Title
Adaptive Local Loop Shaping and Inverse-based Youla-Kucera Parameterization with Application to Precision Control

Author
CHEN, XU

Publication Date
2013

Adaptive Local Loop Shaping and Inverse-based Youla-Kucera Parameterization with Application to Precision Control

by

Xu Chen

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 Masayoshi Tomizuka, Chair
Professor Roberto Horowitz
Professor Andrew Packard
Associate Professor Murat Arcak

Fall 2013

Adaptive Local Loop Shaping and Inverse-based Youla-Kucera Parameterization with Application to Precision Control

Abstract

Adaptive Local Loop Shaping and Inverse-based Youla-Kucera Parameterization with Application to Precision Control

by

Xu Chen

Doctor of Philosophy in Engineering - Mechanical Engineering

University of California, Berkeley

Professor Masayoshi Tomizuka, Chair

In this dissertation we discuss loop-shaping algorithms that bring enhanced servo performance at multiple local frequency regions. These local loop shaping (LLS) algorithms are motivated by several new demands in practical control systems such as hard disk drives in information storage industry, wafer scanners in semiconductor manufacturing, active steering in automotive vehicles, and active suspension in structural vibration rejection. We will examine how knowledge about the disturbance/reference characteristics can be utilized, both offline and online, to customize the servo system for meeting the control challenges.

Along the way, we investigate several design concepts and methodologies. First, in Youla-Kucera (YK) parameterization–the parameterization of all stabilizing linear controllers–we develop plant factorizations based on selective model inversion, which safely inverts a (possibly nonminimum-phase) plant dynamics, using H_∞ minimization and pole/zero modulation. This allows us to obtain a simplified YK algorithm with strong design and tuning intuitions in practical servo. Also, with selective model inversion, it becomes quite approachable to control the waterbed effect, the result of Bode's Integral Formula in fundamental limitations of linear control design. This is achieved by utilizing add-on pole/zero placement and convex-optimization approaches to minimize the disturbance amplification in the sensitivity function, which enables the obtaining of several algorithms for enhanced repetitive control and vibration rejection.

In the third part of the dissertation, we investigate adaptive formulations to achieve online identification of the disturbance characteristics. We study the application of infinite-impulse-response (IIR) filters in YK parameterization, which brings benefits such as minimum-parameter adaptation and better convergence under noisy adaptation environments. We also provide an optimization-based approach to address the problem of robust strict positive real transfer functions, an essential requirement in adaptive control and system identification.

The discussed algorithms are then extended from the control of SISO to MISO (multi-input-single-output) plants, where we formulate a decoupled disturbance observer for estimating the equivalent input disturbance for different actuators in a MISO system.

Simulation and experimental results are obtained on the four classes of systems discussed at the beginning of this abstract. Parts of the results are performed on benchmark problems, studied and compared with the algorithms of peer researchers under extensive test conditions.

To my family and supporting friends

Contents

List of Figures

List of Tables

Nomenclature

DDOB	decoupled disturbance observer
DOB	disturbance observer
DS	decoupled sensitivity
EPS	electrical power steering
ERC	enhanced repetitive control
FIR	finite-impulse-response
HDD	hard disk drive
IIR	infinite-impulse-response
IMP	internal model principle
LLS	local loop shaping
LMI	linear matrix inequality
MIMO	multi-input-multi-output
MISO	multi-input-single-output
PAA	parameter adaptation algorithm
PES	position error signal
RC	repetitive control
SDP	semidefinite programming
SISO	single-input-single-output
SPR	strictly positive real
VGRS	variable-gear-ratio steering
YK parameterization	Youla-Kucera parameterization

Acknowledgments

This dissertation is undoubtedly shaped by many people in the past five years of my graduate school.

First and foremost, I would like to thank my advisor and mentor, Professor Masayoshi Tomizuka. I was just a college kid when I met Tomizuka Sensei in 2008, with a young heart but almost no knowledge of control theory. Professor Tomizuka accepted and guided me, from learning in class, to working on research projects, and to writing proposals and presenting my ideas. His profound knowledge and insightful vision always keep me on track about research. His great care about students, and philosophy of career and life, mark the deepest imprint of memories in my school life.

I am also appreciative of my other dissertation committee members for their time and suggestions on this dissertation. Professor Roberto Horowitz provided great help in my research on hard disk drives (HDD), and shared with me many suggestions about managing post-graduate life; Professor Andrew Packard taught me multivariable control theory, and more importantly, about the attitudes towards teaching and presenting; Professor Murat Arcak gave me great lectures on digital signal processing, something I have been ever using in my graduate school.

Professor Ioan Landau at GIPSA lab (France) made direct and indirect impacts on part of the results in the dissertation. It was very fruitful, and a great pleasure to cooperate with his student Abraham Castellanos Silva on the benchmark on adaptive regulation.

Part of my research funds comes from the Computer Mechanics Lab (CML) directed by Professor David Bogy. I am grateful for his organization of the annual meetings with industrial sponsors, and the chances to present my research and proposals. I am also very grateful for sponsors of my other financial supports: Department Scholarship for Graduate Study in 2009, Frank and Margaret Lucas Fellowship in 2011, and Chinese Government Award for Outstanding Self-financed Students Abroad in 2012.

A lot of thanks go to Dr. Guoxiao Guo and Dr. Li Yi at Western Digital (WD) Corporation. I have always regarded Guoxiao as my "outside advisor". He not only provided generous financial support in the UCB-WD projects, but also taught me the importance of research vision. Li's support to the entire HDD research group was indispensable for maintaining the quality of the CML projects. He also set up a great example of creating changes to a working environment and positively influencing people's career paths. Besides these two directors, I received great help from many other people at WD, to name a just few: Jie Yu, Edward Tu, Jake Kim, Min Chen, Wei Xi, Qixing Zheng, Shang-Chen Wu, Jianguo Zhou, and Alan Young. It has been really fruitful working with them during my internships at WD. It was also great encouragement knowing that my designed algorithms are implemented in industrial mass production.

National Instruments, Nikon Research Corporation of America, and Agilent Technologies provided generous hardware support throughout the research. I appreciate their consistent support in academic projects.

The current and past members in the Mechanical Systems Control Laboratory are incredible researchers and great friends. There are so many great memories and I will definitely be missing the school days with these people: Qixing Zheng, Nancy Dong, Hoday Stearns, Wenjie Chen, Evan Chang-Siu, Kyoungchul Kong, Haifei Cheng, Mike Chan, Kan Kanjanapas, Chi-Shen Tsai, Raechel Tan, Yizhou Wang, Wenlong Zhang, Cong Wang, Pedro Reynoso, Benjamin Fine, Minghui Zheng, Changliu Liu, Xiaowen Yu, Chen-Yu Chan, Yaoqiong Du, Kevin Haninger, Chung-Yen Lin, Junkai Lu, Hiroshi Niki, Joonbum Bae, Kiyonori Inaba, Xuan Fan, Shuwen Yu, Sanggyum Kim, Nora Han, and Takashi Nagata. I also cannot forget to mention my other Berkeley friends: Yudong, Young-Hwan, Theresa, Selina, Frank, Shiyuan, Fu, Yiqi, Heng, Jianbin, Zhichao, Na, Nan, and Liang. I wish them best of luck in pursuing their dreams.

I enjoyed the collaborations with several visiting scholars and students: Atsushi Oshima, Liting Sun, Atsuki Naka, Julian Salt-Llobregat, Sakon Yu, and Sumio Sugita. I wish all of them success in their careers.

I had great fun, and learned a lot, when serving as the graduate student instructor for two advanced control classes: ME 232 in 2012 fall and ME 233 in 2013 spring (thank you again, Tomizuka Sensei). I led discussions for many of them for two full semesters. It is indeed cheerful to see their great passion and eager about learning. The smiles on their faces are the best reward for those long (and really long) hours of grading.

Last but not least, my wife Zhanzhan Jia, my parents, and my parents-in-law have always been supporting me with unlimited love and dedication. They know my thoughts before I speak the words. My heartfelt thanks goes to them for their love, companionship, and sacrifices.

Part I

Background and Introduction

Chapter 1

Introduction

1.1 Practical Servo Control Problems

Technology innovations are urging the penetration of advanced control design in modern engineering systems. In a commercial hard disk drive (HDD), the servo controller must position the read/write head at nano-scale accuracy[1] for data access on the magnetic disks. In 2007, this servo challenge can be mimicked by imagining an airplane flying at 5,000,000 mph above a 100,000-lane highway, to follow the center of a lane whose width is only fraction of an inch [1]! Five years after that, the maximum area density increased by 300%, meaning that there are 300,000 more lanes on the highway, and the lane width is about one forth of its value in 2007 [2]. Along with the density expansion, new compact HDDs are also introduced in the post PC era, for new products such as ultrabooks and slim portable drives, where challenges in servo control are significantly amplified by various new disturbance sources (more details in Section 3.2). In the semiconductor industry, the error upper bound in the motion control of the wafer scanner is even smaller—about one magnitude lower than that of the HDDs in 2007.

Such ultra-high precision and robust performance in modern control systems are achieved by careful consideration of various engineering disciplines. From the viewpoint of system integration, we can classify the design elements to the following four categories:

C1. *hardware and sensing components*: we may want to choose fluid or air bearings for reduced friction, laser interferometers or high-precision encoders for accurate measurements, and/or piezo-electronic/MEMS actuators for fine positioning;

C2. *operation environment*: such as the considerations for friction-isolation tables, clean room, and thermostatic chambers;

[1]In a commercial 2013 HDD, the width of a data track is at the scale of 70 nano meters. In positioning of the read/write heads, servo control has to maintain the position error to be less than 10% track width for acceptable data writing.

C3. *task plans and arrangements*: such as well-designed trajectories, and the arrangement of task repetitions in a manufacturing process;

C4. *servo control algorithms*, i.e., the software development to operate the actuators, and to correct the accumulated errors in the aforementioned three design processes.

A well-designed engineering system reflects optimized considerations in all the above categories. Servo control, as the finalizing step, is responsible for synthesizing and compensating as much as possible the accumulated residual imperfections. Such imperfections are often inevitable. To list a few in each design categories, we have:

C1. *hardware imperfection* which comes from system resonances, imperfections in bearings and gears, torque ripples in motors, periodic disturbance from cooling fans, delays in motor drivers and signal acquisition, and so on;

C2. *environmental disturbance* such as windage (e.g. turbulent airflow in HDDs), and complex structural vibrations that heavily depend on the operation environment (e.g., vibrations induced from high-power audio speakers in laptop computers);

C3. *special errors due to the task nature*, such as repeated trajectories in industrial robots and manufacturing processes.

Figure 1.1 shows several examples of mechanical systems and their typical error spectra collected from experimental tests and benchmark simulations [3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17] (detailed hardware description and the root cause of the errors will be provided in Chapter 3). One immediate common feature we can observe, is that all errors in Figure 1.1 show structured peaks in the spectra.

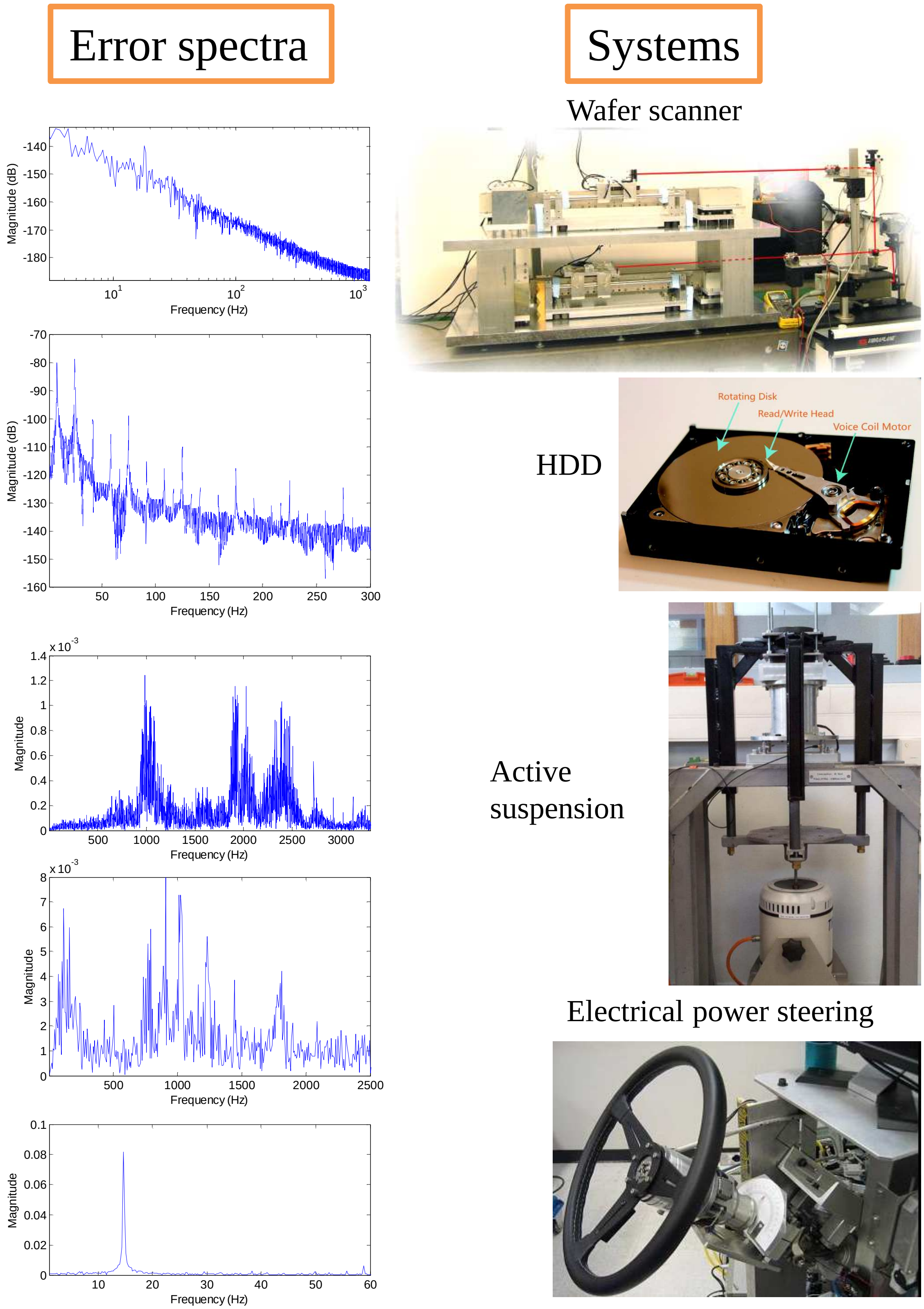


Figure 1.1: Example error spectra in different mechanical systems

Standard feedback controllers, such as lead-lag compensators, PID controllers, and notch filters have already been applied to systems in Figure 1.1, to achieve error-rejection functions similar to that in Figure 1.2. The exact bandwidth varies in each construction. Yet Figure 1.2 reflects a typical shape in practical servo control, where we have good servo performance at low frequencies, and reduced error rejection as the frequency increases.

