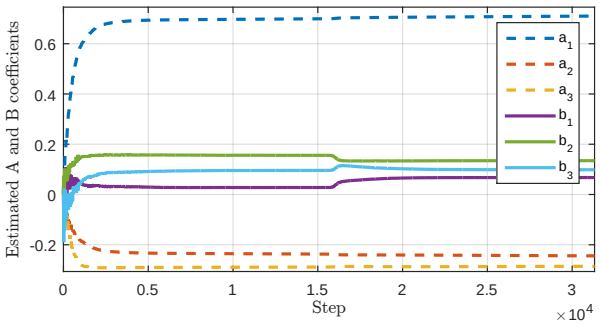
Recall that parameter vector $\hat{\theta}_M(k)$ is an approximator of the error in frequency domain. The evolution of this vector of parameters for the above four cases is shown in Fig. 7.25.



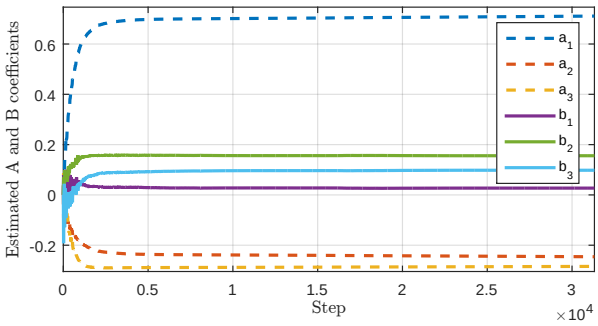
(a) Case 1: from head–0 to head–1 in region 8.



(b) Case 2: from head–0 to head–2 in region 8.



(c) Case 3: from head–0 to head–1 in region 11.



(d) Case 4: from head–0 to head–2 in region 11.

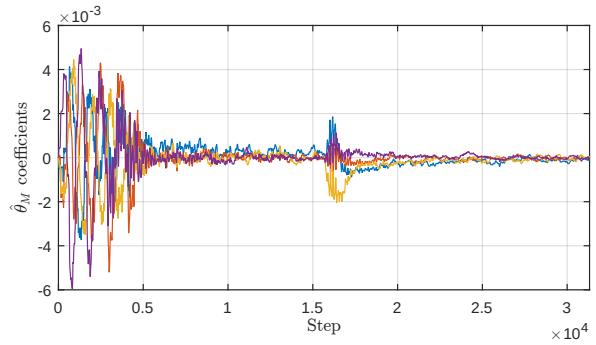
Figure 7.24: Estimated $-\hat{\theta}_A$ and $\hat{\theta}_B$ parameters under abrupt variation of the plant dynamics.

These plots clearly illustrate that after changing the head, the new sets of estimated plant parameters made the adaptive algorithm converge in a direction that the error spectrum at all compensation frequencies converges towards zero.

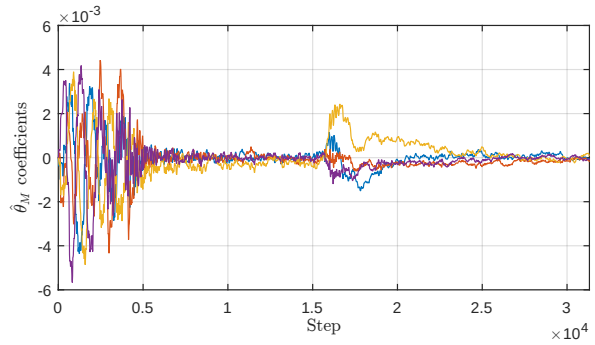
We can also use the excitation signal gain $\alpha^u(k)$ to verify the overall error attenuation. Based on (6.26) this parameters approximate the 2–norm of residual error $\|\hat{\theta}_M(k)\|_2$. This parameter for all four cases is depicted in Fig. 7.26. The figure shows that in all situations the adaptive control algorithm was able to adapt to the new system dynamics and disturbance very quickly and the same, or even better, attenuation level was achieved.

7.2.3 Tracking Repeatable Runout of a BPMR HDD

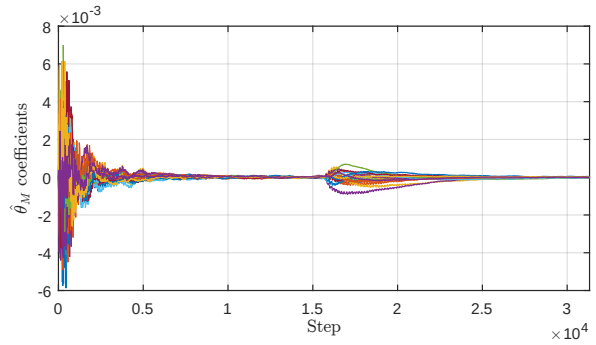
BPMR technology is still in research stage and BPMR HDDs have not been mass–produced yet. However, our industry partners that are pioneers in developing this technology were able to provide us with the noise characteristics of prototype BPMR hard disk drives. It is expected that the hardware of conventional and BPMR HDDs only differ in the media. This implies that using the noise models of a BPMR HDD with the plant dynamics of a conventional HDD can potentially mimic a BPMR HDD characteristics. We extracted the RRO profile at different tracks from the dataset provided by our industry partners. This profile



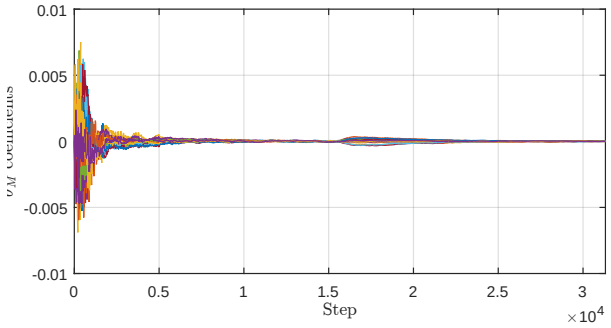
(a) Case 1: from head–0 to head–1 in region 8.



(b) Case 2: from head–0 to head–2 in region 8.



(c) Case 3: from head–0 to head–1 in region 11.

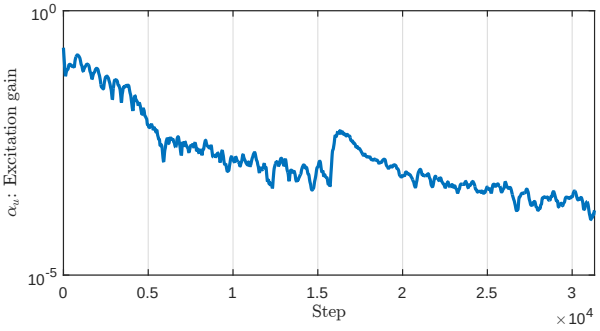


(d) Case 4: from head–0 to head–2 in region 11.

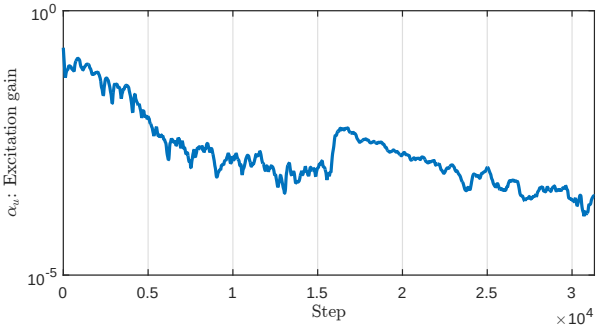
Figure 7.25: Estimated parameters for the residual error $\hat{\theta}_M$ under abrupt variations of system dynamics and disturbance profile (DSP implementation).

was then artificially added to the PES measurements of the 3.5" HDD in our experimental setup.

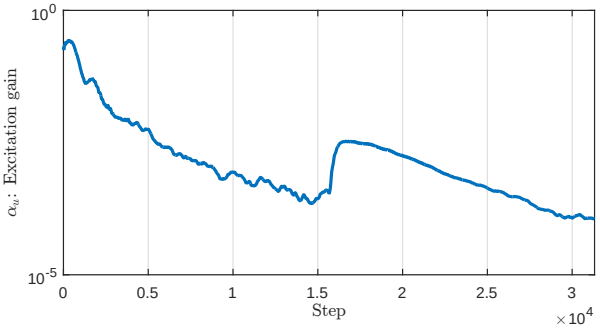
We used the same DSP implementation as the previous section. All parameters and other settings are exactly the same as what was explained earlier. The controller behavior in this case was very similar to tracking the RRO of a conventional HDD. Hence, we only present the error spectrum before and after plugging the controller. Figure. 7.27 illustrates these results. As can be seen from the figure, the RRO spectrum at low frequency is considerably larger than a conventional HDD. This is probably because of the relatively large eccentricity in electron-beam lithography. However, as the plots illustrate, the adaptive control algorithm was able to attenuate the tracking error to significantly small values at all harmonics.



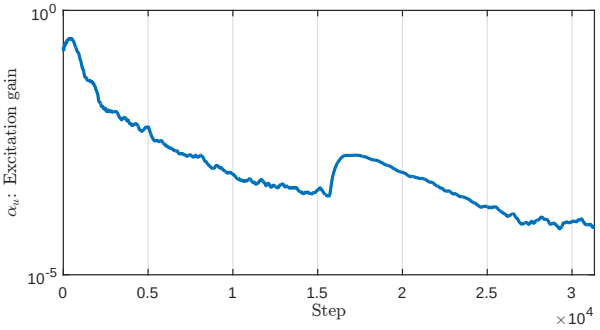
(a) Case 1: from head–0 to head–1 in region 8.



(b) Case 2: from head–0 to head–2 in region 8.

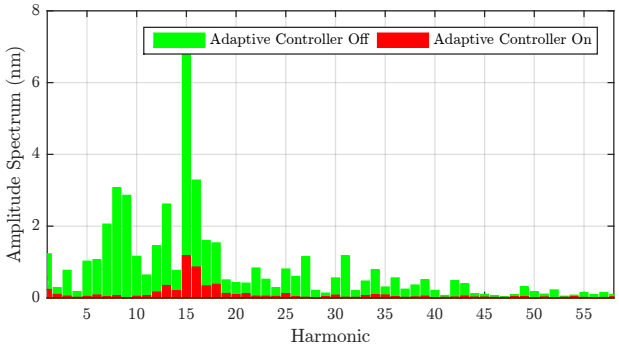


(c) Case 3: from head–0 to head–1 in region 11.

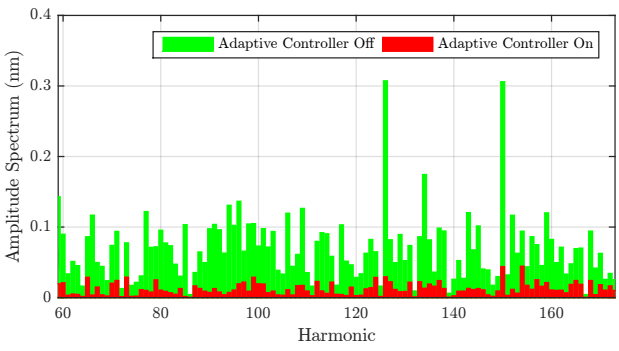


(d) Case 4: from head–0 to head–2 in region 11.

Figure 7.26: Excitation signal gain, $\alpha^u(k)$, in regions 11. Recall that $\alpha^u(k)$ is a smoothed version of $\|\hat{\theta}_M(k)\|_2$ as in (6.26) (DSP implementation).



(a) Compensation through VCM



(b) Compensation through MA

Figure 7.27: Comparison of the position error spectrum before and after plugging the adaptive controller when the RRO profile of a BPMR HDD is artifically added to a conventional HDD (DSP implementation).

Chapter 8

Conclusion and Future Work

In this dissertation, we considered the problem of unknown trajectory tracking or equivalently deterministic (but unknown) disturbance rejection in discrete time systems when the trajectory or the disturbance can be realized (exactly or approximately) by an affine combination of known basis functions. Among various applications of this control task, we were particularly interested in nanopositioning in hard disk drives (HDDs).

We deployed a stochastic gradient descent method to develop an adaptive feedforward control algorithm for compensating multitude narrow–band disturbances or trajectory tracking. A set of conditions on the adaptation step size was derived to guarantee the algorithm convergence and perfect compensation. We proposed a novel adaptive step size to enhance the convergence rate and decrease the steady state error. This analysis was initially carried out in a spectral framework where trigonometric functions were chosen to form an orthogonal basis for the space of real valued square integrable trajectories. Two alternatives in time–domain and in time–frequency–domain were also investigated. It was shown that time–domain approaches can be computationally more plausible than frequency–domain methods. However, the latter type can be more robust to system dynamics uncertainty.

We analyzed the robustness of our proposed adaptive feedforward controller to dynamic mismatches between the actual plant and the model deployed in the algorithm and showed that, unlike many existing methods, the mismatches are only important at excitation frequencies. Moreover, an online system identification architecture was proposed to provide an accurate model of the system dynamics in case a model is not available or accurate. It was shown that under a set of assumptions, the proposed scheme is able to obtain a model that provides necessary conditions for the adaptive controller to achieve perfect compensation. As a result, the identification and compensation mechanisms together formed an "indirect" adaptive controller for the class of problems under our study.

We proposed a special "low–power" excitation signal that was tailored for the HDD application. Besides, we suggested using adaptive band–pass filters on the inputs to the identification unit in order to further reduce the required excitation power. An online method for synthesizing these adaptive filters through *frequency transformation* of a prototype filter was proposed.

In addition to the indirect adaptive control scheme, we proposed a direct adaptive algorithm for the problem of compensating (tracking) periodic disturbances (trajectories) in systems with unknown dynamics. We showed that the algorithm is applicable to non–minimum phase systems without requiring any assumptions. Moreover, the method is applicable in situations with slowly time–varying systems and disturbances. Also, it can be used in applications that the system dynamics or disturbance profile has infrequent abrupt changes. The number of estimated parameters in our proposed algorithm is slightly larger than $2n$ where n is the number of disturbance/trajectory frequency contents, while other methods require $4n$ parameters ($2n$ for system dynamics and $2n$ for control synthesis). As a result, our method estimates significantly less parameters when n is large (e.g. as in the HDD problem where $n = 174$). The convergence of adaptive algorithm parameters to actual values was rigorously analyzed and a set of practical remarks were made for reducing the transient error.

We showed that both of the proposed adaptive control algorithms are "modular", meaning that a controller with a large number of parameters can be split to "smaller" controllers such that each one estimates a portion of the original unknown parameter vector. This is a very appreciated property because a difficult problem can be break into a set of easier control problems without increasing the computational effort. We also proposed a heuristic algorithm inspired by divisive hierarchical clustering in conjunction with a monitoring mechanism to determine the required controller "modules" automatically.

Besides providing theoretical guarantees, we experimentally evaluated our algorithms on a challenging control task for nanopositioning of the read–write head in a dual–stage HDD. The importance of HDDs compared to other types of data storage technologies such as *NAND–flash*–based memory devices and *Tapes* was explained. Moreover, it was described that the control methodologies that can address the problem under our study are crucial for an emerging breakthrough in magnetic recording that is called *Bit–Patterned Media Recording*.

Computer simulation in MATLAB and implementation on a digital signal processor (DSP) unit were performed to compensate for RRO that has narrow–band contents at the HDD spinning frequency (120Hz) and its 173 higher harmonics (up to 20'880Hz). This is a challenging task since it requires estimating a very large number (348) of parameters which is order(s) of magnitude greater than other results reported in the literature. These frequencies span from 120Hz to extremely large frequencies (above 20KHz) where the plant dynamics uncertainties are large and feedback controller amplifies disturbances. We proved by simulations and experiments that both proposed algorithms can be applied to such a mechatronic device and full spectrum compensation can be achieved.

While we have considered disturbances/trajectories that can be realized as an affine combination of a set of known basis functions, these functions may not be available in many real-world scenarios. For instance, in the case of sinusoidal disturbances, the frequencies may drift over time. As such, future work will combine the adaptive algorithm presented here with basis functions estimation techniques. As in the sinusoidal disturbance case, frequency estimation methods such as Wu and Bodson (2003, 2004) can be deployed for applications with time-varying frequency acting on unknown and time-varying systems.

Appendix A

Proofs of Theorems in Chapter 6

We use the method of analysis of stochastic recursive algorithms discussed in Ljung (1977a) to study the convergence and asymptotic behavior of the proposed adaptive algorithm with the update rule given in (6.21). The key idea is that a differential equation counterpart for (6.21) is derived and it is shown that the only convergence point of (6.21) is the stable equilibrium of this differential equation. Moreover, it is shown that the estimated parameters converge with probability one to this equilibrium under some mild assumptions on the excitation signal power or initial values of parameters.

A.1 Preliminaries

A general recursive algorithm can be formalized as

$$\hat{\Theta}(k) = \hat{\Theta}(k-1) + \gamma(k)Q(k; \hat{\Theta}(k-1), \Phi(k)) \tag{A.1}$$

where $\hat{\Theta}(k) \in \Re^n$ is a sequence of n-dimensional column vectors referred as "the estimates". In our particular recursive algorithm, they stand for the parameters of system dynamics in conjunction with the parameters that determine the adaptive control law. $\gamma(k)$ is assumed to be a sequence of positive scalars. The m-dimensional vector $\Phi(k)$ is an observation obtained at time k which is usually called the "regressor". The information contained in $\Phi(k) \in \Re^m$ sequence cause $\hat{\Theta}(k-1)$ to be updated. The map $Q(.;.,.)$ from $R \times R^n \times R^m$ into R^n is a deterministic function with some regularity conditions that guarantee the convergence of the adaptive algorithm to some "desired" points that will be discussed later. From the above equation it can be inferred that $Q(.;.,.)$ together with the choice of the "gain" sequence $\gamma(.)$ determine entirely the algorithm.

The regressor vector $\Phi(k)$ is, in general, a function of the previous measurements (when the system has memory), the previous estimates (when the estimates define an adaptive control law), and the inputs to the system (all other control signals and noises). A very broad class of update rules, as of ours, can be realized by a time varying linear system

$$\Phi(k) = A_\Phi(\hat{\Theta}_{k-1})\Phi(k-1) + B_\Phi(\hat{\Theta}_{k-1})U(k)$$

where A_Φ and B_Φ are $m \times m$ and $m \times r$ matrix functions and $U(k) \in \Re^r$ stands for all control signals and noises entering the system.

In a stochastic framework, $Q(k; \hat{\Theta}(k-1), \Phi(k))$ is a random variable. This implies that convergence can take place only if the noise is rejected by paying less attention to the observations as time passes, i.e. by letting

$$\gamma(k) \to 0 \qquad \text{as} \quad k \to \infty.$$

However, this is not feasible when the actual (system) parameters are time-varying, as in tracking problems. For such problems, $\gamma(k)$ can tend to very small, but still positive, values and the analysis carried out in this section will give some insights regarding the algorithm behavior.

The estimated parameters and the regressor in our adaptive algorithm are

$$\hat{\Theta}(k) := \begin{bmatrix} \hat{\theta}_A(k) \\ \hat{\theta}_B(k) \\ \hat{\theta}_C(k) \\ \hat{\theta}_M(k) \\ \text{col}\,[F(k)] \\ f(k) \end{bmatrix} \qquad \Phi(k) := \begin{bmatrix} \dfrac{\phi_e(k)}{u(k)} \\ \dfrac{\phi_u(k)}{\bar{w}(k)} \\ \dfrac{\phi_{\bar{w}}(k)}{\phi_\epsilon(k)} \\ \dfrac{\phi_R(k)}{\phi_{u_A}(k)} \end{bmatrix}.$$

The A_Φ and B_Φ matrix functions can be written as

$$A_\Phi(\hat{\Theta}_{k-1}) := \left[\begin{array}{ccccccc} \theta_A^T & \begin{bmatrix} 0 & \theta_B^T \end{bmatrix} & \begin{bmatrix} 0 & \theta_C^T \end{bmatrix} & 0_{n_A} & \theta_R^T & \theta_B^T \\ \hline \begin{bmatrix} I_{n_A-1} & 0 \end{bmatrix} & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & \begin{bmatrix} 0 & 0 \\ I_{n_A} & 0 \end{bmatrix} & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & \begin{bmatrix} 0 & 0 \\ I_{n_A} & 0 \end{bmatrix} & 0 & 0 & 0 \\ \hline \hat{\theta}_A^T(k-1) & \begin{bmatrix} 0 & \hat{\theta}_B^T(k-1) \end{bmatrix} & 0 & \hat{\theta}_C^T(k-1) & \hat{\theta}_M^T(k-1) & 0 \\ \hline 0 & 0 & 0 & \begin{bmatrix} I_{n_A-1} & 0 \end{bmatrix} & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & \hat{\theta}_D^T(k-1) & 0 \\ 0 & 0 & 0 & 0 & 0 & \begin{bmatrix} I_{n_A-1} & 0 \end{bmatrix} \end{array}\right] \tag{A.2}$$

$$B_\Phi(\hat{\Theta}_{k-1}) := \begin{bmatrix} 0 & 0 & 0 \\ \begin{bmatrix} I \\ 0 \end{bmatrix} & 0 & 0 \\ 0 & \begin{bmatrix} I \\ 0 \end{bmatrix} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \begin{bmatrix} I \\ 0 \end{bmatrix} \\ 0 & 0 & 0 \end{bmatrix} \tag{A.3}$$

and the input vector is

$$U(k) := \begin{bmatrix} u(k) \\ \bar{w}(k) \\ \phi_R(k) \end{bmatrix}. \tag{A.4}$$

The next part provides certain regularity conditions on functions Q, A_Φ and B_Φ as well as driving "input" term U to analyze the behavior of (A.1).

Before that, we provide two Lemmas about Lipschitz continuity (LC) properties of real valued functions since these results will be used in the sequel. The proofs of both lemmas can be found in Eriksson et al. (2013).

Lemma 5 (Lipschitz continuity of product of functions)
$f(x) = f_1(x)f_2(x)$ is Lipschitz continuous on a bounded set I if $f_1(.)$ and $f_2(.)$ are individually Lipschitz continuous on the same set.

Lemma 6 (Lipschitz continuity of quotient of functions)
$f(x) = \frac{f_1(x)}{f_2(x)}$ is Lipschitz continuous on a bounded set I if $f_1(.)$ and $f_2(.)$ are individually Lipschitz continuous on the same set and there is a positive constant m such that for any $x \in I$, $f_2(x) \geq m$.

A.2 Regularity Conditions

Three sets of regularity conditions are proposed in Ljung (1977a) to analyze a recursive algorithm in the form of (A.1). The first two sets, referred as "Assumptions A" and "Assumptions B", consider $U(k)$ as a sequence of random variables and treat the algorithm in a stochastic framework. In our method, the input signal $U(k)$ (A.4) is consisted of stochastic and deterministic parts, e.g. $\Phi_R(k)$ is a known vector whereas $\bar{w}(k)$ is a random sequence. The third set of assumptions referred as "Assumptions C" is more general and will be used here. Let $\mathscr{D}_S(x)$ be the set of all $A_\Phi(x)$ that have all eigenvalues strictly inside the unit circle and let $\mathscr{D}_R$ be an open connected subset of $\mathscr{D}_S(x)$. "Assumptions C" in Ljung (1977a) are given in the following.

Assumptions C

C.1 The function $Q(k, x, \phi)$ is Lipschitz continuous in x and ϕ. In other words, $|Q(k, x_1, \phi_1) - Q(k, x_2, \phi_2)| < \mathscr{K}(x, \phi, \rho, \nu)\{|x_1 - x_2| + |\phi_1 - \phi_2|\}$ for any x_1 and x_2 in $\mathscr{B}(x, \rho(x))$ ($\mathscr{B}(x, \rho)$ denotes a ρ-neighborhood of x) for some $\rho(x) > 0$ where $x \in \mathscr{D}_R$. Moreover, ϕ_1 and ϕ_2 should be in $\mathscr{B}(\phi, \nu)$ for some $\nu \geq 0$.

C.2 The matrix functions $A_\Phi(x)$ and $B_\Phi(x)$ are Lipschitz continuous in x.

C.3 $z(k, \bar{x})$ as defined by $z(0, \bar{x}) = 0$ and

$$z(k, \bar{x}) = z(k-1, \bar{x}) + \gamma(k)\left[Q(k, \bar{x}, \bar{\phi}(k, \bar{x})) - z(k-1, \bar{x})\right], \quad z(0, \bar{x}) = 0 \tag{A.5}$$

converges for all $\bar{x} \in \mathscr{D}_R$ as $k \to \infty$. Denote the limit by $\mathscr{F}(\bar{x})$.