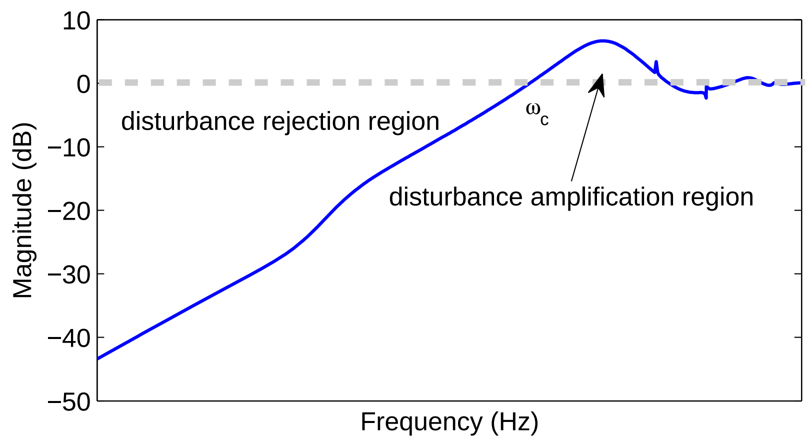


Figure 1.2: A standard magnitude response of the sensitivity function (x axis is in log scale)

From the example spectra in Figure 1.1, we see standard loop shaping is commonly not sufficient in high-precision servo, due to the inevitability of various imperfections in practice. The following error characteristics and servo requirements further amplify the design challenge:

- *errors (particularly vibrations) may occur at frequencies above the servo bandwidth*, and are hence unattenuated or amplified by feedback;[2]

- *disturbances can be uncertain and/or time-varying.* For instance, disturbances from imperfect motor rotation and turbulent air flow are hardware- or trajectory-dependent, and can vary among different products; HDD vibrations from audio speakers in a laptop will depend on the vibration source and the mechanical path between the source and the HDD;

- and lastly, *a baseline system performance*, such as that in Figure *1.2,* has to be maintained to meet general servo performance and robustness; in the meantime, *strong flexibility* has to be built into the servo design, for handling disturbance variations and providing easy tuning options.

[2]Even below the bandwidth, the feedback attenuation may be insufficient in the presence of very strong vibrations.

1.2 Local Loop Shaping

In this dissertation we discuss customized feedback design at local frequency regions. This local loop shaping (LLS) concept, explained in the magnitude response of the output error-rejection/sensitivity function, consists of the following aspects:

Strong local servo enhancement while maintaining a baseline servo performance

For instance in the dashed lines in Figure 1.3, we investigate ways to bring different notch shapes in a local frequency region. Meanwhile, we preserve the achieved performance of the baseline system (solid line), and study ways to minimize the error amplifications (reflected as the magnitude increase of the sensitivity function).

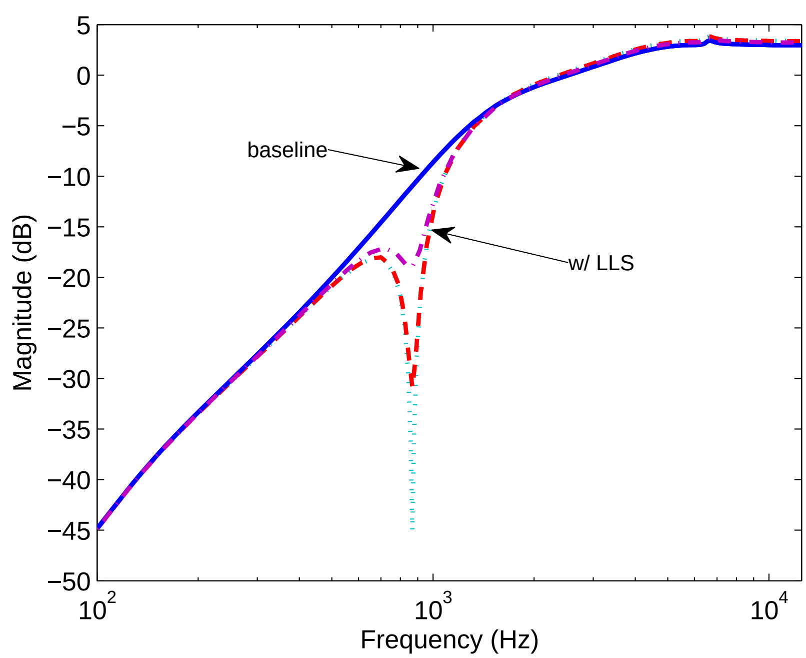


Figure 1.3: The concept of local loop shaping

High flexibility in the achievable range of attenuation

This is particularly required for attenuation of disturbances above the servo bandwidth. Take Figure 1.4—the response of several sensitivity functions in HDD control—as an example. The zero-dB crossover frequency for the baseline sensitivity function (solid line) is around 2000 Hz (a typical value for industrial dual-stage HDDs), while modern products are subject to vibrations at frequencies as high as 3000 Hz [8, 10]. In Figure 1.4, we show six examples of local loop shaping at different frequencies, ranging from 500 Hz to 4000 Hz. Without modifying the entire servo structure, we can use the designs of Figure 1.4 for LLS beyond the servo bandwidth.

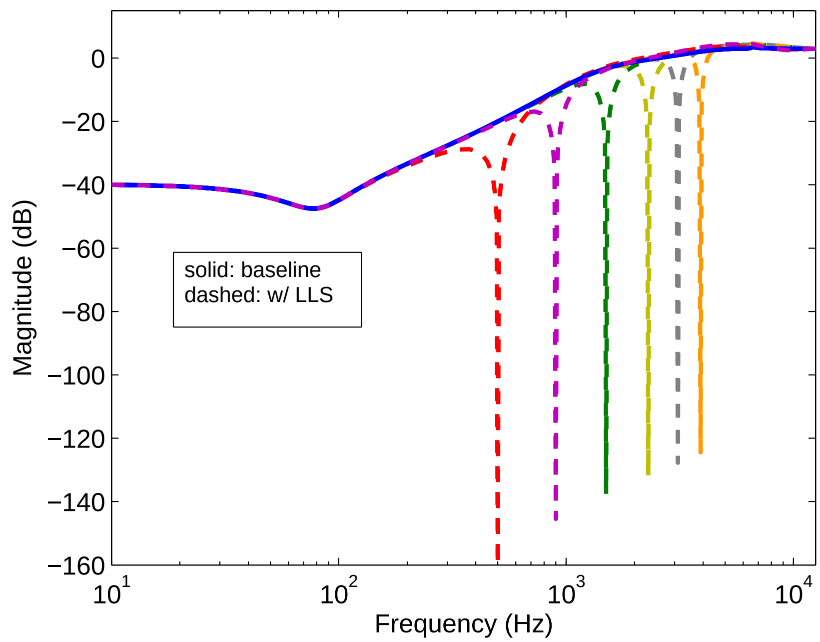


Figure 1.4: LLS configurations at different frequency regions

Multi-band local loop shaping

With the results in Figures 1.3 and 1.4, we study the extension to LLS at multiple frequency regions, as demonstrated in Figures 1.5 and 1.6. In Figure 1.5, very strong attenuation is placed at several frequencies that are very close to each other. In Figure 1.6, we provide repetitive control to attenuate harmonic disturbances that are composed of frequency components at integer multiplications of a fundamental frequency.

Adaptive control

As discussed at the end of Section 1.1, exact characteristics of many practical disturbances may be unknown or time-varying. For LLS in the framework of Figures 1.3 and 1.5, we construct online parameter adaptation algorithms to automatically allocate the desired attenuation regions.

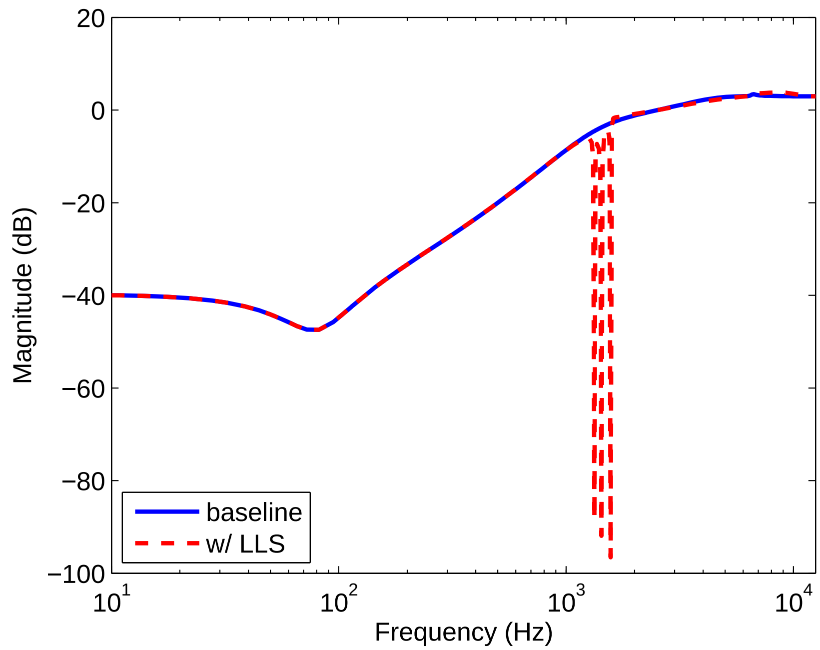


Figure 1.5: Multi-band LLS for rejecting three narrow-band disturbances

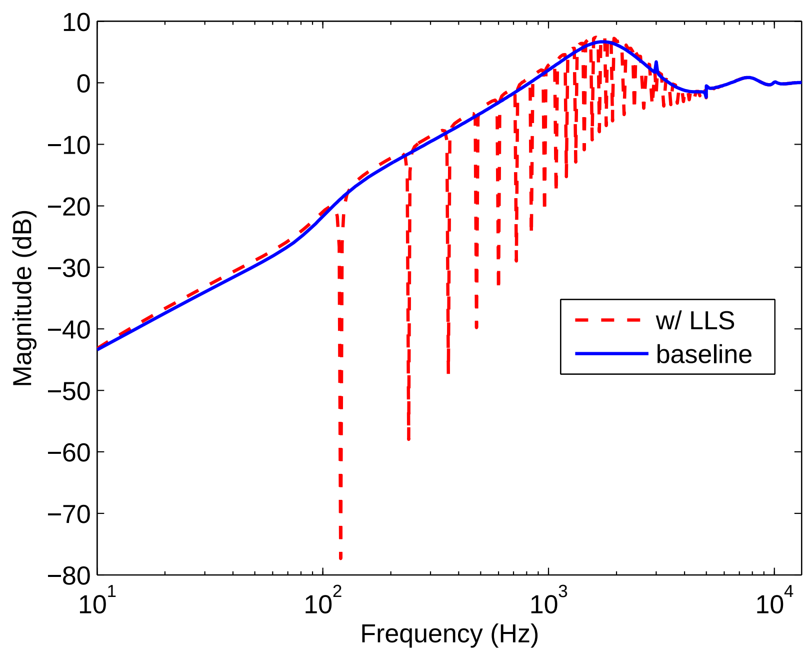


Figure 1.6: LLS for harmonic-error rejection

Remark 1.1. Notice that general bandwidth extension, such as the one shown in Figure 1.7, can be regarded as a special case of local loop shaping. Such a result is relatively easier to achieve. We will discuss the design of it along the analysis of general LLS.

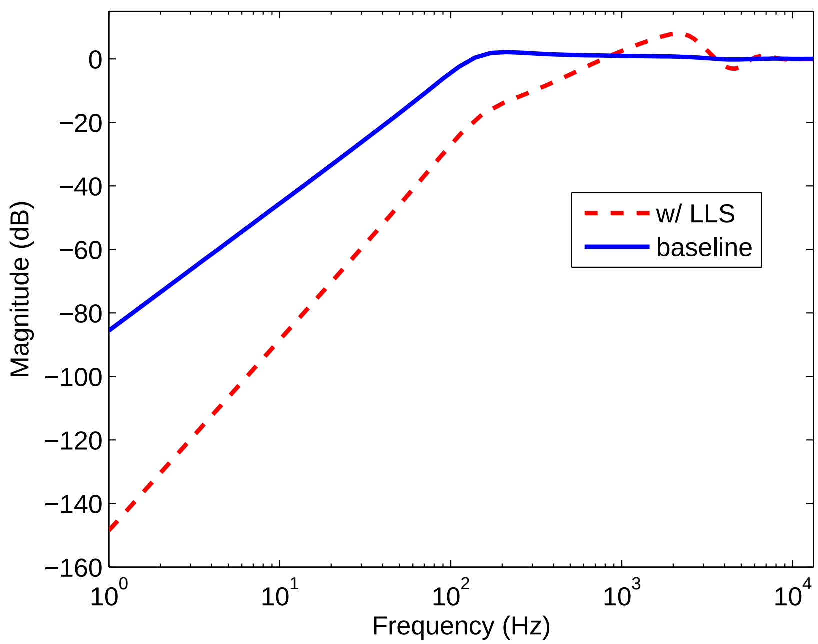


Figure 1.7: LLS for bandwidth extension

Many of the demonstrated LLS concepts are challenging to achieve in both theory and practice. Even for the simpler case of narrow-band disturbance rejection, designers must be careful when using standard loop shaping techniques such as the popular peak-filter method [18]. The main theoretical challenge comes from one fundamental limitation of feedback design, which proves that enhanced local servo performance, under mild assumptions, will always be accompanied by error amplifications at other frequency regions (more details in the next section). From the practical aspect of servo design, the difficulty of loop shaping greatly increases at frequencies above the servo bandwidth, where the loop stability and the robustness against plant uncertainties are much more sensitive compared to that in low-frequency loop shaping.