C.4 $k_v(k, \bar{x}, \lambda, c)$ defined by $k_v(0, \bar{x}, \lambda, c) = 0$ and

$$\begin{aligned} k_v(k, \bar{x}, \lambda, c) = k_v(k-1, \bar{x}, \lambda, c) + \gamma(k) \big[\mathscr{K}\left(\bar{x}, \bar{\phi}(k, \bar{x}), \rho(\bar{x}), \nu(k, \lambda, c)\right) \\ \cdot (1 + \nu(k, \lambda, c)) - k_\nu(k-1, \bar{x}, \lambda, c)] \qquad k_v(0, \bar{x}, \lambda, c) = 0 \end{aligned} \tag{A.6}$$

converges to a finite limit as $k \to \infty$ for all $\bar{x} \in \mathscr{D}_R$, $\lambda < 1$ and $c < \infty$.

C.5 $\sum_{k=1}^{\infty} \gamma(k) = \infty$.

C.6 $\gamma(k) \to 0$ as $k \to \infty$.

Assumption C.1: The function $Q(k; \hat{\Theta}(k-1), \Phi(k))$ in (A.1) in our algorithm is time invariant. We abuse the notation and use $Q(\hat{\Theta}(k-1), \Phi(k))$ in the sequel. This function can be written as

$$Q\left(\hat{\Theta}(k-1), \Phi(k)\right) := \begin{bmatrix} Q_1(\hat{\Theta}(k-1), \Phi(k)) \\ Q_2(\hat{\Theta}(k-1), \Phi(k)) \\ Q_3(\hat{\Theta}(k-1), \Phi(k)) \end{bmatrix} \tag{A.7}$$

where

$$\begin{aligned} Q_1(\hat{\Theta}(k-1), \Phi(k)) := \begin{bmatrix} F^{-1}(k-1) & 0 \\ 0 & f^{-1}(k-1) I \end{bmatrix} M_1 \Phi(k) \\ \cdot \left(M_3^T \Phi(k) - M_2 \hat{\Theta}^T(k-1) M_1 \Phi(k) \right) \end{aligned} \tag{A.8}$$

$$Q_2(\hat{\Theta}(k-1), \Phi(k)) := \mathrm{Col}\left\{ M_4 \Phi(k) \Phi^T(k) M_4^T - F(k-1) \right\} \tag{A.9}$$

$$Q_3(\hat{\Theta}(k-1), \Phi(k)) := \Phi(k)^T M_5^T M_5 \Phi(k) - f(k-1) \tag{A.10}$$

and the matrices M_1, M_2, M_3, M_4 and M_5 are defined as follows

$$\phi(k) := \begin{bmatrix} \phi_e(k) \\ \phi_u(k) \\ \phi_\epsilon(k) \\ \phi_R(k) \end{bmatrix} = \underbrace{\begin{bmatrix} \text{diag}\left\{ \; I \quad \begin{bmatrix} 0 & I \end{bmatrix} \quad \begin{bmatrix} 0 & I \end{bmatrix} \quad I \; \right\} \quad 0 \end{bmatrix}}_{M_1} \Phi(k)$$

$$\hat{\theta}(k) := \begin{bmatrix} \hat{\theta}_A(k) \\ \hat{\theta}_B(k) \\ \hat{\theta}_C(k) \\ \hat{\theta}_M(k) \end{bmatrix} = \underbrace{\begin{bmatrix} I & 0 \end{bmatrix}}_{M_2} \hat{\Theta}(k)$$

$$e(k) = \underbrace{\begin{bmatrix} \theta_A^T & \begin{bmatrix} 0 & \theta_B^T \end{bmatrix} & \begin{bmatrix} 0 & \theta_C^T \end{bmatrix} & 0 & \theta_R^T & \theta_B^T \end{bmatrix}}_{M_3^T} \Phi(k)$$

$$\begin{bmatrix} \phi_e(k) \\ \phi_u(k) \\ \phi_\epsilon(k) \end{bmatrix} = \underbrace{\begin{bmatrix} \text{diag}\left\{ \; I \quad \begin{bmatrix} 0 & I \end{bmatrix} \quad \begin{bmatrix} 0 & I \end{bmatrix} \; \right\} \quad 0 \end{bmatrix}}_{M_4} \Phi(k)$$

$$\phi_R(k) = \underbrace{\begin{bmatrix} 0 & I & 0 \end{bmatrix}}_{M_5} \Phi(k).$$

The dimensions of identity and all zero matrices are omitted since they can be determined from the regressor and estimated parameters dimensions. Equation (A.7) shows that $Q(x, \phi)$ is Lipschitz continuous if $Q_1(x, \phi)$, $Q_2(x, \phi)$ and $Q_3(x, \phi)$ are all Lipschitz continuous. This can be shown by "triangle inequality"

$$|Q(x, \phi)| \leq |Q_1(x, \phi)| + |Q_2(x, \phi)| + |Q_2(x, \phi)|$$

which implies that

$$|Q(x_1, \phi_1) - Q(x_2, \phi_2)| \leq (\mathscr{K}_1 + \mathscr{K}_2 + \mathscr{K}_3) \{|x_1 - x_2| + |\phi_2 - \phi_2|\}$$

where $\mathscr{K}_1$, $\mathscr{K}_2$ and $\mathscr{K}_3$ are the Lipschitz constants of $Q_1(.,.)$, $Q_2(.,.)$ and $Q_3(.,.)$ in a bounded region of interest. With the same argument, we can claim that each Q_i – as a vector valued function – is Lipschitz continuous if each of its individual components is Lipschitz continuous. Equations (A.9) and (A.10) show that $Q_2(.,.)$ and $Q_3(.,.)$ are always Lipschitz continuous by Lemma 5. Moreover, $Q_1(.,.)$ is Lipschitz continuous as long as F is non singular and f is non zero. This can be shown by Lemma 5 and 6, and the fact that matrix inversion incorporates determinants of sub-regions of the original matrix (i.e. cofactors) and inverse of the matrix determinant. It is clear that the cofactor elements are simply polynomials and thus Lipschitz continuous, and the inverse of the matrix determinant is Lipschitz continuous by Lemma 6 as long as it is always nonsingular in $\mathscr{B}(x, \rho) \times \mathscr{B}(\phi, \nu)$.

Assumption C.2: It is clear that $A_\Phi(\hat{\Theta})$ and $B_\Phi(\hat{\Theta})$ given by (A.2) and (A.3) are both (Lipschitz) continuous in $\hat{\Theta}$.

Assumption C.3: By following the same approach as in the proof of Lemma 2 in Ljung (1977a), we can show that when the following conditions are satisfied

1. Assumption **C.1** holds
2. $\mathscr{K}$ in **C.1** is Lipschitz continuous in ϕ and ν

$$|\mathscr{K}(x,\phi_1,\rho,\nu_1)-\mathscr{K}(x,\phi_2,\rho,\nu_2)|\leq\mathscr{H}(x,\phi,\rho,\nu,w)\{|\phi_1-\phi_2|+|\nu_1-\nu_2|\}$$

   for $\phi_1,\phi_2\in\mathscr{B}(\phi,w)$ and $\nu_1,\nu_2\in\mathscr{B}(\nu,w)$

3. For all $x\in\mathscr{D}_R$, the functions Q, $\mathscr{K}$ and $\mathscr{H}$ have bounded p-moments for all $p>1$, $\lambda<1$ and $c<\infty$.

we have

$$\frac{1}{t}\sum_{k=1}^{t}Q\left(\bar{x},\bar{\phi}(k,\bar{x})\right)\rightarrow\frac{1}{t}\sum_{k=1}^{t}\mathbf{E}\left[Q\left(\bar{x},\bar{\phi}(k,\bar{x})\right)\right] \tag{A.11}$$

with probability one as $t\rightarrow\infty$. We have already shown that **C.1** holds and a Lipschitz continuous function $\mathscr{K}$ can be defined. The last condition holds as long as $|F(k)|$ and $|f(k)|$ are non-zero. This expression (A.11) with (A.5) when $\gamma(k)=1/k$ imply

$$z(t,\bar{x})\rightarrow\frac{1}{t}\sum_{k=1}^{t}\mathbf{E}\left[Q\left(\bar{x},\bar{\phi}(k,\bar{x})\right)\right]. \tag{A.12}$$

with probability one as $t\rightarrow\infty$. Therefore, in order to show that regularity condition (C.3) is satisfied, it is enough to prove that the limit

$$\mathscr{F}(\bar{x})=\lim_{t\rightarrow\infty}\frac{1}{t}\sum_{k=1}^{t}\mathbf{E}\left[Q\left(\bar{x},\bar{\phi}(k,\bar{x})\right)\right]$$

exists.

As we will show in the following part, the stochastic process $Q\left(\bar{x},\bar{\phi}(k,\bar{x})\right)$ is cyclostationary and the limit is well defined. Here, we assume that the excitation signal is also periodic with contents focused at frequencies different than the periodic disturbance. We take $\bar{\Theta}\in\mathscr{D}_R$ which implies that $A_\Phi(\bar{\Theta})$ is stable (i.e. has all eigenvalues strictly inside the unit circle). Since $A_\Phi(\bar{\Theta})$ and $B_\Phi(\bar{\Theta})$ are time invariant, we can use superposition principle for this linear time invariant system and decompose the response $\Phi(k)=\Phi_s(k)+\Phi_d(k)$ into stochastic, $\Phi_s(k)$, and deterministic, $\Phi_d(k)$, parts

$$\Phi_s(k)=A_\Phi(\bar{\Theta})\Phi_s(k-1)+B_{\Phi_s}(\bar{\Theta})\bar{w}(k) \tag{A.13}$$

$$\Phi_d(k)=A_\Phi(\bar{\Theta})\Phi_d(k-1)+B_{\Phi_d}(\bar{\Theta})\begin{bmatrix}u(k)\\ \phi_R(k)\end{bmatrix} \tag{A.14}$$

where B_{Φ_s} and B_{Φ_d} can be clearly found from (A.3). Note that when $u(k)$ is not periodic and rather it is wide-sense stationary the only required modification is to remove $u(k)$ from the input signal in (A.13) and instead append the input in (A.14) with $u(k)$. The rest of analysis will be very similar in this case.

Since $A_\Phi(\bar{\Theta})$ is stable, $\Phi_s(k)$ will converge to a zero mean stationary random vector, denoted by $\bar{\Phi}_s(k,\bar{\Theta})$, with bounded covariance and $\Phi_d(k)$ will converge to $\bar{\Phi}_d(k,\bar{\Theta})$ which is a periodic vector that contains sinusoids with the same frequencies as $\phi_R(k)$ and amplitudes/phase-lags that are related to the magnitude and phase of

$$\left(e^{jw_i}I - A_\Phi(\bar{\Theta})\right)^{-1} B_{\Phi_d}(\bar{\Theta})e^{jw_i}.$$

Again, since $A_\Phi(\bar{\Theta})$ is stable, $\bar{\Phi}_d(k,\bar{\Theta})$ is bounded and it implies that $\Phi(k)$ will converge to $\bar{\Phi}(k,\bar{\Theta}) = \bar{\Phi}_s(k,\bar{\Theta}) + \bar{\Phi}_d(k,\bar{\Theta})$. We can now return to (A.12) and show that

$$\lim_{t\to\infty} z(t,\bar{x}) = \lim_{t\to\infty} \frac{1}{t}\sum_{k=1}^{t} \mathbf{E}\left[Q\left(\bar{\Theta},\bar{\Phi}(k,\bar{\Theta})\right)\right] = h\left(\bar{\Theta}\right) \tag{A.15}$$

where $h\left(\bar{\Theta}\right)$ is a bounded and well defined function. Note that

$$\mathbf{E}\left[Q\left(\bar{\Theta},\bar{\Phi}(k,\bar{\Theta})\right)\right] = \begin{bmatrix} \begin{bmatrix} \bar{F}^{-1} & 0 \\ 0 & \bar{f}^{-1} \end{bmatrix} \mathbf{E}\left\{\bar{\phi}(k,\bar{\Theta})\left(\bar{y}(k,\bar{\Theta}) - \bar{\theta}^T\bar{\phi}(k,\bar{\Theta})\right)\right\} \\ Col\left\{\mathbf{E}\left[\bar{\phi}_1(k,\bar{\Theta})\bar{\phi}_1^T(k,\bar{\Theta})\right] - \bar{F}\right\} \\ \mathbf{E}\left[\bar{\Phi}(k,\bar{\Theta})^T M_5^T M_5 \Phi(k) - \bar{f}\right] \end{bmatrix} \tag{A.16}$$

and we can show that the limit given in A.15 is well defined for all the three terms on the right hand side. For instance, for the first term we have

$$\begin{aligned} &\mathbf{E}\left\{\bar{\phi}(k,\bar{\Theta})\left(\bar{y}(k,\bar{\Theta}) - \bar{\theta}^T\bar{\phi}(k,\bar{\Theta})\right)\right\} = \mathbf{E}\left\{\begin{bmatrix} \bar{\phi}_1(k,\bar{\Theta}) \\ \phi_R(k) \end{bmatrix}\left(\bar{y}(k,\bar{\Theta}) - \bar{\theta}^T\bar{\phi}(k,\bar{\Theta})\right)\right\} \\ &= \begin{bmatrix} \mathbf{E}\left[\bar{\phi}_{1s}(k,\bar{\Theta})\bar{y}_s(k,\bar{\Theta})\right] + \bar{\phi}_{1d}(k,\bar{\Theta})\bar{y}_d^T(k,\bar{\Theta}) \\ \phi_R(k)\bar{y}_d(k,\bar{\Theta}) \end{bmatrix} \\ &\quad - \begin{bmatrix} \bar{\theta}_1^T\mathbf{E}\left[\bar{\phi}_{1s}(k,\bar{\Theta})\bar{\phi}_{1s}(k,\bar{\Theta})\right] + \bar{\theta}_1^T\bar{\phi}_{1d}(k,\bar{\Theta})\bar{\phi}_{1d}^T(k,\bar{\Theta}) + \bar{\theta}_M^T\phi_R(k)\bar{\phi}_{1,d}(k,\bar{\Theta}) \\ \phi_R(k)\bar{\Phi}_d^T(k,\bar{\Theta})\bar{\theta} \end{bmatrix} \end{aligned} \tag{A.17}$$

where $\bar{\phi}_{1s}(.,.) := M_4\bar{\Phi}_s(.,.)$, $\bar{\phi}_{1d}(.,.) := M_4\bar{\Phi}_d(.,.)$ and $\bar{\theta}_1^T := \begin{bmatrix}\bar{\theta}_A^T & \bar{\theta}_B^T & \bar{\theta}_C^T\end{bmatrix}$. Similarly, $\bar{y}_d(.,.)$ and $\bar{y}_s(.,.)$ are the values corresponding to y respectively in $\bar{\Phi}_d(.,.)$ and $\bar{\Phi}_s(.,.)$. Note that all the stochastic terms will be stationary when $t\to\infty$ and all deterministic terms become periodic in steady state. Therefore, all the terms in (A.17) when plugged to (A.15) produces well defined limits. Following the same type of analysis, it is easy to show that the second and third terms of (A.16) generate bounded limits when plugged to (A.15).

Assumptions C.4: The following analogies between (A.5) and (A.6) can be exploited to take the same path as the previous part to show that assumption **C.4** holds.

$$\begin{aligned} z(k,\bar{x}) &\leftrightarrow k_v(k,\bar{x},\lambda,c) \\ Q(k,\bar{x},\bar{\phi}(k,\bar{x})) &\leftrightarrow \mathscr{K}\left(\bar{x},\bar{\phi}(k,\bar{x}),\rho(\bar{x}),\nu(k,\lambda,c)\right) \\ \mathscr{K}\left(\bar{x},\bar{\phi}(k,\bar{x}),\rho(\bar{x}),\nu(k,\lambda,c)\right) &\leftrightarrow \mathscr{H}\left(\bar{x},\bar{\phi}(k,\bar{x}),\rho(\bar{x}),\nu(k,\lambda,c),w\right) \end{aligned}$$

Assumptions C.5 and C.6: These assumptions are the same as assumption 2 in theorem 5.

A.3 Convergence Analysis

Under the regularity conditions mentioned above, the parameter adaptation algorithm (6.21) can be associated with a differential equation that will be derived in this sections. Suppose the estimate sequence of $\hat{\theta}(k)$ is fixed to a constant value $\hat{\theta}$. Denote $\bar{e}(t,\hat{\theta})$, $\bar{\bar{y}}(t,\hat{\theta})$ and $\bar{\theta}_M(t,\hat{\theta})$ as the (cyclo) stationary processes defined by (6.17), (6.18) and (6.15). Note that the stationarity of these processes is admissible only when the closed loop system determined by $\hat{\theta}$ is stable. Suppose $\mathscr{D}_s$ is the set of all values of $\hat{\theta}$ that the stability is attained. The stationary $\hat{\theta}_D$, by (6.24), is

$$\bar{\hat{\theta}}_D(\hat{\theta}) = \frac{-\alpha}{1-\beta}\hat{D}_B^{-T}(\hat{\theta})\hat{\theta}_M$$

where $\hat{D}_B(\hat{\theta})$ denotes the value of $\hat{D}_B(k)$ when $\hat{\theta}(k) = \hat{\theta}$. Accordingly the control signal associated with $\hat{\theta}$ is

$$\bar{u}_A(k) = \frac{-\alpha}{1-\beta}\hat{\theta}_M^T\hat{D}_B^{-1}(\hat{\theta})\phi_R(k).$$

Therefore, for any fixed $\hat{\theta}$, the adaptive control algorithm operates in open loop mode and generates a pure feedforward signal. This implies that the closed loop system is stable as long as $\hat{\theta}_A$ corresponds to an $\hat{A}(q^{-1})$ polynomial that has all roots strictly inside the unit circle

$$\mathscr{D}_s := \left\{\hat{\theta}: \ -\hat{\theta}_A^T = [\hat{a}_1,\cdots,\hat{a}_{n_A}],\ 1+\hat{a}_1 q+\cdots+\hat{a}_{n_A}q^{n_A}=0 \Rightarrow |q|>1\right\}.$$

This condition is equivalent to assumption 3 in theorem 5. The control signal is bounded as long as $\hat{\theta}_M$ and $\hat{D}_B^{-1}(\hat{\theta})\phi_R(k)$ are bounded. The second criterion is equivalent to the condition that $\hat{B}(q^{-1})$ associated with $\hat{\theta}_B$ has non-zero magnitude at ω_i's which is true as long as assumption 4 of theorem 5 holds. We have

$$\begin{aligned} \bar{e}(t,\hat{\theta}) &= \theta_A\phi_e(t,\hat{\theta}) + \theta_B^T\phi_u(t) + \theta_C^T\phi_{\bar{w}}(t) + \bar{\theta}_M^T(t,\hat{\theta})\phi_R(t) + \bar{w}(t) \\ \bar{\bar{y}}(t,\hat{\theta}) &= \hat{\theta}_A^T\phi_e(t,\hat{\theta}) + \hat{\theta}_B^T\phi_u(t) + \hat{\theta}_C^T\phi_\epsilon(t,\hat{\theta}) + \hat{\theta}_M^T\phi_R(t,\hat{\theta}) \end{aligned}$$

In the first expression, we have used $w_t(t,\hat{\theta}) = 0$ since the term $(z-1)[\theta(k)]$ in lemma 4 is zero once $\hat{\theta}(k)$ is frozen. The stationary estimation error defined in (6.19) is

$$\begin{aligned}\bar{\epsilon}(t,\hat{\theta}) &= \bar{e}(t,\hat{\theta}) - \bar{\hat{y}}(t,\hat{\theta}) \\ &= \left(\theta_A^T - \hat{\theta}_A^T\right)\phi_e(t,\hat{\theta}) + \left(\theta_B^T - \hat{\theta}_B^T\right)\phi_u(t) + \left(\theta_C^T - \hat{\theta}_C^T\right)\phi_\epsilon(t,\hat{\theta}) - \theta_C\phi_\epsilon(t,\hat{\theta}) \\ &\qquad + \theta_C^T\phi_{\bar{w}}(t) + \bar{\theta}_M^T(t,\hat{\theta})\phi_R(t) - \hat{\theta}_M^T\phi_R(t,\hat{\theta}) + \bar{w}(t).\end{aligned}$$

By using $\theta_C^T\phi_{\bar{w}}(t) + \bar{w}(t) = C(q^{-1})\bar{w}(t)$ and $\theta_C^T\phi_\epsilon(t,\hat{\theta}) + \bar{\epsilon}(t,\hat{\theta}) = C(q^{-1})\bar{\epsilon}(t,\hat{\theta})$ we have

$$\begin{aligned}\bar{\epsilon}(t,\hat{\theta}) &= \frac{1}{C(q^{-1})}\Big[\left(\theta_A^T - \hat{\theta}_A^T\right)\phi_e(t,\hat{\theta}) + \left(\theta_B^T - \hat{\theta}_B^T\right)\phi_u(t) + \left(\theta_C^T - \hat{\theta}_C^T\right)\phi_\epsilon(t,\hat{\theta}) \\ &\qquad + \bar{\theta}_M^T(t,\hat{\theta})\phi_R(t) - \hat{\theta}_M^T\phi_R(t,\hat{\theta})\Big] + \bar{w}(t) \\ &= \frac{1}{C(q^{-1})}\begin{bmatrix}\phi_e(t,\hat{\theta}) \\ \phi_u(t) \\ \phi_\epsilon(t,\hat{\theta}) \\ \phi_R(t)\end{bmatrix}^T\begin{bmatrix}\theta_A - \hat{\theta}_A \\ \theta_B - \hat{\theta}_B \\ \theta_C - \hat{\theta}_C \\ \bar{\theta}_M(t,\hat{\theta}) - \hat{\theta}_M\end{bmatrix}\end{aligned} \tag{A.18}$$

The stationary process $\bar{\theta}_M(t,\hat{\theta})$ can be derived from (6.15) and (6.24)

$$\begin{aligned}\theta_M(k+1) &= D_B^T\hat{\theta}_D(k+1) + \theta_R \\ &= \beta D_B^T\hat{\theta}_D(k) - \alpha D_B^T\hat{D}_B^{-T}(k)\hat{\theta}_M(k) + \theta_R \\ &= \beta\left(D_B^T(k)\hat{\theta}_D(k) + \theta_R\right) - \alpha D_B^T\hat{D}_B^{-T}(k)\hat{\theta}_M(k) + (1-\beta)\theta_R \\ &= \beta\theta_M(k) - \alpha\Delta^T(\hat{\theta})\hat{\theta}_M(k) + (1-\beta)\theta_R\end{aligned}$$

where $\Delta(\hat{\theta})$ is the stationary matrix for $\Delta(k)$ and

$$\begin{aligned}\Delta(k) :&= \hat{D}_B^{-1}(k)D_B \\ &= \begin{bmatrix}\hat{D}_{B1}^{-1}(k) & 0 & \cdots & 0 \\ 0 & \hat{D}_{B2}^{-1}(k) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \hat{D}_{Bn}^{-1}(k)\end{bmatrix}\begin{bmatrix}D_{B1} & 0 & \cdots & 0 \\ 0 & D_{B2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & D_{Bn}\end{bmatrix} \\ &= \begin{bmatrix}\hat{D}_{B1}^{-1}(k)D_{B1} & 0 & \cdots & 0 \\ 0 & \hat{D}_{B2}^{-1}(k)D_{B2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \hat{D}_{Bn}^{-1}(k)D_{Bn}\end{bmatrix}\end{aligned}$$

$$\begin{aligned}\hat{D}_{Bi}^{-1}(k)D_{Bi} &= \left(\frac{1}{\hat{m}_{Bi}}R\left(-\hat{\delta}_{Bi}(k)\right)\right)\left(m_{Bi}R\left(\delta_{Bi}\right)\right) \\ &= \frac{m_{Bi}}{\hat{m}_{Bi}(k)}R(\delta_{Bi} - \hat{\delta}_{Bi}(k))\end{aligned}$$

Therefore,

$$\begin{aligned}\bar{\theta}_M(t,\hat{\theta}) &= \beta\bar{\theta}_M(t,\hat{\theta}) - \alpha\Delta^T(\hat{\theta})\hat{\theta}_M + (1-\beta)\,\theta_R \\ &= \theta_R - \frac{\alpha}{1-\beta}\Delta^T(\hat{\theta})\hat{\theta}_M.\end{aligned}$$

Using this definition and the expression for $\bar{\theta}_M$ we have

$$\bar{\theta}_M^T(t,\hat{\theta})\phi_R(t) - \hat{\theta}_M^T\phi_R(t,\hat{\theta}) = \phi_R^T(t,\hat{\theta})\left(I + \frac{\alpha}{1-\beta}\Delta^T(\hat{\theta})\right)\left[\theta_M^* - \hat{\theta}_M\right]$$

where

$$\theta_M^* = \left(I + \frac{\alpha}{1-\beta}\Delta(\hat{\theta})\right)^{-T}\theta_R. \tag{A.19}$$

The expression $I + \frac{\alpha}{1-\beta}\Delta(\hat{\theta})$ has a singularity point at

$$\begin{aligned}\hat{\delta}_{B_i} &= \delta_{B_i} - \pi \\ \hat{m}_{B_i} &= \frac{\alpha}{1-\beta}m_{B_i}.\end{aligned} \tag{A.20}$$

By choosing appropriate α and β and limiting $\hat{m}_{B_i}$ from above, the singularity point can be avoided. This condition is guaranteed by assumption 4 in theorem 5. However, as we will see in the sequel, the algorithm avoids such a large phase difference $\delta_{B_i} - \hat{\delta}_{B_i} = \pi$ with probability one even when no upper bound is applied to $\hat{m}_{B_i}$.