1.3 Organization and Contributions of this Dissertation

Practical Youla Parameterization

It is easy to see the influence of a feedback controller C on the open-loop system response $L = PC$ (P is the transfer function of an LTI plant); less straightforward, however, to see the impact of C on closed-loop quantities such as the sensitivity function $S = 1/(1+PC)$ and the complementary sensitivity function $T = PC/(1+PC)$. One main part of this dissertation is devoted to making loop shaping such as those in Figures 1.3 to 1.7 intuitively achievable, and in the meantime with minimum design effort.

A fundamental concept in achieving the goal is Youla-Kucera (YK) parameterization—the parameterization of all stabilizing controllers. It provides that any stabilizing controllers can be parameterized using a fixed structure, if a perfect model of the plant is available. The design of the Q filter in this scheme however does not have a commonly agreed rule. General discrete-time YK parametrization usually applies an unstructured finite-impulse-response (FIR) filter [19, 20, 21, 22]. In the continuous-time case, discussions on using a linear combination of some basis transfer functions [23, 24] have been explored.

One main reason for the variance in designing the Q filter is the heavy dependence on the coprime factorization of the plant model, particularly in multi-input-multi-output (MIMO) systems. For this unsolved problem, we discuss structural choices of the plant factorization based on selective model inversion, and form a pseudo and robust YK parameterization, with more intuitive and uniform Q design [Chapter 4]. This pseudo YK parameterization has been the core structure for successful applications to various mechanical systems including hard disk drives [3, 5, 4], wafer scanners [4, 13], active suspensions [7, 16, 25], and electrical power steering [8].

Inverse-based Feedback

Inverse-based control has long been used in servo design (e.g., in feedforward control), but not extensively explored in direct YK parameterization. This is particularly true for MIMO control systems, where plant inversion itself is a nontrivial concept [Chapter 9]. We discuss two ideas of safe inversion of a (possibly nonminimum-phase) plant for general servo design [Chapter 7]: one using unstable-zero modulation, and another based on H_∞ optimization.

Generalized Disturbance Observer

A third contribution of the dissertation is to provide a generalized concept of the disturbance observer (DOB) [26]—a well-known robust control design tool [27, 28, 29, 30, 31]. In the discrete-time control framework, we show that DOB is a special case of Youla-Kucera parameterization in regulation control, and discuss the class of stabilizing controllers that is missing in the DOB scheme [Chapter 4]. We also provide generalized disturbance observers for periodic [Chapter 5] and harmonic signals [Chapter 8]. Finally, DOB theory has been mainly developed in the SISO control framework [10]. We provide one extension of DOB to MISO plants, and develop ideas to decouple the disturbance-estimation problem to individual actuators in a MISO system [Chapter 9].

Adaptive Vibration Rejection

One main application of LLS is for high-performance vibration rejection. Practical systems that involve rotational motions are inevitably subjected to periodic references or disturbances. The problem of adaptive narrow-band disturbance[3] rejection has thus attracted great research attention in control engineering, and recently been extensively studied in the multi-band situation [25]. Adaptive noise cancellation [32] uses additional sensors and stochastic-gradient-based adaptation to cancel the disturbance effect. Adaptive feedforward cancellation [33] handles sinusoidal disturbances by composing an estimate of the disturbance with trigonometric functions. From the perspective of feedback loop shaping, controllers can be customized to incorporate the disturbance structure to asymptotically reject the vibrations. This internal-model-principle [34] based perspective has been investigated in feedback control algorithms in [35, 20, 36, 37, 38, 39, 40], among which [37, 36] used state-space designs; and [35, 20, 38, 39, 40] applied YK Parameterization with adaptive FIR Q filters. Alternatively, the disturbance frequency can be firstly estimated and then applied for controller design. This indirect-adaptive-control perspective has been considered in [41, 42, 43].

Indeed, frequency identification of narrow-band signals is a problem that receives great research attention itself. Among the related literatures we can find: (i) methods using nonparametric spectrum estimation or eigen analysis (subspace methods) [44, 45, 46]; (ii) online adaptive identification approaches [47, 48, 49, 50, 51, 52, 53, 54]. Among references in group (ii), for the identification of n frequency components, adaptive notch filters in the orders of $5n-1$ [53], $2n+6$ [54], $3n$ [48, 49, 50, 51], and $2n$ [47], have been discussed.

Different from the FIR formulations, we base the construction of discrete-time YK parameterization on infinite-impulse-response (IIR) filters. By using additionally the internal model principle, we are able to obtain a YK parameterization that requires the minimum number of adaptation parameters [Chapters 5 and 10]. An additional consideration is that adaptation on IIR structures enables direct adaptive control with parallel predictors that use the *a posteriori* adaptation information, which is essential for accurate parameter convergence under noisy environments [55]. The importance of this aspect can also be seen from the aforementioned literature on frequency identification.

Along the way, we provide a new study of the problem about strictly positive real transfer functions [Chapter 12]—an essential problem in stability analysis, adaptive control, and system identification [55, 56].

Control of the Waterbed Effect

Generalizing from narrow- to wider-band error rejection, we provide several Q-filter designs for enhanced band-limited disturbance rejection [3, 6, 7, 8, 9, 10, 16] and enhanced repetitive control [4, 5]. Some of the results are compared with algorithms of peer researchers in benchmark problems [4, 25]. Of particular difference is the handling of the "waterbed" effect in feedback loop shaping [8, 25]. From Bode's Integral Formula, enhanced servo performance

[3]Disturbances whose energies are concentrated within narrow frequency bands.

at certain frequencies commonly results in deteriorated loop shapes at other frequencies (details in Section 2.2). In the pseudo YK parameterization, the proof of the waterbed effect takes a much easier form, and we provide systematic ways to control the waterbed via add-on pole/zero placement or automatic convex-optimization formulations [Chapter 6]. These considerations are useful for, e.g., the new and more challenging wide-band disturbance rejection problem [Section 3.2], and the repetitive control problem under strong non-periodic noises [4].

MISO Control Design

Multi-input-single-output (MISO) plants are important (and easy) extensions of SISO systems, with yet great potential of enhanced plant dynamics for servo design (see one example in Section 3.2). However, specific MISO loop shaping design is much less discussed in literature. As long as the dimension of the plant output is unity, the feedback loop will have a scalar sensitivity function, and we show that the rich SISO loop shaping algorithms based on YK parameterization can be readily extended to MISO systems [Chapter 9].

From the viewpoint of problem solving and algorithm implementation, the rest of the dissertation is organized as follows:

In Chapter 2 of Part I, we review several foundational results of linear control, digital signal processing, and convex optimization. Chapter 3 explains details of the mentioned examples in Figure 1.1, and analyzes the root cause of various errors in practical control systems.

In Part II, we discuss SISO and MISO local loop shaping algorithms when the disturbance or reference structure is *a priori* known. Chapter 5 presents the basic design of the Q filter, which is the heart of LLS. In Chapter 6, advanced Q-filter design concepts are introduced for enhanced control of the waterbed effect. In Chapter 7 we discuss practical aspects of inverse design. Chapter 8 generalizes the loop shaping idea and presents an extension of Q-filter design for enhanced repetitive control.

Part III of the dissertation provides adaptive LLS and practical implementation of the algorithms. We design online identification and adaptive control algorithms in Chapter 10, to obtain the disturbance/reference characteristics. Chapter 11 provides suggestions on algorithm implementation and several more example applications. Chapter 12 is about the achieving of the strictly positive real (SPR) condition. In Chapter 13, we summarize the dissertation and discuss some future works.

1.4 Format and Notations

$G(s)$ and $G(z)$ denote, respectively, continuous- and discrete-time transfer functions. If the index is omitted, then the result is valid if G is either $G(z)$ or $G(s)$. Consider, for instance, the statement "the sensitivity function is defined by $S \triangleq 1/(1+PC)$".

We use $G(e^{j\omega})\left(\triangleq G(z)|_{z=e^{j\omega}}\right)$ and $G(j\omega) = \left(G(s)|_{s=j\omega}\right)$ for the frequency responses of $G(z)$ and $G(s)$. $G(\omega)$ denotes the general frequency response that holds for both $G(z)$ and $G(s)$.

Following conventions in adaptive control and sampled-data control systems, we denote q^{-1} as the one-step delay operator (also known as the backward shift operator) that satisfies

$$u(k) = q^{-1}u(k+1)$$

For instance, if

$$y(k) = -0.9y(k-1) + u(k-1)$$

then we can write

$$y(k) = -0.9q^{-1}y(k) + q^{-1}u(k)$$
$$\Rightarrow y(k) = \frac{q^{-1}}{1+0.9q^{-1}}u(k)$$

where $q^{-1}/(1+0.9q^{-1})$ is the pulse transfer function from $u(k)$ to $y(k)$.

More generally, we can use $y(k) = G(q^{-1})\,u(k)$ to describe the input-output relation in Figure 1.8, under the assumption of zero initial conditions. In the Z domain, a corresponding statement is $Y(z) = G(z)\,U(z)$, where $Y(z)$ and $U(z)$ are the Z transforms of $y(k)$ and $u(k)$.

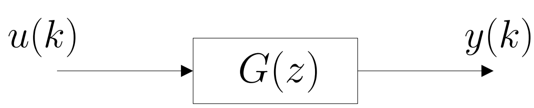


Figure 1.8: Input-output modeling of an LTI system

We sometimes use the full expression $Y(z) = G_{yu}(z)U(z)$, where the sub indices in $G_{yu}(z)$ explicitly explains that G_{yu} is the transfer function from u to y.

Unless explicitly stated, we assume all signals are causal, namely, the value of the signal is zero before time zero.

SISO is the abbreviation for single-input single-output. MISO and DISO denote, respectively, "multi-input single-output" and "dual-input single-output".

An FIR filter, such as $1 + z^{-1} + 2z^{-2}$, has a finite impulse response; IIR filters have nonzero poles and hence infinite-length impulse responses.

Chapter 2

Elemental Tools and Concepts

In this chapter, we review several fundamental results in linear control theory.

2.1 Performance Goals in Feedback Design

Consider the block diagram of a standard feedback system in Figure 2.1. The closed-loop transfer function from the disturbance d_o to the plant output y, is the sensitivity function

$$S \triangleq (I + PC)^{-1}$$

which is the same as the closed-loop transfer function from the reference r to the feedback error e.

The complementary sensitivity function is defined as

$$T \triangleq I - S = PC(I + PC)^{-1}$$

i.e., T is the transfer function from the reference r to y. Additionally, $-T$ is the transfer function from the measurement noise n to y.

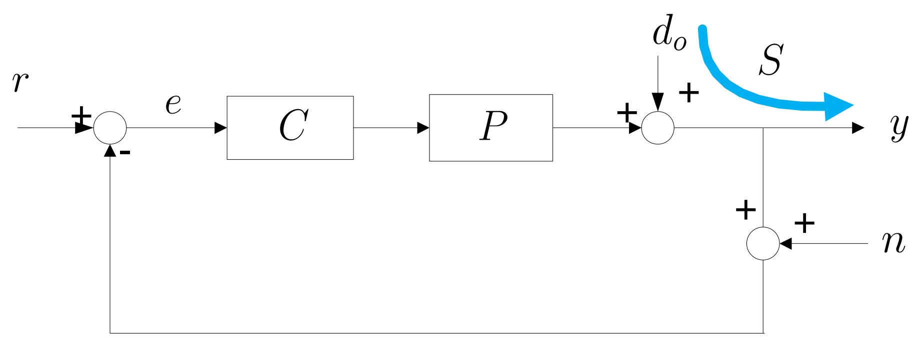


Figure 2.1: A standard closed-loop system under feedback control

The sensitivity function measures the closed-loop disturbance-attenuation property, while the complementary sensitivity function defines how the system responds to the reference input as well as the sensor noise. For SISO systems, a typical magnitude response of S is shown in Figure 2.2. Below the cross-over frequency ω_c, the magnitude response of S is less than 1, hence the attenuation of d_o in Figure 2.1.

Ultimately, to make $S = (I + PC)^{-1}$ small at certain frequencies, we need the open-loop transfer function $L \triangleq PC$ to be large in the same region. This is the concept of high-gain feedback control. The term "loop shaping" conventionally refers to designing the magnitude response of L to stabilize the loop and satisfy the desired performance metrics.

Wherever the magnitude of $S(\omega)$ is small, $T(\omega)$ will be close to unity due to the fundamental relationship $S+T = I$. This means that at frequencies where we achieve good disturbance rejection, we also obtain improved reference tracking and amplified sensor noise. Advances in hardware design are providing sensors that are more and more accurate. Modern laser interferometers can reach a measurement resolution at the picometer level. Micrometer-resolution linear encoders are also commercially available in the market. Hence, to some extent, the problem of disturbance rejection is a more important issue than sensor noises attenuation in modern precision systems.

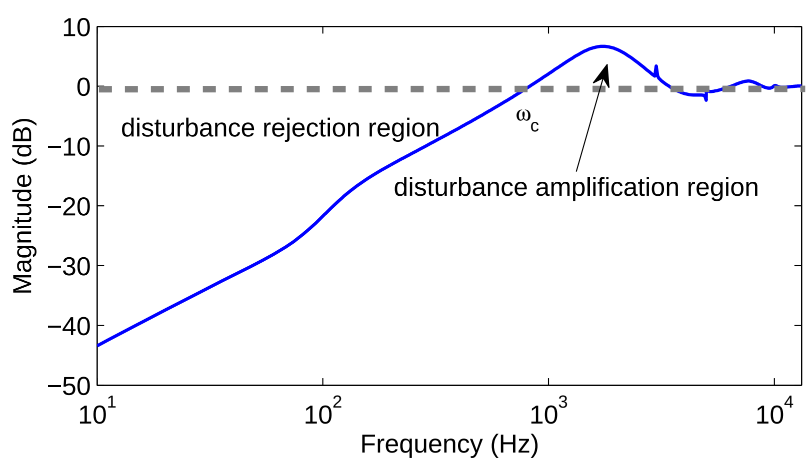


Figure 2.2: A typical magnitude response of the sensitivity function (x axis is in log scale)

2.2 Fundamental Limitations for LTI Systems

Mathematically, $G(s)$ and $G(z)$ are rational complex functions. Complex analysis provides several guidance and fundamental limitations about the achievable performance in LTI systems. We review a few that are most relevant to this dissertation.

Bandwidth Limitation: The bandwidth ω_c in Figure 2.2 can not be pushed to be arbitrarily large. Intuitively, a mechanical system would not be able to respond to arbitrarily fast control inputs, due to hardware (such as motors, gears, etc) limitations. The plant behavior will become uncertain when the frequency of the input is pushed high enough. For these reasons, it is common practice to keep the gain of the controller small at high frequencies.

From the theoretical perspective, under mild conditions, it is inevitable to have $|S(\omega')| > 1$ at certain frequencies if $|S(\omega)| < 1$ holds over some frequency interval, namely, when certain disturbance components are attenuated in the feedback system, some other disturbance components will be amplified (see the areas above and below the 0dB line in Figure 2.2). This is the *waterbed effect* which comes from Bode's Integral Theorem:

Theorem 2.1. *Let the open-loop transfer function $L(s)$ be a stable, proper, scalar rational transfer function in a single-input single-output system. Let $S(s) = 1/(1+L(s))$ and assume that $S(s)$ has no poles in the right half plane, and $k_s = \lim_{s\to\infty} sL(s)$. Then*

$$\frac{1}{\pi}\int_0^\infty \ln|S(j\omega)|d\omega = \frac{-1}{2}k_s \tag{2.1}$$

If in addition the relative degree of $L(s)$ is no less than 2, then $k_s = 0$, yielding

$$\frac{1}{\pi}\int_0^\infty \ln|S(j\omega)|d\omega = 0 \tag{2.2}$$

The above theorem indicates that the sum of the area (relative to the 0dB line) under the log magnitude response of the sensitivity function, is zero under the stated conditions for (2.2). The x axis in Figure 2.2 has a log scale. Observing Figure 2.3, the linear-scale equivalent of Figure 2.2, we can clearly see the conservation of area from Theorem 2.1.

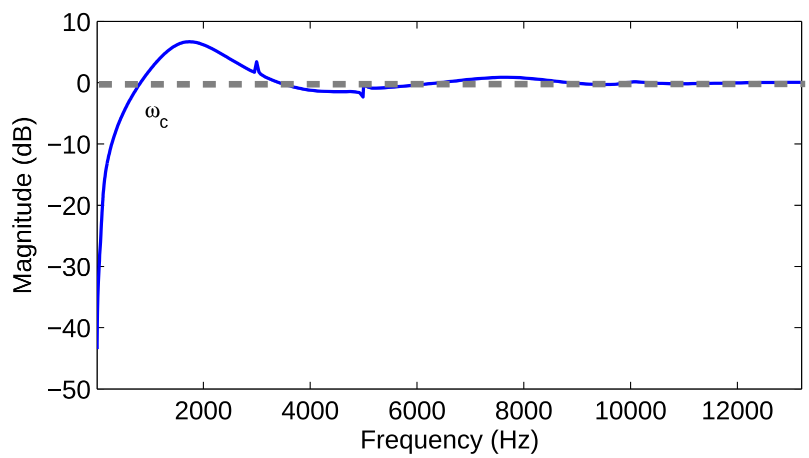


Figure 2.3: A typical magnitude response of the sensitivity function (x axis is in linear scale)

The condition of the relative degree being larger than or equal to two is usually not difficult to satisfy in practice. Practical systems have small gains at high frequencies. The rate of high-frequency rolloff is usually higher than (or at least as fast as) $1/s$. It is also natural, at least for motion-control systems, to have integrator-type plant dynamics, as motors commonly take force/torque as the input, and generate (angular) position or velocity as the output.[1] An exceptional case is when the system behaves simply like a constant gain. For instance, piezoelectric (PZT) actuators are constructed to have an overall flat magnitude response with resonances at very high frequencies. The relative degree for such systems can be less than two, hence greatly relaxing the waterbed effect.[2] This is one of the reasons that make PZT actuators popular for high-bandwidth designs in high-precision servo.

The waterbed effect gets even worse if the open-loop system has unstable poles.

Theorem 2.2. *(Bode's integral formula for continuous-time SISO systems) Let $L(s)$ be a proper, scalar rational transfer function, of relative degree larger than 1. Let $S(s) = 1/(1 + L(s))$ and assume that $S(s)$ has no poles in the right half plane, and has $q \geq 0$ zeros in the closed right half plane, at locations p_1, p_2, ..., p_q. Then*

$$\int_0^\infty \ln |S(j\omega)| \, d\omega = \pi \sum_{k=1}^{q} p_k \tag{2.3}$$

Proof. There are several ways to prove Bode's integral formula. We provide one version in Appendix A.1. □

From the above theorem, when open-loop unstable poles (which becomes zeros of $S(s)$ in the closed right half plane) exist, disturbance amplification will always happen regardless of the presence of any disturbance attenuation.

For continuous-time systems, the waterbed effect holds if the relative degree of the loop transfer function $L = PC$ is no less than two. In the discrete-time case, the waterbed effect is more inevitable:

Theorem 2.3. *(Bode's integral formula for discrete-time systems) For all closed-loop stable discrete-time feedback systems, the sensitivity function has to satisfy the following integral constraint:*

$$\int_0^\pi \ln \left|S\left(e^{j\omega}\right)\right| d\omega = \pi \sum_{k=1}^{q} \ln |p_k| \tag{2.4}$$

where p_k are the open-loop unstable poles of the system, and q is the total number of such poles.

[1] By Newton's laws we then have $f = ma = m\ddot{y}$ or $\tau = J\ddot{\theta}$, yielding a nominal transfer function of $1/\left(ms^2\right)$ or $1/(Js^2)$.

[2] This great hardware advantage is, however, also accompanied by the drawback that PZT actuators have very limited actuation range. For such reasons, PZT actuators are usually combined with long-range actuators to form a dual-stage system for practical high-precision servo.

Proof. See [57]. □

One intuition for the drop of the relative degree requirement here is that the zero order hold has a low-pass type dynamics and introduces high-frequency rolloffs in the magnitude response.

Limitations From Nonminimum-phase Zeros: Besides the waterbed effect from Bode's Integral Formula, nonminimum-phase zeros also place fundamental limitations in feedback design. Take a simple example where $P(z)$ has $1 - z^{-1}$ in the numerator, then any constant input will have null effect in the output of $P(z)$! Other limitations from general nonminimum-phase zeros include:

- the sensitivity function will always have magnitudes larger than one. In

$$S(\sigma) = 1/(1 + P(\sigma) C(\sigma))$$

  (σ denotes s or z), if σ_o is a nonminimum-phase zero of $P(\sigma)$, then $P(\sigma_o) = 0$ and $S(\sigma_o) = 1/(1 + 0 \times C(\sigma_o)) = 1$, regardless of the design of $C(\sigma)$. If S is stable—hence analytic in the right-half plane (or outside the unit circle)—then the maximum modulus theorem[3] indicates that the maximum of $|S(\sigma)|$ is achieved on the imaginary axis (or the unit circle). Hence $\max|S(\sigma)| \geq |S(\sigma_o)| = 1$. There will thus always be a frequency region where we can not achieve high servo performance, particularly around the frequencies of the nonminimum-phase zeros.

- unbounded input can produce zero output at the steady state of a nonminimum-phase system. Consider, for instance, $u(t) = e^{2t}$ and $G(s) = (s-2)/(s+1)^2$.

- step responses can have initial undershoot. Moreover, zero crossovers occur for step responses. Consider [58]

$$Y(s) = \frac{G(s)}{s} = \int_0^\infty y(t) e^{-st} dt$$

  If $G(\sigma_o) = 0$ where $\sigma_o > 0$ is the unstable zero, then

$$Y(\sigma_o) = \int_0^\infty y(t) e^{-\sigma_o t} dt = 0$$

  As $e^{-\sigma_0 t}$ is positive and decreasing, $y(t)$ must change signs for $t \in [0, \infty)$.

[3] *Maximum modulus theorem*: if a complex function $S(\sigma)$ is defined and continuous on a closed bounded set $\mathbf{\Omega}$, and it is analytic on the interior of $\mathbf{\Omega}$, then $|S(\sigma)|$ can not attain the maximum in the interior of $\mathbf{\Omega}$ unless it is a constant.

Resonant and Anti-resonant Modes: Resonances and anti-resonances, if very sharp in the magnitude response, also create *practical* difficulty in control design, since very small mismatch in the resonant frequency (due to e.g. temperature change) will make the identified model behave differently from the actual system near the resonant frequencies.

2.3 Internal Models of Signals

Many disturbance and reference signals can be regarded as the output of a system excited by an impulse signal. For instance, a constant $d(k) = d$ is the impulse response of the scaled integrator $d/(1-z^{-1})$; $\sin(\omega_0 k)$ and $\cos(\omega_0 n)$ are respectively impulse responses of the filters $z^{-1}\sin(\omega_0)/(1-2z^{-1}\cos(\omega_0)+z^{-2})$ and $(1-z^{-1}\cos(\omega_0))/(1-2z^{-1}\cos(\omega_0)+z^{-2})$. If we write

$$d(k) = \frac{B_d(q^{-1})}{A_d(q^{-1})}\delta(k) \tag{2.5}$$

where $\delta(k)$ is a Dirac impulse; $B_d(q^{-1})$ and $A_d(q^{-1})$ are coprime polynomials of the one-step delay operator q^{-1}, then the internal model principle (IMP) [34, 59] states that the disturbance can be asymptotically rejected if the polynomial $A_d(q^{-1})$ is absorbed in the feedback controller. The proof of IMP is actually quite simple. One version is provided in Appendix A.2.

We will mostly be using partial results of IMP, that

$$A_d(q^{-1})d(k) = B_d(q^{-1})\delta(k) \to 0 \tag{2.6}$$

where the transient response is specified by the polynomial $B_d(q^{-1})$. We call (2.6) the internal model (IM) of the signal $d(k)$. For common disturbance and reference signals, we summarize $A_d(q^{-1})$ and the transient length in Table 2.1

Table 2.1: Internal models of common signals in control engineering

Signals	$A_d(q^{-1})$	Convergence
a^k, $\alpha \leq 1$	$1-aq^{-1}$	$k \geq 1$
$\cos(\omega_0 k)$ and $\sin(\omega_0 k)$	$1-2q^{-1}\cos(\omega_0)+q^{-2}$	$k \geq 2$
$a^k\cos(\omega_0 k)$ and $a^k\sin(\omega_0 k)$	$1-2aq^{-1}\cos(\omega_0)+a^2q^{-2}$	$k \geq 2$
periodic $d(k)$ with $d(k) = d(k-N)$	$1-q^{-N}$	$k \geq N$

There are two ways to understand Table 2.1. Taking the example of $\cos(\omega_0 k)$, in the time domain, we have

$$\begin{aligned}
&\left(1-2q^{-1}\cos(\omega_0)+q^{-2}\right)\cos(\omega_0 k)\\
=&\left[1-q^{-1}\left(e^{j\omega_0}+e^{-j\omega_0}\right)+q^{-2}\right]\frac{e^{j\omega_0 k}+e^{-j\omega_0 k}}{2}\\
=&\frac{1}{2}\left[e^{j\omega_0 k}-e^{j\omega_0(k-1)}\left(e^{j\omega_0}+e^{-j\omega_0}\right)+e^{j\omega_0(k-2)}\right]\\
&+\frac{1}{2}\left[e^{-j\omega_0 k}-e^{-j\omega_0(k-1)}\left(e^{j\omega_0}+e^{-j\omega_0}\right)+e^{-j\omega_0(k-2)}\right]\\
=&\frac{1}{2}\left[e^{j\omega_0 k}-\left(e^{j\omega_0 k}+e^{j\omega_0(k-2)}\right)+e^{j\omega_0(k-2)}\right]\\
&+\frac{1}{2}\left[e^{-j\omega_0 k}-\left(e^{-j\omega_0(k-2)}+e^{-j\omega_0 k}\right)+e^{-j\omega_0(k-2)}\right]\\
=&0\ \forall k\geq 2
\end{aligned}$$

In the Z domain, the filter

$$1-2z^{-1}\cos(\omega_0)+z^{-2}=\left(1-e^{j\omega_0}z^{-1}\right)\left(1-e^{-j\omega_0}z^{-1}\right)$$

has two zeros at $e^{\pm j\omega_0}$, hence zero magnitude response at ω_0. Any input at this frequency will thus yield null effect in the output.

2.4 Discrete-time Plant Delay

Delays are not uncommon in practical control systems. Even if the continuous-time plant has no input delays, the discrete-time plant model (with zero order holder) will have at least one-step delay. This is because, due to sampling and the zero order holder, the output $y(k+1)$ depends only on $u(i)$ up to $i=k$ and not on $u(k+1)$ (see Figure 2.4), namely, there is always a default one-step input delay.

When the plant has resonances, additional phase delays are introduced if we use notch filters, as shown in Figure 2.6, to reduce the loop gain at the resonant frequencies. This is from Bode's Phase Formula—another fundamental theorem in loop shaping—which proves that the phase of a stable and *minimum-phase* transfer function is determined uniquely by its magnitude response. Roughly speaking, a large slope in magnitude response corresponds to a large phase value. A positive/negative slope corresponds to a positive/negative phase angle.

Theorem 2.4. *(Bode's phase formula) If L is a minimum-phase continuous-time transfer function, then its phase is uniquely defined by its gain, according to*

$$\angle L(j\omega)=\int_{-\infty}^{\infty}\frac{d\ln\left|L\left(e^{\nu}\omega\right)\right|}{d\nu}\psi(\nu)\,d\nu$$

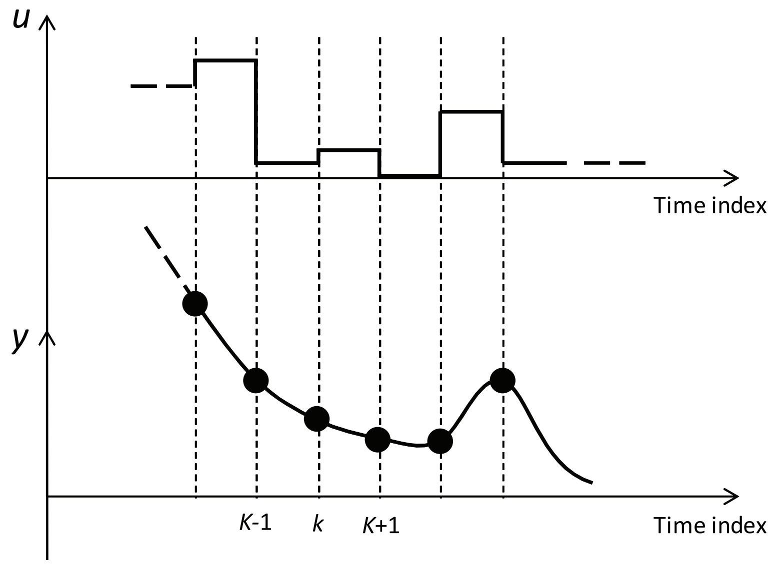


Figure 2.4: Discrete-time sampled-data input and output

where

$$\psi(\nu) = \frac{1}{\pi} \ln \frac{e^{|\nu|/2} + e^{-|\nu|/2}}{e^{|\nu|/2} - e^{-|\nu|/2}}.$$

The function $\psi(\nu)$ has the characteristics as shown in Fig. 2.5. The main contribution to the integral is made in the region $\nu \approx 0$. Hence the integral mainly depends on

$$\frac{d \ln |L(e^{\nu}\omega)|}{d\nu}$$

with $\nu \approx 0$, therefore the discussed results in the paragraph before Theorem 2.4.