Returning to (A.18), we have

$$\bar{\epsilon}(t,\hat{\theta}) = \frac{1}{C(q^{-1})}\left[\begin{array}{c}\phi_e(t,\hat{\theta})\\ \phi_u(t)\\ \phi_\epsilon(t,\hat{\theta})\\ \hdashline \phi_R(t)\end{array}\right]^T \left[\begin{array}{c:c} I & 0 \\ \hdashline 0 & I + \frac{\alpha}{1-\beta}\bar{\Delta}^T(\hat{\theta})\end{array}\right]\left[\begin{array}{c}\theta_A - \hat{\theta}_A\\ \theta_B - \hat{\theta}_B\\ \theta_C - \hat{\theta}_C\\ \hdashline \theta_M^* - \hat{\theta}_M\end{array}\right] \tag{A.21}$$

When Δ is bounded, by choosing β close enough to 1 and $\alpha \gg 1-\beta$ the inverse term is not subjected to numerical issues and $\theta_M^* \ll \theta_R$.

Introduce

$$\tilde{G}(\hat{\theta}) := \frac{1}{t}\sum_{k=1}^{t}\left(\left[\begin{array}{c}\phi_e(k,\hat{\theta})\\ \phi_u(k)\\ \phi_\epsilon(k,\hat{\theta})\\ \phi_R(k)\end{array}\right]\left[\begin{array}{c:c}\frac{1}{C(q^{-1})} & 0\\ \hdashline 0 & \frac{1+H(q^{-1},\hat{\theta})}{C(q^{-1})}\end{array}\right]\left[\begin{array}{c}\phi_e(k,\hat{\theta})\\ \phi_u(k)\\ \phi_\epsilon(k,\hat{\theta})\\ \phi_R(k)\end{array}\right]^T\right) \tag{A.22a}$$

$$G(\hat{\theta}) := \frac{1}{t}\sum_{k=1}^{t}\left[\begin{array}{c}\phi_e(k,\hat{\theta})\\ \phi_u(k)\\ \phi_\epsilon(k,\hat{\theta})\end{array}\right] \tag{A.22b}$$

$$g(\hat{\theta}) := \frac{1}{t}\sum_{k=1}^{t}\phi_R^T(k)\phi_R(k) \tag{A.22c}$$

and decompose $\phi_e(k)$ to a stochastic and deterministic part as $\phi_e(k) = \phi_{ed}(k) + \phi_{es}(k)$. The fact that all stochastic processes are zero mean along with (A.11) imply

$$\tilde{G}(\hat{\theta}) \to \tilde{G}_s(\hat{\theta}) + \tilde{G}_d(\hat{\theta}) \qquad \text{w.p. 1 as } t \to \infty \tag{A.23}$$

$$\tilde{G}_s(\hat{\theta}) := \mathbf{E}\left(\begin{bmatrix} \phi_{es}(k,\hat{\theta}) \\ 0 \\ \phi_\epsilon(k,\hat{\theta}) \\ 0 \end{bmatrix} \frac{1}{C(q^{-1})} \begin{bmatrix} \phi_{es}(k,\hat{\theta}) \\ 0 \\ \phi_\epsilon(k,\hat{\theta}) \\ 0 \end{bmatrix}^T \right) \tag{A.24}$$

$$\tilde{G}_d(\hat{\theta}) := \frac{1}{t}\sum_{k=1}^{t} \left(\begin{bmatrix} \phi_{ed}(k,\hat{\theta}) \\ \phi_u(k) \\ 0 \\ \bar{\phi}_R(k) \end{bmatrix} \left[\begin{array}{c|c} \frac{1}{C(q^{-1})} & 0 \\ \hline 0 & \frac{1+H(q^{-1};\hat{\theta})}{C(q^{-1})} \end{array} \right] \begin{bmatrix} \phi_{ed}(k,\hat{\theta}) \\ \phi_u(k) \\ 0 \\ \bar{\phi}_R(k) \end{bmatrix}^T \right) \tag{A.25}$$

Here, $H(q^{-1},\hat{\theta})$ is a stable and causal transfer function, e.g. an FIR LTI system, that at ω_i's has magnitude and phase responses equal to the corresponding magnitudes and phases in $\frac{\alpha}{1-\beta}\Delta^T(\hat{\theta})$. It is trivial that such an LTI transfer function can be found for any $\Delta^T(\hat{\theta})$ by considering at most $2n$ coefficients.

Under the regularity conditions and conditions on the stochastic processes $\bar{w}(k)$ and $u(k)$ that were mentioned above the parameter adaptation algorithm in (6.21) can be associated with the differential equation Ljung (1977a)

$$\frac{d}{dt}\hat{\theta}(t) = \begin{bmatrix} F^{-1}(t) & 0 \\ 0 & f^{-1}(t)I \end{bmatrix} \tilde{G}(\hat{\theta}(t)) \begin{bmatrix} \theta_A - \hat{\theta}_A \\ \theta_B - \hat{\theta}_B \\ \theta_C - \hat{\theta}_C \\ \theta_M^* - \hat{\theta}_M \end{bmatrix} \tag{A.26a}$$

$$\frac{d}{dt}F(t) = G(\hat{\theta}(t)) - F(t) \tag{A.26b}$$

$$\frac{d}{dt}f(t) = g(\hat{\theta}(t)) - f(t). \tag{A.26c}$$

Moreover, the convergence point(s) of parameter adaptation algorithm (6.21) can be related to the set of stationary points, say $\mathscr{D}_c$, of the differential equations (A.26) through corollary 1 in Ljung (1977a). The following conditions, that are already shown to be satisfied partially, are required to use the corollary results

- Regularity conditions mentioned above should be satisfied.
- $\hat{\theta}(k) \in \mathscr{D}_s$ infinitely often with probability one.
- $\|\phi(k)\|$ is bounded infinitely often whenever $\hat{\theta}(k)$ belongs to $\mathscr{D}_s$

- There is a Lyapunov function $V\left(\hat{\theta}(t), F(t), f(t)\right)$ such that

$$\frac{d}{dt}V\left(\hat{\theta}(t), F(t), f(t)\right) \leq 0 \quad \text{for } \hat{\theta}(t) \in \mathscr{D}_s,\ F(t) > 0,\ f(t) > 0$$
$$\frac{d}{dt}V\left(\hat{\theta}(t), F(t), f(t)\right) = 0 \quad \text{for } \hat{\theta}(t) \in \mathscr{D}_c$$

We have already shown that the first condition is satisfied when a projection scheme is considered. The third condition can be satisfied by restricting $\hat{\theta}_B(k)$ to

$$\mathscr{D}_B := \left\{ \hat{\theta}_B : |b_1 e^{-j\omega_m} + \cdots + b_{n_A} e^{-jn_A\omega_m}| > 0 \right. \\ \left. \forall m \in \{1, \ldots, n\} \right\}.$$

This means that the magnitude of $\hat{B}(q^{-1}; k)$ at excitation frequencies should be bounded from below by a small positive value. This is equivalent to assumption 4 in theorem 5. In case $\hat{\theta}_B(k)$ does not belong to $\mathscr{D}_B$, the parameters can be projected to $\mathscr{D}_B$ by scaling $\hat{\theta}_B(k)$. An implication of this restriction is that $\Delta(\hat{\theta}(k))$ is also always bounded.

Analogous to Ljung (1977b) an admissible Lyapunov function can be defined when $\tilde{G}(\hat{\theta}) + \tilde{G}^T(\hat{\theta})$ is positive definite. This is possible when both $\frac{1}{C(q^{-1})}$ and $\frac{H(q^{-1},\hat{\theta})}{C(q^{-1})}$ are strictly positive real. The second condition can be relaxed further since the response of $H(q^{-1}; \hat{\theta})$ is only important at ω_i's. Therefore, $\tilde{G}(\hat{\theta}) + \tilde{G}^T(\hat{\theta})$ is positive definite if

1. ϕ_u is persistently exciting of order $2n$.
2. $\frac{1}{C(q^{-1})}$ is strictly positive real (SPR).
3. Absolute phase of $\frac{H(e^{-j\omega_i};\hat{\theta})}{C(e^{-j\omega_i})}$ is less than 90 degrees.

The third condition can be justified by comparing (A.22a) with

$$\begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix} \begin{bmatrix} \frac{1}{C} & 0 \\ 0 & \frac{1+H}{C} \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix}^T = \begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix} \frac{1}{C} \begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix}^T + \begin{bmatrix} 0 & 0 \\ 0 & \phi_2 \frac{H}{C} \phi_2^T \end{bmatrix}.$$

If $1/C(q^{-1})$ is strictly positive real, the first term on the left will make a positive definite matrix when it is summed with its transpose. Similarly, when $H(q^{-1})/C(q^{-1})$ is SPR, summation of the second term with its transpose results in a positive semi-definite matrix. This implies that the summation of the two transfer functions results in a positive definite matrix when condition 2 and 3 hold. Note that these conditions are equivalent to assumptions 1,5 and 6 of theorem 5 respectively.

Proof **(Theorem 5)**
Theorem 2 in Ljung (1977a) implies that the convergence point(s) of the parameter adaptation algorithm (6.21) are among the stable stationary points of (A.26). For an SPR $1/C(q^{-1})$, the cross covariance matrix

$$\frac{1}{t}\sum_{k=1}^{t}\left(\begin{bmatrix}\phi_e(k,\hat{\theta})\\ \phi_u(k)\\ \phi_\epsilon(k,\hat{\theta})\\ \phi_R(k)\end{bmatrix}\frac{1}{C(q^{-1})}\begin{bmatrix}\phi_e(k,\hat{\theta})\\ \phi_u(k)\\ \phi_\epsilon(k,\hat{\theta})\\ \phi_R(k)\end{bmatrix}^T\right)$$

added to its transpose makes a positive definite matrix. Therefore, from (A.21), the only stationary point of (A.26a) is given by

$$\left[\begin{array}{c|c} I & 0 \\ \hline 0 & I+\frac{\alpha}{1-\beta}\bar{\Delta}^T(\hat{\theta}) \end{array}\right]\begin{bmatrix}\theta_A-\hat{\theta}_A\\ \theta_B-\hat{\theta}_B\\ \theta_C-\hat{\theta}_C\\ \hline \theta_M^*-\hat{\theta}_M\end{bmatrix}=0.$$

The left matrix singularity points are described in (A.20). It is trivial that the parameter $\hat{\theta}_B$ corresponding to such a singularity point does not fall in the null space of the left matrix. Hence, the only stationary point of (A.26) is

$$\theta^*=\begin{bmatrix}\hat{\theta}_A\\ \hat{\theta}_B\\ \hat{\theta}_C\\ \hat{\theta}_M\end{bmatrix}=\begin{bmatrix}\theta_A\\ \theta_B\\ \theta_C\\ \frac{1-\beta}{1-\beta+\alpha}\theta_R\end{bmatrix}$$

where the last term comes from (A.19) when $\hat{\theta}_B=\theta_B$. Moreover, it can be shown that this is a locally stable stationary point. Suppose $\mathscr{B}(\theta_B,\rho(\theta_B))$ be a neighborhood of θ_B in which $|\delta_{B_i}-\hat{\delta}_{B_i}|<90$. The quadratic function

$$V(\hat{\theta}):=\hat{\theta}^T\left(\tilde{G}(\hat{\theta})+\tilde{G}^T(\hat{\theta})\right)\hat{\theta}.$$

is positive definite and has negative definite Lie derivative along (A.26) trajectories as long as $\hat{\theta}\in\mathscr{B}(\theta^*,\rho(\theta_B))$. This is because $\|\theta_B-\hat{\theta}_B\|\leq\|\theta^*-\hat{\theta}\|$ implies that starting the trajectory from $\hat{\theta}(0)\in\mathscr{B}(\theta^*,\rho(\theta_B))$, the estimated $\hat{\theta}_B$ always stays in a neighborhood with a smaller radius around θ_B. Therefore, decreasing $\|\theta^*-\hat{\theta}\|$ results in keeping $\hat{\theta}_B$ in $\mathscr{B}(\theta^*,\rho(\hat{\theta}_B))$ which itself implies more reduction in $\|\theta^*-\hat{\theta}\|$. Accordingly, $V(\hat{\theta})$ is a Lyapunov function for (A.26) and the equilibrium state θ^* is stable in the sense of Lyapunov.

Proof **(Theorem 6)**
The proof of theorem 6 is similar to the proof of theorem 5 up to equation (A.21). Starting from (A.22), one needs to substitute all instances of $C(q^{-1})$ by $C(q^{-1})/\bar{C}(q^{-1})$ to show the results.

Proof **(Theorem 7)**
The proof is based on to the proof of theorem 5. When there is no prior knowledge regarding $B(q^{-1})$ *in hand, it may not be possible to set the initial conditions of* (6.21) *and accordingly* (A.26) *to guarantee SPR condition on* $H(q^{-1};\hat{\theta})/C(q^{-1})$ *term. However, the region of attraction for differential equation* (A.26) *can be enlarged in practice by increasing the excitation energy. Consider the upper states of* (A.26a)

$$\frac{d}{dt}\begin{bmatrix}\hat{\theta}_A\\ \hat{\theta}_B\\ \hat{\theta}_C\end{bmatrix} = F^{-1}(t)\frac{1}{t}\sum_{k=1}^{t}\left(\begin{bmatrix}\phi_e(k,\hat{\theta})\\ \phi_u(k)\\ \phi_\epsilon(k,\hat{\theta})\end{bmatrix}\frac{1}{C(q^{-1})}\begin{bmatrix}\phi_e(k,\hat{\theta})\\ \phi_u(k)\\ \phi_\epsilon(k,\hat{\theta})\end{bmatrix}^T\right)\begin{bmatrix}\theta_A-\hat{\theta}_A\\ \theta_B-\hat{\theta}_B\\ \theta_C-\hat{\theta}_C\end{bmatrix} + F^{-1}(t)\frac{1}{t}\sum_{k=1}^{t}\begin{bmatrix}\phi_e(k,\hat{\theta})\frac{1+H(q^{-1};\hat{\theta})}{C(q^{-1})}\phi_R^T(k)\\ 0\\ 0\end{bmatrix}\theta_M^*-\hat{\theta}_M \tag{A.27}$$

when $1/C(q^{-1})$ *is SPR, the first term on the right attracts the states to* $\begin{bmatrix}\theta_A^T & \theta_B^T & \theta_C^T\end{bmatrix}^T$. *However, the second term might be repulsive when* $H(q^{-1};\hat{\theta})/C(q^{-1})$ *is not SPR. In* (A.27), *the second term on the right hand side is bounded when assumption 10 in theorem 7 holds true. As a result, by increasing* α_{dc} *the attraction of the first term increases which can lead to bringing* $\hat{\theta}_B$ *into* $\mathscr{D}_B$ *when* α_{dc} *is chosen large enough. This means that by choosing enough excitation in such a case, the convergence rate of the first term can be increased such that the first term brings* $\hat{\theta}_B$ *into a vicinity of* θ_B *in which the second term is also attractive.*

Appendix B

Model Parameters

The transfer function coefficients for a set of models mentioned in the previous sections are given in this appendix. All transfer functions are in the form of

$$G(z) = \frac{b_1 z^n + b_2 z^{n-1} + b_3 z^{n-2} + \cdots + b_n z + b_{n+1}}{a_1 z^n + a_2 z^{n-1} + a_3 z^{n-2} + \cdots + a_n z + b_{n+1}}$$

and in most (if not all) of the cases the denumerator leading coefficient is $a_1 = 1$.

Table B.1: Transfer function coefficients for closed loop model from the MA input to the PES

a_1	a_2	a_3	a_4	a_5
$1.000e+00$	$7.243e-01$	$-3.290e-01$	$-6.334e-01$	$7.694e-01$
a_6	a_7	a_8	a_9	a_{10}
$1.276e+00$	$9.027e-01$	$-2.364e-01$	$-3.912e-01$	$1.266e-01$
a_{11}	a_{12}	a_{13}	a_{14}	a_{15}
$1.091e+00$	$6.582e-01$	$-1.468e-01$	$-2.107e-01$	$2.325e-01$
a_{16}	a_{17}	a_{18}		
$1.226e-01$	$1.200e-01$	$1.620e-01$		
b_1	b_2	b_3	b_4	b_5
$0.000e+00$	$-3.164e-02$	$8.523e-03$	$2.950e-03$	$2.652e-02$
b_6	b_7	b_8	b_9	b_{10}
$-2.923e-02$	$-4.232e-03$	$-1.080e-02$	$2.798e-02$	$9.970e-03$
b_{11}	b_{12}	b_{13}	b_{14}	b_{15}
$4.318e-03$	$-1.362e-02$	$4.054e-03$	$1.650e-02$	$2.123e-03$
b_{16}	b_{17}	b_{18}		
$4.095e-03$	$6.023e-04$	$4.392e-03$		

Table B.2: Transfer function coefficients for closed loop model from the MA input to the PES

a_1	a_2	a_3	a_4	a_5
$1.000e+00$	$-4.249e+00$	$9.056e+00$	$-1.227e+01$	$1.155e+01$
a_6	a_7	a_8	a_9	a_{10}
$-8.484e+00$	$6.469e+00$	$-6.604e+00$	$6.817e+00$	$-5.507e+00$
a_{11}	a_{12}	a_{13}	a_{14}	a_{15}
$3.391e+00$	$-2.030e+00$	$2.176e+00$	$-3.511e+00$	$5.012e+00$
a_{16}	a_{17}	a_{18}	a_{19}	a_{20}
$-4.943e+00$	$2.567e+00$	$2.486e-01$	$-4.136e-01$	$-2.320e+00$
a_{21}	a_{22}	a_{23}	a_{24}	a_{25}
$4.593e+00$	$-3.730e+00$	$1.210e+00$	$-4.791e-01$	$2.409e+00$
a_{26}	a_{27}	a_{28}	a_{29}	a_{30}
$-4.765e+00$	$5.322e+00$	$-4.095e+00$	$2.397e+00$	$-1.184e+00$
a_{31}	a_{32}	a_{33}	a_{34}	a_{35}
$7.326e-01$	$-9.286e-01$	$1.178e+00$	$-7.553e-01$	$-2.222e-01$
a_{36}	a_{37}	a_{38}	a_{39}	a_{40}
$8.061e-01$	$-5.051e-01$	$-9.935e-03$	$-1.781e-01$	$9.067e-01$
a_{41}	a_{42}	a_{43}	a_{44}	a_{45}
$-1.128e+00$	$2.591e-01$	$1.225e+00$	$-2.437e+00$	$2.866e+00$
a_{46}	a_{47}	a_{48}	a_{49}	a_{50}
$-2.558e+00$	$1.841e+00$	$-1.035e+00$	$4.007e-01$	$-7.961e-02$
a_{51}				
$7.253e-04$				
b_1	b_2	b_3	b_4	b_5
$0.000e+00$	$5.728e-04$	$-3.502e-03$	$2.181e-02$	$-2.307e-02$
b_6	b_7	b_8	b_9	b_{10}
$1.184e-02$	$2.428e-03$	$3.693e-02$	$-1.017e-01$	$1.165e-01$
b_{11}	b_{12}	b_{13}	b_{14}	b_{15}
$-6.439e-02$	$-9.396e-03$	$2.662e-02$	$-2.232e-02$	$1.158e-02$
b_{16}	b_{17}	b_{18}	b_{19}	b_{20}
$-9.251e-03$	$8.612e-04$	$-2.658e-03$	$1.715e-02$	$2.690e-03$
b_{21}	b_{22}	b_{23}	b_{24}	b_{25}
$-5.162e-02$	$8.742e-02$	$-4.887e-02$	$-1.704e-02$	$4.850e-02$
b_{26}	b_{27}	b_{28}	b_{29}	b_{30}
$-2.695e-02$	$4.293e-03$	$-1.252e-02$	$2.813e-02$	$-3.110e-02$
b_{31}	b_{32}	b_{33}	b_{34}	b_{35}
$1.957e-02$	$-1.902e-02$	$1.667e-02$	$-1.080e-02$	$1.109e-02$
b_{36}	b_{37}	b_{38}	b_{39}	b_{40}
$-3.447e-02$	$6.441e-02$	$-6.432e-02$	$3.449e-02$	$-5.880e-03$
b_{41}	b_{42}	b_{43}	b_{44}	b_{45}
$-1.515e-03$	$2.801e-03$	$-1.587e-02$	$2.739e-02$	$-2.851e-02$
b_{46}	b_{47}	b_{48}	b_{49}	b_{50}
$2.176e-02$	$-2.220e-02$	$2.373e-02$	$-1.916e-02$	$9.094e-03$
b_{51}				

Bibliography

Åström, K. (1980). Direct methods for nonminimum phase systems. In *1980 19th IEEE Conference on Decision and Control including the Symposium on Adaptive Processes*, number 19, pages 611–615.

Abramovitch, D. and Franklin, G. (2002). A brief history of disk drive control. *IEEE Control Systems Magazine*, 22(3):28–42.

Akhtar, M. T., Abe, M., and Kawamata, M. (2006). A new variable step size lms algorithm-based method for improved online secondary path modeling in active noise control systems. *Audio, Speech, and Language Processing, IEEE Transactions on*, 14(2):720–726.

Al Mamun, A., Guo, G., and Bi, C. (2007). *Hard disk drive: mechatronics and control*, volume 23. CRC PressI Llc.

Albrecht, T., Arora, H., Ayanoor-Vitikkate, V., Beaujour, J.-M., Bedau, D., Berman, D., Bogdanov, A., Chapuis, Y.-A., Cushen, J., Dobisz, E., et al. (2015). Bit patterned magnetic recording: Theory, media fabrication, and recording performance.

Albrecht, T. R., Hellwing, O., Ruiz, R., Schabes, M. E., Terris, B. D., and Wu, X. Z. (2009). Bit-patterned magnetic recording: Nanoscale magnetic islands for data storage. In *Nanoscale Magnetic Materials and Applications*, pages 237–274. Springer.

Aphale, S. S., Devasia, S., and Moheimani, S. R. (2008). High-bandwidth control of a piezoelectric nanopositioning stage in the presence of plant uncertainties. *Nanotechnology*, 19(12):125503.

Åström, K. J. (1975). Theory and applications of self-tuning regulators. In *Control Theory, Numerical Methods and Computer Systems Modelling*, pages 669–680. Springer.

Bagherieh, O., Shahsavari, B., and Horowitz, R. (2015). Online identification of system uncertainties using coprime factorization with application to hard disk drives. In *ASME 2015 Dynamic Systems and Control Conference*. American Society of Mechanical Engineers.

Bagherieh, O., Shahsavari, B., Keikha, E., and Horowitz, R. (2014). Observer design for non-uniform sampled systems using gain-scheduling. In *ASME 2014 Conference on Information Storage and Processing Systems*, pages V001T03A004–V001T03A004. American Society of Mechanical Engineers.

Bai, E.-W., Fu, L.-C., and Sastry, S. S. (1988). Averaging analysis for discrete time and sampled data adaptive systems. *Circuits and Systems, IEEE Transactions on*, 35(2):137–148.

Bengtsson, G. (1977). Output regulation and internal modelsâĂŤa frequency domain ap-

proach. *Automatica*, 13(4):333–345.

Beranek, L. L. and Ver, I. L. (1992). Noise and vibration control engineering-principles and applications. *Noise and vibration control engineering-Principles and applications John Wiley & Sons, Inc., 814 p.*, 1.

Bitmead, R. R. (1984). Persistence of excitation conditions and the convergence of adaptive schemes. *Information Theory, IEEE Transactions on*, 30(2):183–191.

Bittanti, S. and Colaneri, P. (2009). *Periodic Systems: Filtering and Control*. Communications and Control Engineering. Springer-Verlag, London.

Bittanti, S. and Moiraghi, L. (1994). Active control of vibrations in helicopters via pole assignment techniques. *Control Systems Technology, IEEE Transactions on*, 2(4):343–351.