Back to the notch-filter induced delays. Second-order notch filters have the structure of

$$\frac{s^2 + 2\beta\Omega_0 s + \Omega_0^2}{s^2 + 2\alpha\Omega_0 s + \Omega_0^2}$$

with $1 > \alpha > \beta > 0$ and α, β being close to zero; or

$$k \frac{1 - 2\alpha \cos(\Omega_0 T_s) z^{-1} + \alpha^2 z^{-2}}{1 - 2\beta \cos(\Omega_0 T_s) z^{-1} + \beta^2 z^{-2}}$$

($k = \frac{1-2\beta\cos(\Omega_0 T_s)+\beta^2}{1-2\alpha\cos(\Omega_0 T_s)+\alpha^2}$ for unity DC gain) with $1 > \alpha > \beta > 0$ and α, β being close to one. The stability and minimum-phase assumptions are both satisfied here and hence Bode's Phase Formula holds. The magnitude-phase relationship in Figure 2.6 clearly matches the implications of Bode's Phase Formula (see, in particular, the responses near the center frequency of the notch).

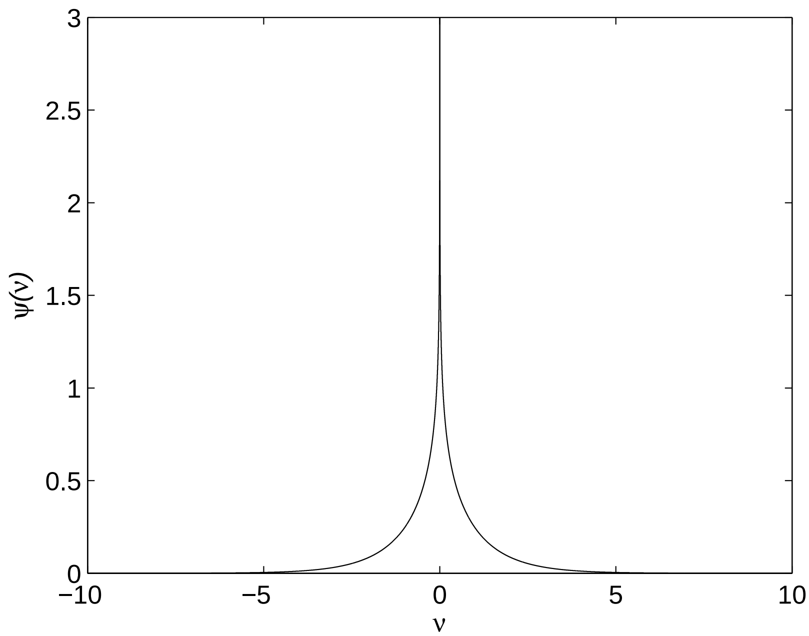


Figure 2.5: Characteristics of the function $\psi(\nu)$ in Bode's Phase Formula

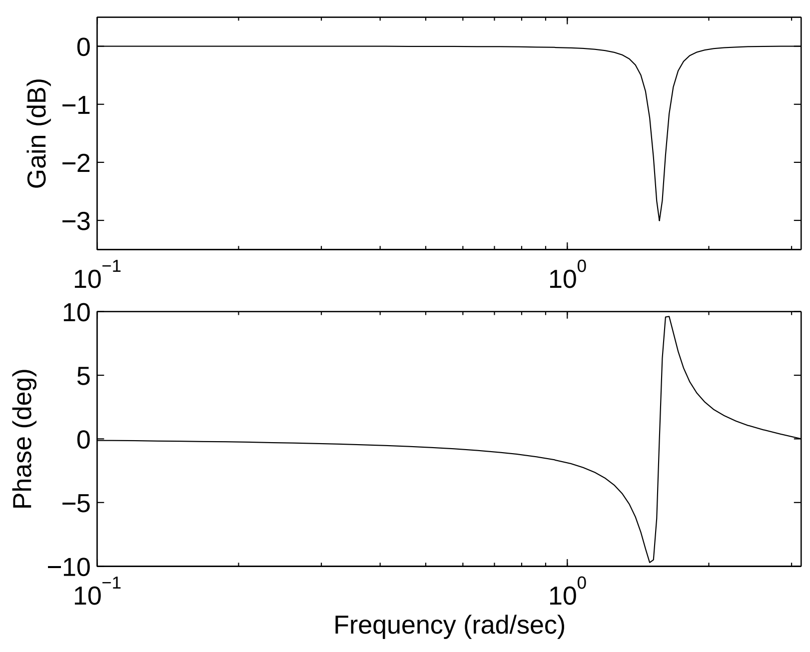


Figure 2.6: Example frequency response of a discrete-time notch filter

Influences of Delays: Digital delays increases the relative degree of the plant. For instance, $G(z) = 1/(1-z^{-1}) = z/(z-1)$ has a zero relative degree and $z^{-1}G(z) = 1/(z-1)$ has a relative degree of one. Delays are also "feedback invariant". Consider the case where the discrete-time plant has m steps of delays, and assume that the controller does not introduce additional delays. Then the open-loop transfer function can be factorized as $L(z) = z^{-m}L_m(z)$, yielding the complementary sensitivity function $T(z) = z^{-m}L_m(z) / [1 + z^{-m}L_m(z)]$.

2.5 Convex Optimization

Optimization theory has rich mathematical foundations and is a vast research field itself. We summarize the minimum required information for this dissertation, and refer readers to [60] for a full introduction of the theory and algorithms.

The standard optimization problem has the form of

$$\begin{array}{lll} \min\limits_{x} & f_0(x) & \ldots\ldots \text{objective function/cost} \\ \text{s.t.} & f_i(x) \leq 0,\ i = 1, \ldots, m & \ldots\ldots \text{constraints} \end{array}$$

where $x \in \mathbb{R}^n$ is called the decision variable; "s.t." denotes "subject to"; and

$$X = \{x \in \mathbb{R}^n :\ f_i(x) \leq 0,\ i = 1, \ldots, m\}$$

is called the feasibility set. There may be global and local optimal points depending on the nature of the problem. We are mostly interested in convex problems where the cost function is convex ("bowl-shaped") and the constraints form a convex feasibility set. This way the solution is guaranteed to exist and can be efficiently found using well-developed algorithms such as simplex and interior point methods.

Common functions such as linear and quadratic functions easily fall into the convex category. Some standard convex optimization problems are summarized in Table 2.2.

The semidefinite programing is in particular relevant to control engineering. For instance, consider the following matrix inequality (which is closely related to the strictly positive real condition of a continuous-time transfer function)

$$\begin{bmatrix} A_p^T M + M A_p & -C_p^T + M B_p \\ -C_p + B_p^T M & -D_p^T - D_p \end{bmatrix} \prec 0 \tag{2.7}$$

where the matrices $A_p \in \mathbb{R}^{n\times n}$, $B_p \in \mathbb{R}^{n\times 1}$, and $D_p \in \mathbb{R}$ are given; and $M \in \mathbb{R}^{n\times n}$, $M = M^T$, and $C_p = [c_0, c_1, \ldots, c_n]$ are the decision variables. Inequality (2.7) can be transformed to

Table 2.2: Standard convex optimization problems relevant to this dissertation

Name	Objective	Constraints
Linear programming (LP)	$\min_x \; c^T x$	$a^T x \leq b$
Quadratic programming (QP)	$\min_x \; x^T Q x + c^T x$, where $Q \succeq 0$	$a^T x \leq b$
Quadratically constrained QP (QCQP)	$\min_x \; x^T Q x + c^T x$, where $Q \succeq 0$	$x^T P_i x + c_i^T x \leq b$ $P_i \succeq 0$
Second order cone programming (SOCP)	$\min_x \; c^T x$	$\\|A_i x + b_i\\|_2 \leq c_i^T x_i + d_i, \; i = 1, ..., m$
Semidefinite programming (SDP)	$\min_x \; c^T x$	$F_0 + \sum_{i=1}^m x_i F_i \succeq 0$ where F_i's are given symmetric matrices

the linear matrix inequality (LMI):

$$
\begin{bmatrix} A_p^T M + M A_p & 0_{1\times n} \\ 0_{n\times 1} & -D_p^T - D_p \end{bmatrix} + \begin{bmatrix} 0_{n\times n} & M B_p \\ B_p^T M & 0 \end{bmatrix}
$$

$$
+ c_0 \begin{bmatrix} 0_{n\times n} & [-1, 0, \ldots, 0]^T \\ [-1, 0, \ldots, 0] & 0 \end{bmatrix} + c_1 \begin{bmatrix} 0_{n\times n} & [0, -1, \ldots, 0]^T \\ [0, -1, \ldots, 0] & 0 \end{bmatrix} +
$$

$$
\cdots + c_n \begin{bmatrix} 0_{n\times n} & [0, 0, \ldots, -1]^T \\ [0, 0, \ldots, -1] & 0 \end{bmatrix} \prec 0
$$

$$
M = m_{11} \begin{bmatrix} 1 & 0 & \ldots & \ldots & 0 \\ 0 & 0 & \ldots & \ldots & 0 \\ \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & \ldots & \ldots & \ldots & 0 \end{bmatrix} + m_{12} \begin{bmatrix} 0 & 1 & 0 & \ldots & 0 \\ 1 & 0 & \ldots & \ldots & 0 \\ 0 & \vdots & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & \ldots & \ldots & \ldots & 0 \end{bmatrix}
$$

$$
+ \cdots + m_{nn} \begin{bmatrix} 0 & 0 & \ldots & \ldots & 0 \\ 0 & 0 & \ldots & \ldots & 0 \\ \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & & 0 & 0 \\ 0 & \ldots & \ldots & 0 & 1 \end{bmatrix}
$$

and readily solved in semidefinite programming problems.

Notes and References

Many of the ideas about loop shaping and performance limits are from [23, 61]. [62, 63] are two richer references containing also MIMO control theory.

Details of the convex optimization problems and their solutions are provided in [60].

Chapter 3

Application Examples

Recall the three error sources we mentioned in Chapter 1: hardware imperfection, operation environment, and task arrangement. Part of the errors—such as those from imperfect motor rotation, repeated trajectories, and fan noises—are repeatable once the hardware and the trajectory are fixed. Other errors—such as those caused by (environmental) vibrations—although may vary case by case, are at least structured. Structured errors are common in practice. Otherwise, if errors are pure white noises, little can be done for servo improvement. In this chapter, we analyze the error sources in several practical examples, and provide the corresponding plant characteristics. Verification of the local loop shaping algorithms will be performed on these systems in later chapters.

3.1 Advanced Manufacturing

We discuss first the wafer-scanning process in lithography. This is one key step for circuit fabrication in the semiconductor industry. To print the circuit, wafers are exposed to patterned ultraviolet lights that come through a mask carried by a reticle stage. The wafer and the reticle stages move the wafer and the mask in a synchronized manner. Due to limited size of the lens, only a small part of the wafer is exposed at each scan, and the wafer is moved from one field to another in between the scans. The scanning process is repeated until all required areas on the wafer have been exposed under the light.

Figure 3.1 illustrates an experimental setup of the key wafer-scanner components at the Mechanical Systems Control Laboratory, UC Berkeley [64]. There are two stages in the system, mounted on air bearings and actuated by epoxy-core linear permanent magnet motors (LPMMs). The stage positions are measured by laser interferometers. A LabVIEW real-time system with field-programmable gate array (FPGA) is used to execute the control commands with a sampling time of 0.0004 sec. The frequency response of the reticle stage is shown in Figure 3.2. The input and the output are respectively the voltage command into the actuator and the position of the moving stage.

Figure 3.3 shows the movement of the reticle stage during a scaled scanning process.

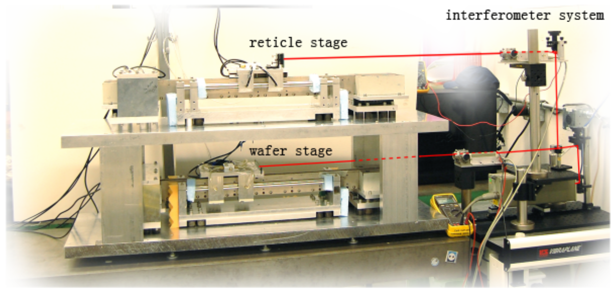


Figure 3.1: An experimental testbed for wafer scanners

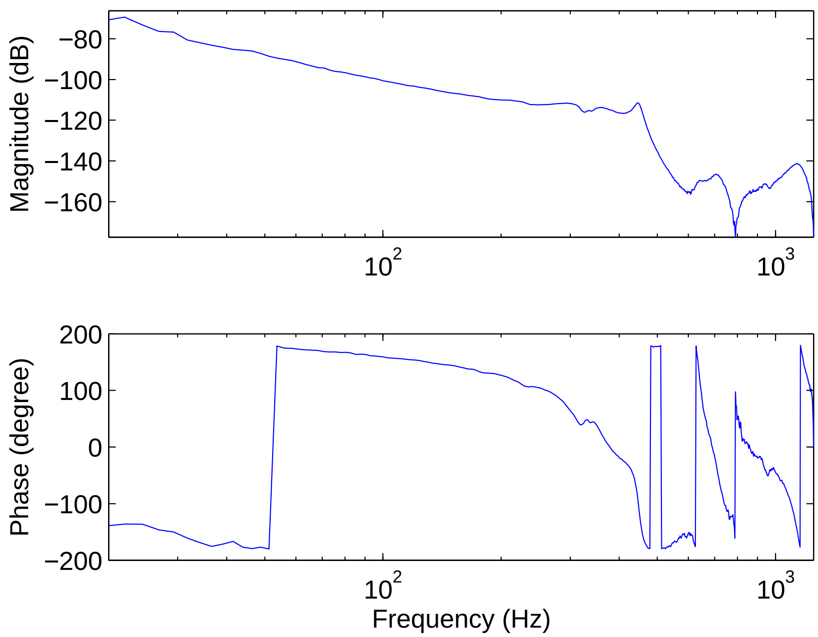


Figure 3.2: Frequency responses of the reticle stage

This is the baseline result with a simple PID controller, without any feedforward algorithms or customized controller parameterizations. From the repetitive nature of the process, if we append all errors at continuous iterations, the position error in Figure 3.3 will be replicated to yield a time sequence in Figure 3.4. The fast Fourier transform (FFT) of this error signal is shown in Figure 3.5, where we observe strong energy components at multiples of a fundamental frequency. Figures {3.3, 3.4, 3.5} analyze just one simple example encountered in applications. However, regardless of the shape of the trajectory, the periodic error structure would not change. The fundamental frequency can be analytically obtained by $f_o = 1/T_{\text{trajectory}}$, where $T_{\text{trajectory}}$ is the period of one iteration. This is a direct result of the Fourier-series theory. In the example in Figures {3.3, 3.4, 3.5}, $f_o = 1/(300 \times 0.0004) \approx 8.33$

Hz, which can be seen to match the FFT result in Figure 3.5.