Bobtsov, A. A., Kolyubin, S., Kremlev, A. S., and Pyrkin, A. (2012). An iterative algorithm of adaptive output control with complete compensation for unknown sinusoidal disturbance. *Automation and Remote Control*, 73(8):1327–1336.

Bode, H. W. et al. (1945). Network analysis and feedback amplifier design.

Bodson, M. and Douglas, S. C. (1997). Adaptive algorithms for the rejection of sinusoidal disturbances with unknown frequency. *Automatica*, 33(12):2213–2221.

Camacho, E. F. and Alba, C. B. (2013). *Model predictive control*. Springer Science & Business Media.

Challener, W., Peng, C., Itagi, A., Karns, D., Peng, W., Peng, Y., Yang, X., Zhu, X., Gokemeijer, N., Hsia, Y.-T., et al. (2009). Heat-assisted magnetic recording by a near-field transducer with efficient optical energy transfer. *Nature photonics*, 3(4):220–224.

Chandrasekar, J., Liu, L., Patt, D., Friedmann, P. P., and Bernstein, D. S. (2006). Adaptive harmonic steady-state control for disturbance rejection. *Control Systems Technology, IEEE Transactions on*, 14(6):993–1007.

Chaplin, B. (1980). The cancellation of repetitive noise and vibration. In *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, volume 1980, pages 699–702. Institute of Noise Control Engineering.

Chen, X. and Tomizuka, M. (2012). A minimum parameter adaptive approach for rejecting multiple narrow-band disturbances with application to hard disk drives. *Control Systems Technology, IEEE Transactions on*, 20(2):408–415.

Chen, X. and Tomizuka, M. (2013). Selective model inversion and adaptive disturbance observer for time-varying vibration rejection on an active-suspension benchmark. *European Journal of Control*, 19(4):300–312.

Chen, Y. Q., Moore, K. L., Yu, J., and Zhang, T. (2006). Iterative learning control and repetitive control in hard disk drive industry-a tutorial. In *Decision and Control, 2006 45th IEEE Conference on*, pages 2338–2351. IEEE.

Chew, K.-K. and Tomizuka, M. (1989). Digital control of repetitive errors in disk drive systems. In *American Control Conference, 1989*, pages 540–548. IEEE.

Chou, S. Y., Wei, M. S., Krauss, P. R., and Fischer, P. B. (1994). Single-domain magnetic pillar array of 35 nm diameter and 65 gbits/in. 2 density for ultrahigh density quantum magnetic storage. *Journal of Applied Physics*, 76(10):6673–6675.

Connolly, A. J., Green, M., Chicharo, J. F., and Bitmead, R. R. (1995). The design of lqg and h_∞ controllers for use in active vibration control and narrow band disturbance rejection. In *Decision and Control, 1995., Proceedings of the 34th IEEE Conference on*, volume 3, pages 2982–2987. IEEE.

Constantinides, A. G. (1970). Spectral transformations for digital filters. In *Proceedings of the Institution of Electrical Engineers*, volume 117, pages 1585–1590. IET.

De Roover, D. and Bosgra, O. H. (2000). Synthesis of robust multivariable iterative learning controllers with application to a wafer stage motion system. *International Journal of Control*, 73(10):968–979.

De Wit, C. C. and Praly, L. (2000). Adaptive eccentricity compensation. *Control Systems Technology, IEEE Transactions on*, 8(5):757–766.

Dijkstra, B. G. (2004). *Iterative learning control, with applications to a wafer-stage*. TU Delft, Delft University of Technology.

Diniz, P. S. (1997). *Adaptive filtering*. Springer.

Doh, T.-Y., Ryoo, J., and Chung, M. (2006). Design of a repetitive controller: an application to the track-following servo system of optical disk drives. *IEE Proceedings-Control Theory and Applications*, 153(3):323–330.

Eisenmenger, J. and Schuller, I. K. (2003). Magnetic nanostructures: overcoming thermal fluctuations. *Nature materials*, 2(7):437–438.

Elliott, H. (1981). Direct adaptive pole placement with application to nonminimum phase systems. In *Decision and Control including the Symposium on Adaptive Processes, 1981 20th IEEE Conference on*, pages 531–536. IEEE.

Elliott, S. and Nelson, P. (1985). The application of adaptive filtering to the active control of sound and vibration. *NASA STI/Recon Technical Report N*, 86:32628.

Elliott, S. J. and Darlington, P. (1985). Adaptive cancellation of periodic, synchronously sampled interference. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 33(3):715–717.

Elliott, S. J. and Nelson, P. A. (1993). Active noise control. *Signal Processing Magazine, IEEE*, 10(4):12–35.

Emborg, U. (1998). Cabin noise control in the saab 2000 high-speed turboprop aircraft. In *PROCEEDINGS OF THE INTERNATIONAL SEMINAR ON MODAL ANALYSIS*, volume 1, pages 13–26. KATHOLIEKE UNIVERSITEIT LEUVEN.

Eriksson, K., Estep, D., and Johnson, C. (2013). *Applied Mathematics: Body and Soul: Calculus in Several Dimensions*. Springer Science & Business Media.

Eriksson, L. and Allie, M. (1989). Use of random noise for on-line transducer modeling in an adaptive active attenuation system. *The Journal of the Acoustical Society of America*, 85(2):797–802.

Feng, G. and Palaniswami, M. (1992). A stable adaptive implementation of the internal model principle. *Automatic Control, IEEE Transactions on*, 37(8):1220–1225.

Feuer, A. and Weinstein, E. (1985). Convergence analysis of lms filters with uncorrelated gaussian data. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 33(1):222–230.

Feyh, G., Franchitti, J.-C., and Mullis, C. T. (1986). All-pass filter interpolation and frequency transformation problem. In *Proc./20th Asilomar Conf. on Signals, Systems and Computers, Pacific Grove*, pages 164–168.

Fontana, R. E., Hetzler, S. R., and Decad, G. (2012). Technology roadmap comparisons for tape, hdd, and nand flash: Implications for data storage applications. *Magnetics, IEEE Transactions on*, 48(5):1692–1696.

Francis, B. A. and Wonham, W. M. (1976). The internal model principle of control theory. *Automatica*, 12(5):457–465.

Friedmann, P. P. and Millott, T. A. (1995). Vibration reduction in rotorcraft using active control-a comparison of various approaches. *Journal of Guidance, Control, and Dynamics*, 18(4):664–673.

Giri, F., Dion, J., M'Saad, M., and Dugard, L. (1987). A globally convergent pole placement indirect adaptive controller. In *Decision and Control, 1987. 26th IEEE Conference on*, volume 26, pages 1–6. IEEE.

Goodzeit, N. E. and Phan, M. Q. (1997). System and periodic disturbance identification for feedforward-feedback control of flexible spacecraft. In *Proceedings of the 35th Aerospace Sciences Meeting and Exhibit*.

Guan, Y. H., Shepard Jr, W. S., Lim, T. C., and Li, M. (2004). Experimental analysis of an active vibration control system for gearboxes. *Smart materials and structures*, 13(5):1230.

Gupta, N. K. (1980). Frequency-shaped cost functionals-extension of linear-quadratic-gaussian design methods. *Journal of Guidance, Control, and dynamics*, 3(6):529–535.

Harris, C. M. (1991). *Handbook of acoustical measurements and noise control*. McGraw-Hill New York.

Herzog, R. (1994). Active versus passive vibration absorbers. *Journal of dynamic systems, measurement, and control*, 116(3):367–371.

Hirano, T., Yang, H., Pattanaik, S., White, M., and Arya, S. (2003). A microactuator for tracking servo. In *STLE/ASME 2003 International Joint Tribology Conference*, pages 21–25. American Society of Mechanical Engineers.

Hong, J. and Bernstein, D. S. (1998). Bode integral constraints, collocation, and spillover in active noise and vibration control. *Control Systems Technology, IEEE Transactions on*, 6(1):111–120.

Horowitz, R., Li, Y., Oldham, K., Kon, S., and Huang, X. (2007). Dual-stage servo systems and vibration compensation in computer hard disk drives. *Control Engineering Practice*, 15(3):291–305.

Houtzager, I., van Wingerden, J.-W., and Verhaegen, M. (2013). Rejection of periodic wind disturbances on a smart rotor test section using lifted repetitive control. *Control Systems Technology, IEEE Transactions on*, 21(2):347–359.

Kamen, E. W. and Heck, B. S. (2000). *Fundamentals of signals and systems: using the Web and MATLAB*. Prentice Hall.

Keikha, E., Shahsavari, B., Al-Mamun, A., and Horowitz, R. (2013). A probabilistic approach to robust controller design for a servo system with irregular sampling. In *Control and Automation (ICCA), 2013 10th IEEE International Conference on*, pages 1790–1795.

IEEE.

Kempf, C., Messner, W., Tomizuka, M., and Horowitz, R. (1993). Comparison of four discrete-time repetitive control algorithms. *IEEE Control Systems Magazine*, 13(6):48–54.

Khizroev, S. and Litvinov, D. (2006). *Perpendicular magnetic recording*. Springer Science & Business Media.

Knospe, C., Hope, R., Tamer, S., and Fedigan, S. (1996). Robustness of adaptive unbalance control of rotors with magnetic bearings. *Journal of Vibration and Control*, 2(1):33–52.

Knospe, C. R., Fedigan, S. J., Hope, R. W., and Williams, R. D. (1997). A multitasking dsp implementation of adaptive magnetic bearing control. *Control Systems Technology, IEEE Transactions on*, 5(2):230–238.

Konstanzer, P., Enenkl, B., Aubourg, P., and Cranga, P. (2008). Recent advances in eurocopter's passive and active vibration control. In *ANNUAL FORUM PROCEEDINGS-AMERICAN HELICOPTER SOCIETY*, volume 64, page 854. AMERICAN HELICOPTER SOCIETY, INC.

Krukowski, A., Cain, G. D., and Kale, I. (1995). Custom designed high-order frequency transformations for iir filters. In *Circuits and Systems, 1995., Proceedings., Proceedings of the 38th Midwest Symposium on*, volume 1, pages 588–591. IEEE.

Kuo, S. M. and Morgan, D. (1995). *Active noise control systems: algorithms and DSP implementations*. John Wiley & Sons, Inc.

Landau, I. D., Alma, M., Constantinescu, A., Martinez, J. J., and Noë, M. (2011a). Adaptive regulationâĂŤrejection of unknown multiple narrow band disturbances (a review on algorithms and applications). *Control Engineering Practice*, 19(10):1168–1181.

Landau, I. D., Constantinescu, A., and Rey, D. (2005). Adaptive narrow band disturbance rejection applied to an active suspensionâĂŤan internal model principle approach. *Automatica*, 41(4):563–574.

Landau, I. D., Landau, Y. D., and Zito, G. (2006). *Digital control systems: design, identification and implementation*. Springer Science & Business Media.

Landau, I. D., Lozano, R., M'Saad, M., and Karimi, A. (2011b). *Adaptive control: algorithms, analysis and applications*. Springer Science & Business Media.

Landau, I. D., Lozano, R., MâĂŹSaad, M., and Karimi, A. (2011c). Indirect adaptive control. In *Adaptive Control*, pages 409–456. Springer.

Landau, I. D., Silva, A. C., Airimitoaie, T.-B., Buche, G., and Noe, M. (2013). Benchmark on adaptive regulationâĂŤrejection of unknown/time-varying multiple narrow band disturbances. *European Journal of control*, 19(4):237–252.

Lau, J., Joshi, S. S., Agrawal, B. N., and Kim, J.-W. (2006). Investigation of periodic-disturbance identification and rejection in spacecraft. *Journal of guidance, control, and dynamics*, 29(4):792–798.

Leal, R. L. and Landau, I. (1982). Quasi-direct adaptive control for nonminimum phase systems. *Journal of Dynamic Systems, Measurement, and Control*, 104(4):311–316.

Li, M., Lim, T. C., Shepard Jr, W. S., and Guan, Y. (2005). Experimental active vibration control of gear mesh harmonics in a power recirculation gearbox system using a piezoelectric stack actuator. *Smart materials and structures*, 14(5):917.

Ljung, L. (1977a). Analysis of recursive stochastic algorithms. *Automatic Control, IEEE Transactions on*, 22(4):551–575.

Ljung, L. (1977b). On positive real transfer functions and the convergence of some recursive schemes. *Automatic Control, IEEE Transactions on*, 22(4):539–551.

Ljung, L. and Söderström, T. (1983). Theory and practice of recursive identification.

Lovera, M., Colaneri, P., Malpica, C., and Celi, R. (2003). Closed-loop aeromechanical stability analysis of hhc and ibc, with application to a hingeless rotor helicopter. In *29 th European Rotorcraft Forum*.

Lueng, P. (1936). Process of silencing sound oscillations. US Patent 2,043,416.

MacMartin, D. G. (1994). A feedback perspective on the lms disturbance feedforward algorithm. In *American Control Conference, 1994*, volume 2, pages 1632–1636. IEEE.

Malang, K. and Hutsell, L. (2005). Method and apparatus for calibrating piezoelectric driver in dual actuator disk drive. US Patent 6,975,123.

Malpica, C. A. (2008). *Contributions to the dynamics of helicopters with active rotor controls*. ProQuest.

Marino, R., Santosuosso, G. L., and Tomei, P. (2003). Robust adaptive compensation of biased sinusoidal disturbances with unknown frequency. *Automatica*, 39(10):1755–1761.

McEver, M. A., Cole, D. G., and Clark, R. L. (2004). Adaptive feedback control of optical jitter using q-parameterization. *Optical Engineering*, 43(4):904–910.

Mehra, R. K. (1970). On the identification of variances and adaptive kalman filtering. *Automatic Control, IEEE Transactions on*, 15(2):175–184.

Miu, D. K. and Bhat, S. P. (1991). Minimum power and minimum jerk position control and its applications in computer disk drives. *Magnetics, IEEE Transactions on*, 27(6):4471–4475.

Moon, J.-H., Lee, M.-N., and Chung, M. J. (1998). Repetitive control for the track-following servo system of an optical disk drive. *Control Systems Technology, IEEE Transactions on*, 6(5):663–670.

Moore, J. B. and Mingori, D. L. (1987). Robust frequency-shaped lq control. *Automatica*, 23(5):641–646.

M'Saad, M., Dugard, L., and Hammad, S. (1993). A suitable generalized predictive adaptive controller case study: control of a flexible arm. *Automatica*, 29(3):589–608.

M'saad, M., Ortega, R., and Landau, I. (1985). Adaptive controllers for discrete-time systems with arbitrary zeros: an overview. *Automatica*, 21(4):413–423.

M'saad, M. and Sanchez, G. (1992). Partial state reference model adaptive control of multivariable systems. *Automatica*, 28(6):1189–1197.

Myers, K., Tapley, B. D., et al. (1976). Adaptive sequential estimation with unknown noise statistics. *Automatic Control, IEEE Transactions on*, 21(4):520–523.

Nair, P. B. and Keane, A. J. (2001). Passive vibration suppression of flexible space structures via optimal geometric redesign. *AIAA journal*, 39(7):1338–1346.

New, R., Pease, R., and White, R. (1994). Submicron patterning of thin cobalt films for magnetic storage. *Journal of Vacuum Science & Technology B*, 12(6):3196–3201.

Nowrouzian, B. and Constantinides, A. (1990). Prototype reference transfer function pa-

rameters in the discrete-time frequency transformations. In *Circuits and Systems, 1990., Proceedings of the 33rd Midwest Symposium on*, pages 1078–1082. IEEE.

Nummiaro, K., Koller-Meier, E., and Van Gool, L. (2003). An adaptive color-based particle filter. *Image and vision computing*, 21(1):99–110.

Odelson, B. J., Rajamani, M. R., and Rawlings, J. B. (2006). A new autocovariance least-squares method for estimating noise covariances. *Automatica*, 42(2):303–308.

Oppenheim, A. V., Schafer, R. W., Buck, J. R., et al. (1989). *Discrete-time signal processing*, volume 2. Prentice-hall Englewood Cliffs.

Panda, S. P. and Lu, Y. (2003). Tutorial on control systems design in tape drives. In *American Control Conference, 2003. Proceedings of the 2003*, volume 1, pages 1–17. IEEE.

Pantazi, A., Jelitto, J., Bui, N., and Eleftheriou, E. (2012). Track-following in tape storage: Lateral tape motion and control. *Mechatronics*, 22(3):361–367.

Patt, D., Liu, L., Chandrasekar, J., Bernstein, D. S., and Friedmann, P. P. (2005). Higher-harmonic-control algorithm for helicopter vibration reduction revisited. *Journal of guidance, control, and dynamics*, 28(5):918–930.

Pearson, J. T. and Goodall, R. M. (1994). Adaptive schemes for the active control of helicopter structural response. *Control Systems Technology, IEEE Transactions on*, 2(2):61–72.

Pigg, S. and Bodson, M. (2006). Adaptive rejection of sinusoidal disturbances of known frequency acting on unknown systems. In *American Control Conference, 2006*, pages 5–pp. IEEE.

Pigg, S. and Bodson, M. (2010). Adaptive algorithms for the rejection of sinusoidal disturbances acting on unknown plants. *Control Systems Technology, IEEE Transactions on*, 18(4):822–836.

Regalia, P. (1994). *Adaptive IIR filtering in signal processing and control*, volume 90. CRC Press.

Rober, S. and Shin, Y. (1996). Control of cutting force for end milling processes using an extended model reference adaptive control scheme. *Journal of Manufacturing Science and Engineering*, 118(3):339–347.

Rokach, L. and Maimon, O. (2005). Clustering methods. In *Data mining and knowledge discovery handbook*, pages 321–352. Springer.

Sacks, A. H., Bodson, M., and Messner, W. (1995). Advanced methods for repeatable runout compensation [disc drives]. *Magnetics, IEEE Transactions on*, 31(2):1031–1036.

Shahsavari, B., Bagherieh, O., Mehr, N., Tomlin, C., and Horowitz, R. (2016). Optimal mode–switching and control synthesis for floating offshore wind turbines.

Shahsavari, B., Conway, R., Keikha, E., and Horowitz, R. (2012). Robust control design for hard disk drives with irregular sampling. In *APMRC, 2012 Digest*, pages 1–2. IEEE.

Shahsavari, B., Conway, R., Keikha, E., and Horowitz, R. (2013a). Limits of performance in systems with periodic irregular sampling and actuation rates. In *Proc. of 6th IFAC Symposium on Mechatronic Systems, Hangzhou, China*.

Shahsavari, B., Conway, R., Keikha, E., Zhang, F., and Horowitz, R. (2013b). h_∞ control design for systems with periodic irregular sampling using optimal h_2 reference controller-

s. In *ASME 2013 Conference on Information Storage and Processing Systems*, pages V001T03A008–V001T03A008. American Society of Mechanical Engineers.

Shahsavari, B., Conway, R., Keikha, E., Zhang, F., and Horowitz, R. (2013c). Robust track-following controller design for hard disk drives with irregular sampling. *IEEE TRANSACTIONS ON MAGNETICS*, 49(6).

Shahsavari, B., Keikha, E., Zhang, F., and Horowitz, R. (2014a). Adaptive repetitive control using a modified filtered-x lms algorithm. In *ASME 2014 Dynamic Systems and Control Conference*, pages V001T13A006–V001T13A006. American Society of Mechanical Engineers.

Shahsavari, B., Keikha, E., Zhang, F., and Horowitz, R. (2014b). Repeatable runout following in bit patterned media recording. In *ASME 2014 Conference on Information Storage and Processing Systems*, pages V001T03A001–V001T03A001. American Society of Mechanical Engineers.

Shahsavari, B., Keikha, E., Zhang, F., and Horowitz, R. (2015a). Adaptive repetitive control design with online secondary path modeling and application to bit-patterned media recording. *Magnetics, IEEE Transactions on*, 51(4):1–8.

Shahsavari, B., Maasoumy, M., Sangiovanni-Vincentelli, A., and Horowitz, R. (2015b). Stochastic model predictive control design for load management system of aircraft electrical power distribution.

Stol, K. A. and Balas, M. J. (2003). Periodic disturbance accommodating control for blade load mitigation in wind turbines. *Journal of solar energy engineering*, 125(4):379–385.

Sun, L., Chen, X., and Tomizuka, M. (2013). Neural-network based automatic pid gain tuning in the presence of time-varying disturbances in hard disk drives. In *ASME 2013 Conference on Information Storage and Processing Systems*, pages V001T03A007–V001T03A007. American Society of Mechanical Engineers.

Sun, L., Chen, X., and Tomizuka, M. (2014). Adaptive suppression of high-frequency wide-spectrum vibrations with application to disk drive systems. In *ASME 2014 Dynamic Systems and Control Conference*, pages V003T52A003–V003T52A003. American Society of Mechanical Engineers.

Tao, G. (2003). *Adaptive control design and analysis*, volume 37. John Wiley & Sons.

Tomizuka, M. (1987). Zero phase error tracking algorithm for digital control. *Journal of Dynamic Systems, Measurement, and Control*, 109(1):65–68.

Tomizuka, M., Chew, K.-K., and Yang, W.-C. (1990). Disturbance rejection through an external model. *Journal of dynamic systems, measurement, and control*, 112(4):559–564.

Tsao, T. and Pong, K. (1991). Control of radial runout in multi-tooth face milling. *Transactions of the North American Manufacturing Research Institute of SME*, pages 183–190.

Tsao, T.-C. and Bentsman, J. (1996). Rejection of unknown periodic load disturbances in continuous steel casting process using learning repetitive control approach. *Control Systems Technology, IEEE Transactions on*, 4(3):259–265.

Tsypkin, Y. Z. (1997). Stochastic discrete systems with internal models. *Journal of Automation and Information Sciences*, 29(4-5).

Ungerboeck, G. (1972). Theory on the speed of convergence in adaptive equalizers for digital

communication. *IBM Journal of Research and Development*, 16(6):546–555.

Voulgaris, P. G., Dahleh, M. A., and Valavani, L. S. (1994). HâĹđ and h2 optimal controllers for periodic and multirate systems. *Automatica*, 30(2):251–263.

WIDROW, B., HOFF, M. E., et al. (1960). Adaptive switching circuits.

Wilby, J. (1996). Aircraft interior noise. *Journal of Sound and Vibration*, 190(3):545–564.

Wu, B. and Bodson, M. (2003). A magnitude/phase-locked loop approach to parameter estimation of periodic signals. *Automatic Control, IEEE Transactions on*, 48(4):612–618.

Wu, B. and Bodson, M. (2004). Multi-channel active noise control for periodic sources - indirect approach. *Automatica*, 40(2):203–212.

Wu, S.-C. and Tomizuka, M. (2006). Repeatable runout compensation for hard disk drives using adaptive feedforward cancellation. In *American Control Conference, 2006*, pages 382–387.

Yang, J. K., Chen, Y., Huang, T., Duan, H., Thiyagarajah, N., Hui, H. K., Leong, S. H., and Ng, V. (2011). Fabrication and characterization of bit-patterned media beyond 1.5 tbit/in2. *Nanotechnology*, 22(38):385301.

Zhang, F., Keikha, E., Shahsavari, B., and Horowitz, R. (2014a). Adaptive mismatch compensation for rate integrating vibratory gyroscopes with improved convergence rate. In *ASME 2014 Dynamic Systems and Control Conference*, pages V003T45A003–V003T45A003. American Society of Mechanical Engineers.

Zhang, F., Keikha, E., Shahsavari, B., and Horowitz, R. (2014b). Adaptive mismatch compensation for vibratory gyroscopes. In *Inertial Sensors and Systems (ISISS), 2014 International Symposium on*, pages 1–4. IEEE.

Zhang, M., Lan, H., and Ser, W. (2000). An improved secondary path modeling method for active noise control systems. *IEEE Signal Processing Letters*, 7(4):73–75.

Zhang, M., Lan, H., and Ser, W. (2001). Cross-updated active noise control system with online secondary path modeling. *Speech and Audio Processing, IEEE Transactions on*, 9(5):598–602.

Zheng, M., Chen, X., and Tomizuka, M. (2014a). Discrete-time frequency-shaped sliding mode control for audio-vibration rejection in hard disk drivesâŃĘ. *Mechanical Engineering*.

Zheng, M., Chen, X., and Tomizuka, M. (2014b). A nonlinear feedback control scheme for transient performance enhancement in hard disk drives. In *ASME 2014 Conference on Information Storage and Processing Systems*, pages V001T03A003–V001T03A003. American Society of Mechanical Engineers.

Zheng, M., Chen, X., Wang, H., Kim, Y.-H., Xi, W., and Tu, K.-Y. (2015). Data storage device comprising slew rate anti-windup compensation for microactuator. US Patent 9,007,714.