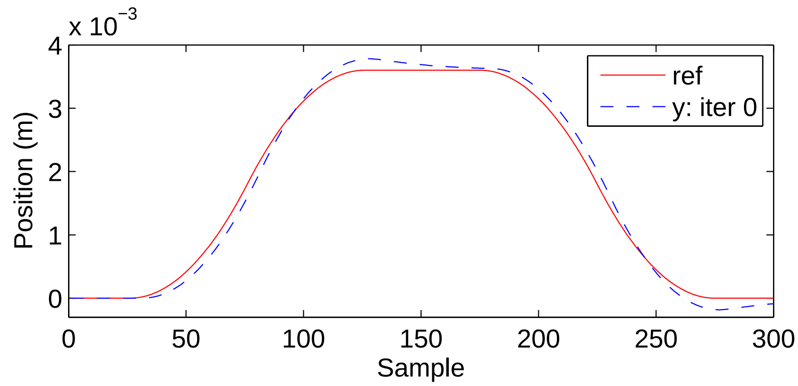


Figure 3.3: An example scanning trajectory

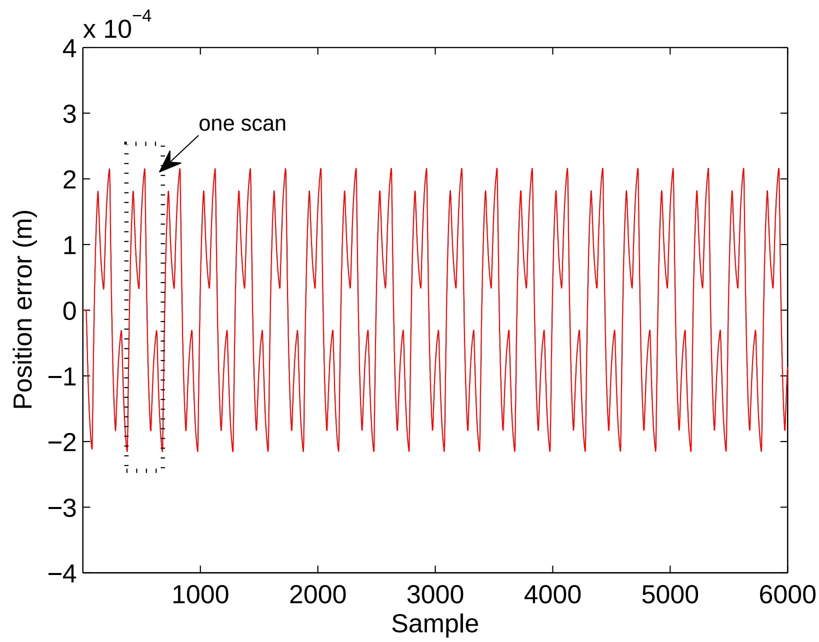


Figure 3.4: Errors in a repeated scanning process

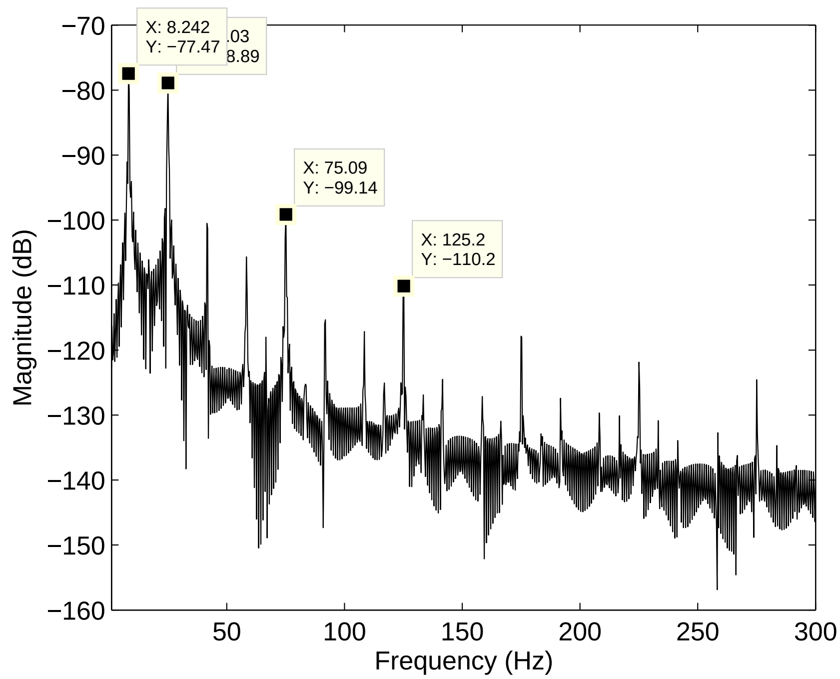


Figure 3.5: Error spectrum of the data in Figure 3.4

Figure 3.6 shows a section of an experimental scanning result when we applied a longer reference trajectory. The control of the position errors is particularly important when the stage is moving at a constant speed. Actual wafer scanning occurs here. A zoomed-in view at the constant-speed region indicates that small ripples exist in the position error. The ripple structure is more clearly explained in Figure 3.7, where we have plotted the spectrum of the position error collected at the constant-speed regions. The peak at around 18 Hz is the main contributor to the ripples in Figure 3.6 (we can confirm the frequency by computing the period of the ripple in the time domain). The error source is the environmental vibrations from the motor and motor drivers. Despite that there is a passive vibration isolation table, vibrations still have a strong residual effect on the servo system.

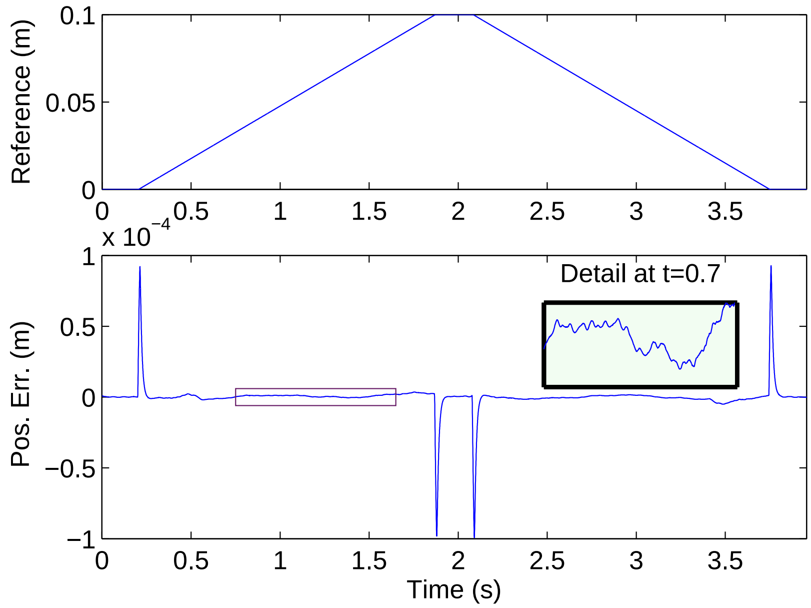


Figure 3.6: Time-domain tracking result for a long scanning process

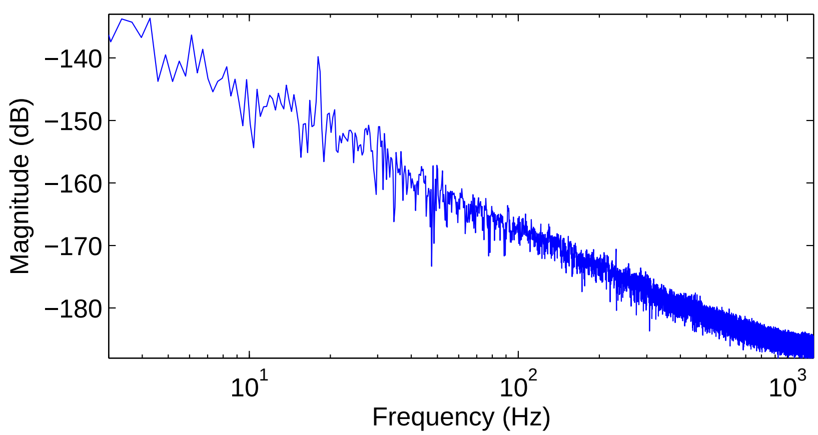


Figure 3.7: Spectrum of the errors in the constant-speed scanning region

3.2 Hard Disk Drives

Hard disk drives (HDDs) are essential data storage devices with deep integrations of nano-scale tribology, dynamics, and control engineering.

In a modern HDD, data/information is stored in circular patterns of magnetization known as data tracks or simply, tracks. During reading and writing of the data, the disk spins and the read/write head is controlled to follow the circular tracks, as shown in Figure 3.8. This creates the track-following problem, where the servo system performs regulation control to position the read/write head at the desired track, with as low variance as possible.

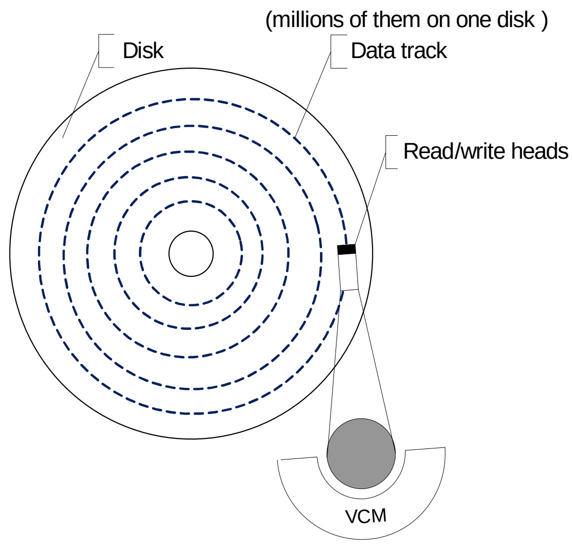


Figure 3.8: Control-related components in a hard disk drive

The actuator in a single-stage HDD is powered by a voice coil motor (VCM). Figure 3.9 shows a typical frequency response of an HDD plant that includes the power amplifier, the voice-coil motor, and the actuator mechanics. The input and the output are the voltage to the VCM and the position of the read/write head mounted at the end of the actuator. Similar to Figure 3.2, by Newton's law, the dynamics again has a nominal response of a double integrator. The low-frequency mode at around 80 Hz is due to friction. The multiple high-frequency modes above 3000 Hz are due to structural resonances. In this example, the disk has a rotational speed of 7200 revolutions per minute (rpm). At every revolution of the disks, 220 measurements are obtained, at a sampling frequency of $220 \times 7200/60\,(= 26,400)$ Hz.

With the ever increasing demand of larger capacity in HDDs, piezoelectric-based dual-stage actuation has become an essential technique to break the bottleneck of the servo performance in single-actuator HDDs [65, 66]. A dual-stage HDD applies an additional piezoelectric microactuator (MA) at the end of the VCM actuator, as shown in Figure 3.10. The MA stage has much smaller moving range but greatly improved positioning speed and accuracy. Its dynamical response is also much easier to control, as shown in the frequency response in Figure 3.11. Compared to the VCM actuator, the MA has enhanced mechanical performance in the high-frequency region, providing the capacity to greatly increase the servo bandwidth and disturbance-attenuation capacity. Only the position error of the read/write head is measurable in practice. The plant is hence a dual-input-single-output system.

Both Figure 3.9 and Figure 3.11 are from benchmark problems of HDD control, and have been frequently used in the disk drive community for algorithm verification and performance

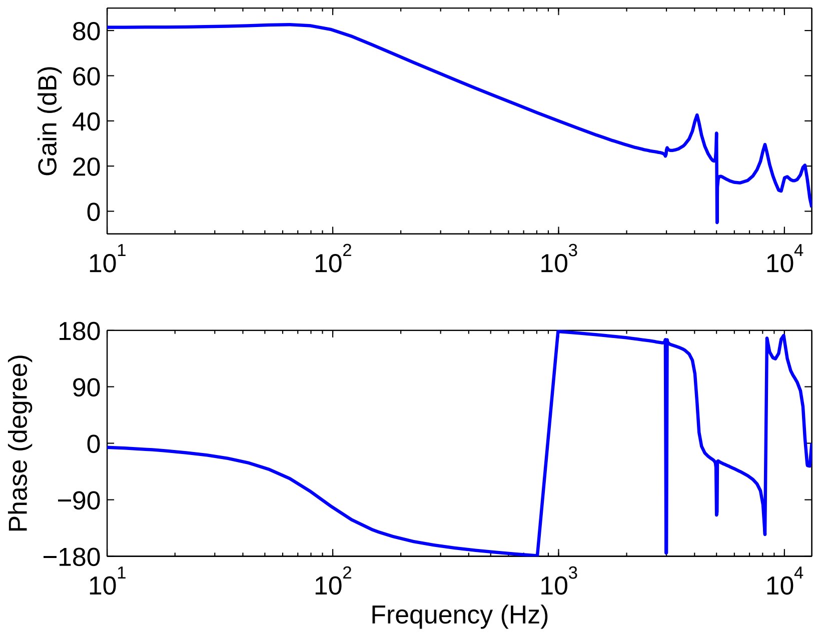


Figure 3.9: Typical frequency response of the plant in a single-stage HDD

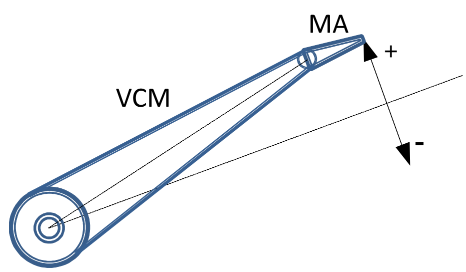


Figure 3.10: Schematic structure of dual-stage HDDs

comparison. Specifically, the system in Figure 3.11 is from page 195 of the book [1]. It is the model of a commercial Maxtor 51536U3 dual-stage HDD. Figure 3.9 is from [67], which is prepared by IEE Japan Technical Committee for Novel Nanoscale Servo Control.

There are many control problems in HDD systems. We discuss two of them below.

Audio-vibration Rejection

With the ever increasing demand of HDD applications in multimedia environments, external vibrations generated from sounds and environments are becoming a dominating disturbance source in various HDD products. These audio vibrations from, e.g., the computer/TV speakers, deteriorate the HDD servo performance by introducing strong and wide peaks to the

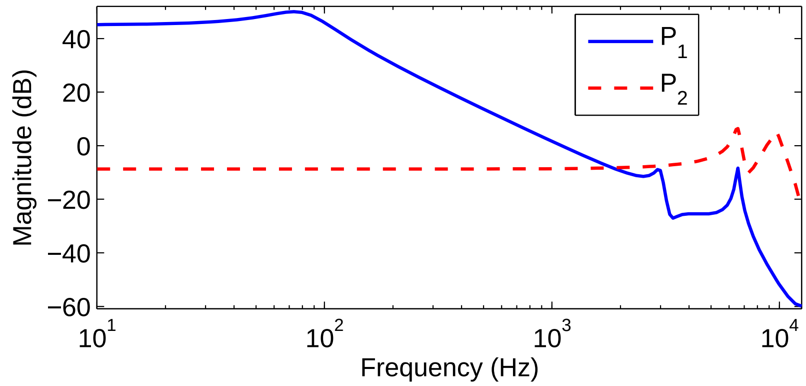


Figure 3.11: Magnitude responses of an dual-stage HDD plant

PES spectrum. Due to the nature of the disturbances [68], these vibrations occur in several concentrated bands of frequencies, near or even above the bandwidth of the servo system. Figure 3.12 demonstrates the effect of a set of measured audio vibrations projected on the benchmark problem [67]. We can observe that strong vibrations occur at as high as 2500 Hz, which is much higher than the current bandwidth (below 1500 Hz) of typical single-stage HDD systems.

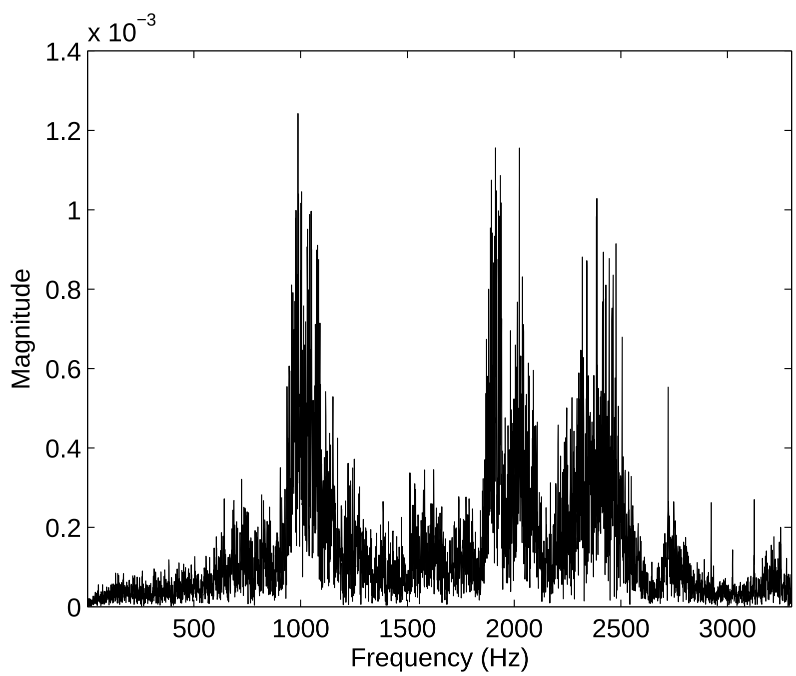


Figure 3.12: A typical HDD error spectrum under audio vibrations

Multiple Narrow-band Disturbance Rejection

Narrow-band disturbances, as depicted in Figure 3.13, show up in the PES spectrum as very sharp spikes compared to the audio vibrations in Figure 3.12. The sources of the narrow-band disturbances in HDD include track eccentricity/repeatable runout, turbulent air flow in the compact disk enclosure, and environmental vibrations such as fan noise in computers. Essentially, at nano-scale precision, tracks can no longer be treated as perfectly circular; and the disk behaves not as a perfect plate but has flutter/vibration motions in the axial direction. When audio vibrations are not present, narrow-band vibrations contribute to a large portion of the PES [69, 70, 71]. They occur at different frequencies both below and above the servo bandwidth, with different characteristics in different brands of products.

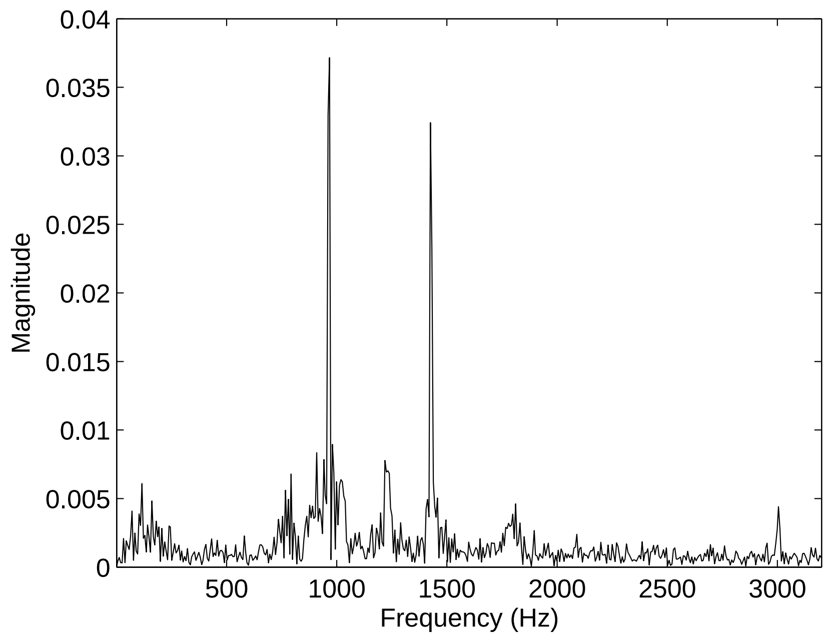


Figure 3.13: An HDD error spectrum under narrow-band disturbances

3.3 Active Suspension

Suspension systems are important components in various vibration-isolation systems (e.g., in automobiles). Figure 3.14 shows a test bed of an active suspension system built in GIPSA-Lab, Grenoble, France. It was used as a benchmark for adaptive regulation in a special session about vibration rejection in the 2013 European Control Conference [72] and in 2013 European Journal of Control [25].

Vibrations are generated from a mechanical shaker at the bottom of the system. An inertial actuator is controlled to provide forces to counteract the vibrations. The residuals are then transmitted to a passive damper (on the top of Figure 3.14), and measured by a force sensor. The input and the output of the system are respectively the input to the inertial actuator and the measured residual force. The system is controlled by MATLAB in real time, at a sampling frequency of $F_s = 800$ Hz.

The solid line in Figure 3.15 presents the frequency response of the plant. It can be observed that the plant has a group of resonant and anti-resonant modes, especially at around 50 Hz and 100 Hz. Additionally, the system is open-loop stable but has multiple lightly damped mid-frequency zeros and high-frequency non-minimum phase zeros. These characteristics place additional challenges not just for adaptive disturbance rejection, but also for general feedback control.

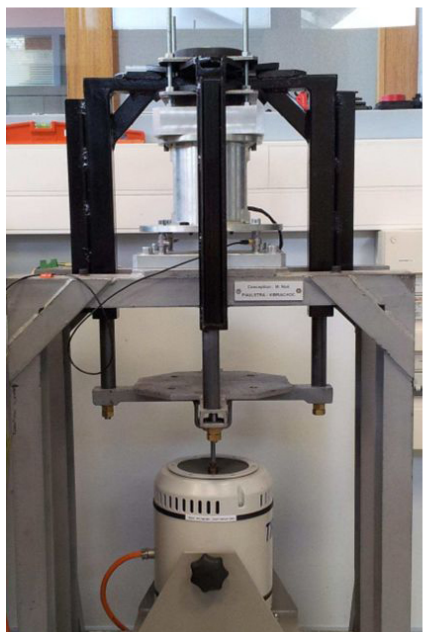


Figure 3.14: Picture of an active suspension system

A set of disturbance profiles, consisting of narrow-band vibrations at unknown and/or time-varying frequencies, are applied to the system for evaluation of the control scheme. Two example profiles are shown in Figures 3.16 and 3.17. Both the disturbance-injection time and the timing of the frequency variation, are unknown to the designers. A comprehensive set of tests at different frequency values is performed. The results are compared with six other participants of the benchmark.

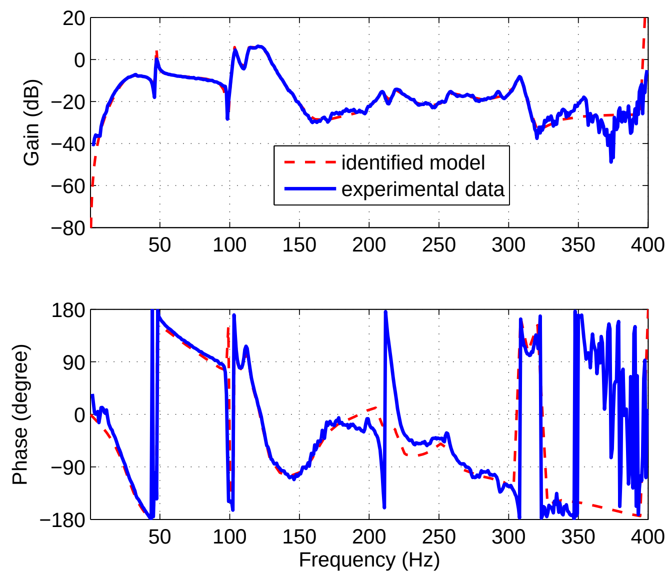


Figure 3.15: Frequency response of the plant in Figure 3.14

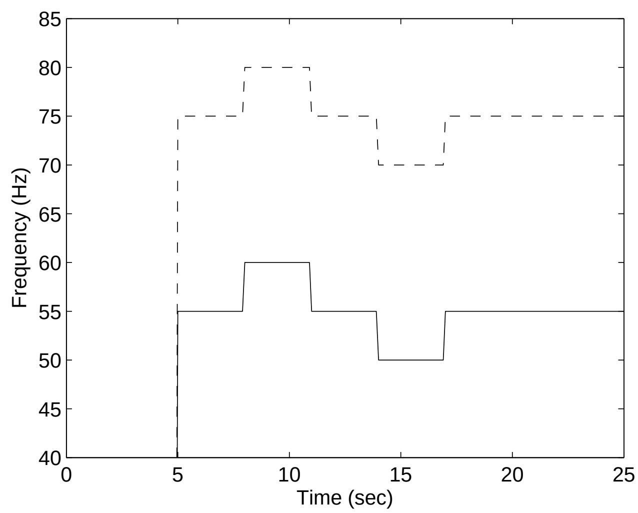


Figure 3.16: Frequencies of time-varying vibrations on an active suspension: example one

3.4 Active Steering in Automotive Vehicles

Variable-gear-ratio steering (VGRS) is a speed-dependent steering system in automotive vehicles. It uses a variable actuator to change the relative angle between the steering wheel and the steering shaft. This advanced steering feature brings various advantages. For example,

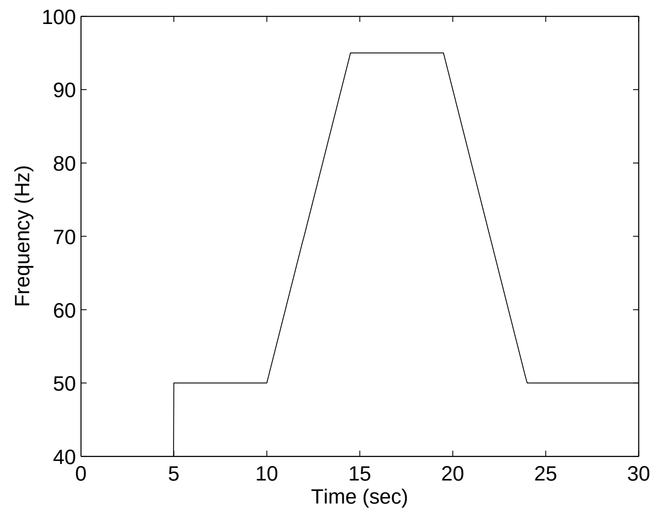


Figure 3.17: Frequencies of time-varying vibrations on an active suspension: example two

the steering gear ratio is set to be high when the vehicle speed is low. This allows a small steering-wheel motion to provide a large steering angle, which is desirable in situations such as parking. On the other hand, at high speeds, a smaller steering gear ratio can be used, therefore enhancing vehicle stability in situations such as driving on the freeway.

A serious problem of this advanced steering system is that unexpected torque is relayed to the driver whenever the variable-gear-ratio control is activated. To see this, suppose a command is given to change $\Delta\theta$ in Figure 3.18 by a certain degree. The variable actuator needs to generate torque to cause this change of angle. As a consequence, a reaction torque will be transmitted to the steering wheel, which is unexpected from the human driver. If the human tries to maintain his/her steering position, (s)he will feel a strong torque on the steering wheel. Besides in VGRS systems, unnatural torque is common when other active steering schemes are used to assist the drivers. For instance, in lateral-wind compensation, an active change of steering-shaft angle is applied to compensate perturbations in lateral acceleration, which may cause unnaturalness for the drivers.

One way to address the problem is to configure the VGRS and the electrical power steering (EPS) system as shown in Figure 3.18. The EPS system uses electrical motors[1] to provide assistive torque on the pinion shaft. The assistive torque can be used for regular steering assistance and also unnatural-torque compensation.

Figure 3.19 shows an experimental steering system with EPS and VGRS. The VGRS system is realized by a harmonic gear and a DC brushless motor. The control algorithm is evaluated on a hardware-in-the-loop (HIL) simulator. The real-time operation environment is realized by National Instruments' LabVIEW Real-Time for ETS targets, which displays simulated road views on computer screens. In this system, the command for changing the

[1]in contract to hydraulic power steering.

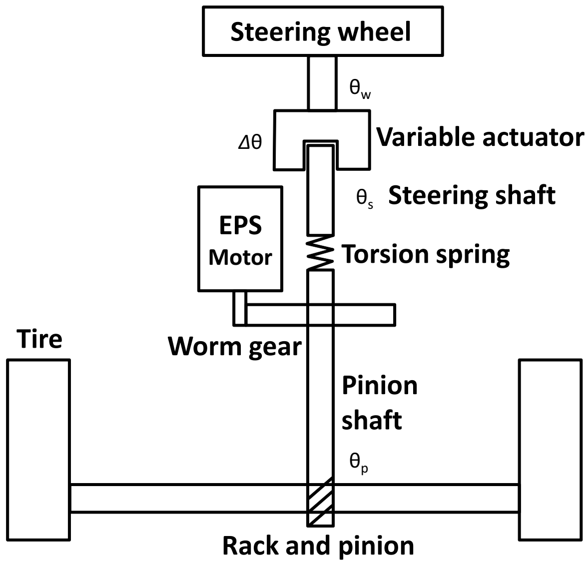


Figure 3.18: Structure of a variable-gear-ratio steering system

gear ratio in VGRS is determined by some high-level control algorithms. The low-level EPS control system is designed such that the velocity of the variable actuator does not affect the steering torque.

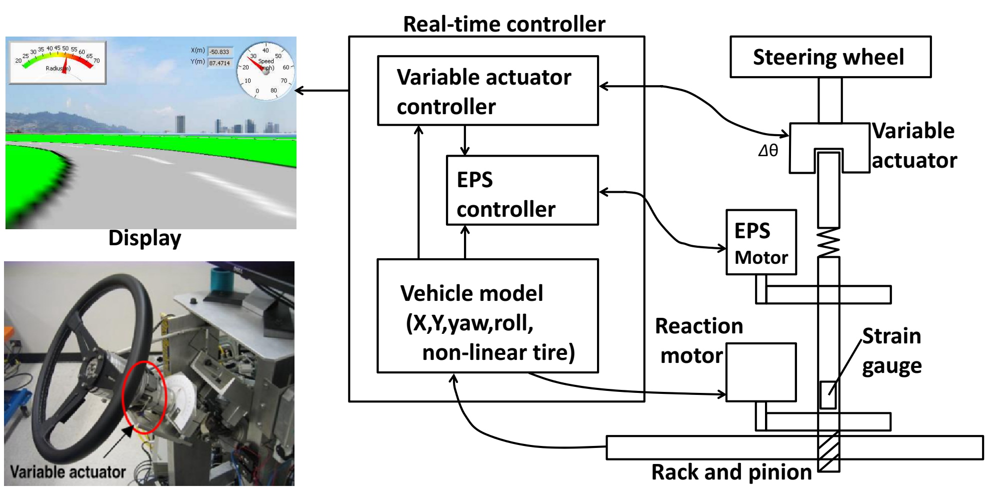


Figure 3.19: Experimental setup of a steering system with EPS and VGRS

The problem of structured-error compensation occurs due to mainly hardware imperfections and the task nature. With normal EPS, Figure 3.20 shows an example of the unnatural torque during steering. Here, the driver tries to maintain a constant steering-wheel angle and the VGRS system assists to apply a periodic steering angle. Ideally, the driver should feel a constant steering torque on the wheel. However, from the solid line, we observe that the actual steering torque has various vibrations/ripples accounting for unnaturalness during driving. This is just one common test for the steering system. With different assistant steering characteristics, the ripple frequencies also change. One source of such errors is the imperfect interactions between magnetic fields and conductors in the motor, which can be amplified in the change of gear ratios by VGRS.

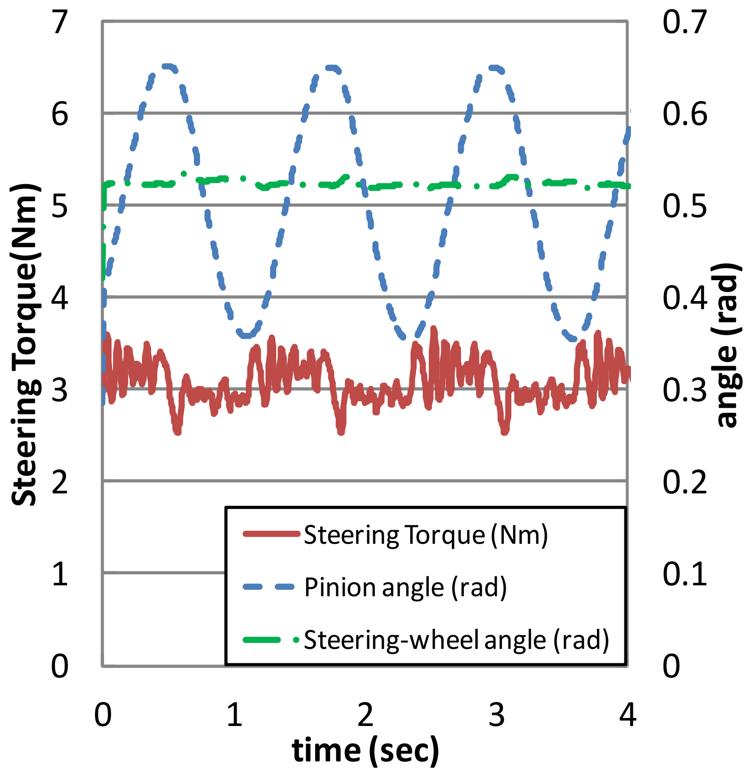


Figure 3.20: Demonstration of unnatural torque in VGRS systems

Part II

Deterministic Local Loop Shaping

Chapter 4

A Pseudo Youla-Kucera Parameterization

Motivated by the practical control problems in previous chapters, we now discuss the realization methods of local loop shaping (LLS). The main goal of this chapter is to establish the general framework for flexible LLS.

We start with a special case of YK parameterization:

4.1 Motivational Example

Consider a *stable* negative feedback loop where the plant P and the controller C are rational and proper. Standard techniques such as PID, lead-lag, and H_∞ control, can be applied to design this controller C, to yield a baseline loop shape as discussed in Section 2.1.

Assume that P is *stable* in this section, and consider the following new controller

$$\tilde{C} = \frac{C + Q}{1 - PQ} \tag{4.1}$$

where Q is a rational transfer function that is *stable* and proper. The new sensitivity and complementary sensitivity functions are

$$\tilde{S} = \frac{1}{1 + P\tilde{C}} = \frac{1 - PQ}{1 + PC} = S_o \times (1 - PQ) \tag{4.2}$$

$$\tilde{T} = 1 - \tilde{S} = \frac{PC + PQ}{1 + PC} = T_o + S_o PQ \tag{4.3}$$

where $S_o \triangleq 1/(1 + PC)$ and $T_o \triangleq PC/(1 + PC)$ are the sensitivity and complementary sensitivity functions of the baseline closed loop. The enlarged closed-loop system has the following advantageous properties:

1. $\tilde{S}$ and $\tilde{T}$ are both affine functions of Q. Moreover, $\tilde{S}$ is decomposed to be the product of the original sensitivity function $1/(1 + PC)$ and the add-on element $1 - PQ$.

2. By observation, when P and Q are stable, $\tilde{S}$ and $\tilde{T}$ are also stable. Actually, the closed-loop system with (4.1) has guaranteed stability, namely, $\tilde{C}$ is a stabilizing controller. A standard stability proof in the literature of YK parameterization is to derive the four transfer functions that describe all signal relations in a feedback loop in Figure 4.1.

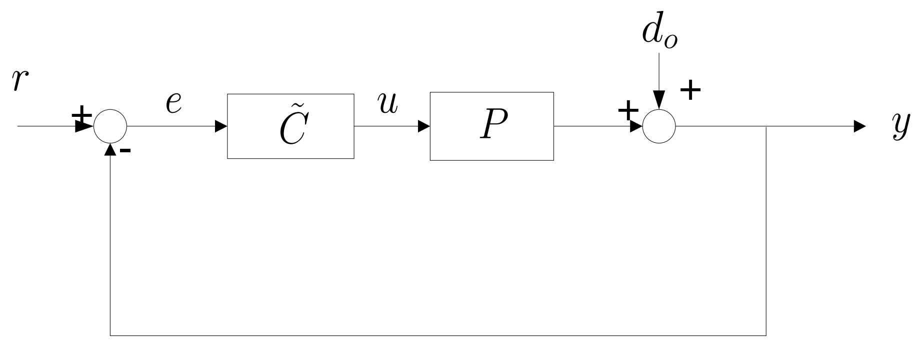


Figure 4.1: General block diagram for feedback control

We compute

$$\begin{bmatrix} y \\ u \end{bmatrix} = \begin{bmatrix} \tilde{G}_{yr} & \tilde{G}_{yd} \\ \tilde{G}_{ur} & \tilde{G}_{ud} \end{bmatrix} \begin{bmatrix} r \\ d_o \end{bmatrix}$$

When $\tilde{C} = (C+Q)/(1-PQ)$, we have

$$\tilde{G}_{yd} = \tilde{S} = \frac{1-PQ}{1+PC} \tag{4.4}$$

$$\tilde{G}_{yr} = \tilde{T} = \frac{PC+PQ}{1+PC} \tag{4.5}$$

$$\tilde{G}_{yu} = P\tilde{S} = \frac{P(1-PQ)}{1+PC} \tag{4.6}$$

$$\tilde{G}_{ud} = \frac{\tilde{C}}{1+P\tilde{C}} = \frac{C+Q}{1+PC} \tag{4.7}$$

From the stable baseline feedback loop, we know that $G_{yd} = 1/(1+PC)$, $G_{yr} = PC/(1+PC)$, $G_{yu} = P/(1+PC)$, and $G_{ud} = C/(1+PC)$, are all stable. Hence equations (4.4) to (4.7)—composed by direct multiplications or summations of G_{yd}, G_{yr}, G_{yu}, G_{ud}, P, and Q—are all stable if P and Q are stable.

3. If C is additionally stable,[1] then *any* feedback controller that stabilizes P can be expressed in the form of (4.1), by choosing a proper Q that is stable and rational. Proof for this result can be easily done after we introduce some theorems in Section

[1]Note: in the first two properties, the only assumptions we made are that the plant and the baseline feedback loop are stable.

4.2. At the moment, we provide the following intuition: let C be the baseline stabilizing controller and $\tilde{C}$ be an arbitrary stabilizing controller. Solving (4.1) for Q we can get

$$Q = \frac{\tilde{C}}{1+P\tilde{C}} - \frac{C}{1+P\tilde{C}} \tag{4.8}$$

We have proven that any stable Q in the discussed controller structure will yield a stable closed-loop system. Notice that $\tilde{C}/\left(1+P\tilde{C}\right)$ and $1/\left(1+P\tilde{C}\right)$ in (4.8) are both stable (since they are two of the closed-loop transfer functions in the stable loop consisting of P and $\tilde{C}$). When C is stable, Q is clearly stable. Hence with (4.8) we can obtain the stabilizing controller $\tilde{C}$.

Since stability has already been guaranteed, the affine form of Q in $\tilde{S} = (1-PQ)\,/\,(1+PC)$ makes it quite easy to directly shape the sensitivity function. The main concept is that we can use well-developed tools to design the baseline sensitivity $1/(1+PC)$, and consider $1-PQ$ for the add-on local loop shaping. In a sense, the original feedback design problem is then transformed to a feedforward (and hence simpler and more intuitive) one.

4.2 Standard Youla-Kucera Parameterization

In order to present the full results of Youla-Kucera parameterization, we review several definitions first:

Definition. Let

$$\mathcal{S} \triangleq \{\text{stable, proper, rational transfer functions}\}$$
$$\mathcal{R} \triangleq \{\text{proper, rational transfer functions}\}$$

Definition 4.1. (**coprime**) Suppose $N \in \mathcal{S}$, $D \in \mathcal{S}$. The pair (N, D) is called coprime over $\mathcal{S}$ if there exists $U \in \mathcal{S}$, $V \in \mathcal{S}$ such that

$$UN + VD = 1 \tag{4.9}$$

If in addition $G = N/D$ and D^{-1} is realizable, then (N, D) is called a coprime factorization of G.

Since D^{-1} is realizable and $D \in \mathcal{S}$ (hence D is proper/realizable), we have the following immediate result:

Fact 4.1. *For a transfer function G with the coprime factorization $G = N/D$, the relative degree of D must be zero.*

Example 4.1. If

$$G(z) = z^{-1}\frac{1+z^{-1}}{1-0.9z^{-1}}$$

then $N(z) = z^{-1}\frac{1+z^{-1}}{1+0.95z^{-1}}$ and $D(z) = \left(\frac{1+0.95z^{-1}}{1-0.9z^{-1}}\right)^{-1}$ form a coprime factorization for $G(z)$, with

$$\frac{0.9^2}{1.9}N(z) + \left(1+\frac{0.9}{1.9}z^{-1}\right)D(z) = 1.$$

Remark 4.1. Coprime factorizations are not unique. For instance, $N(z) = z^{-1}(1+z^{-1})$ and $D(z) = 1-0.9z^{-1}$ also form a valid coprime pair for $G(z)$ in Example 4.1.

The general Youla-Kucera parameterization for SISO systems is presented in the next Theorem.

Theorem 4.1. *If a plant $P = N/D$ can be stabilized by a negative-feedback controller $C = X/Y$, with $(N,\ D)$ and $(X,\ Y)$ being coprime factorizations over $\mathcal{S}$, then any stabilizing feedback controller can be parameterized as*

$$C_{all} = \frac{X+DQ}{Y-NQ}:\ Q\in\mathcal{S},\ Y(\infty)-N(\infty)Q(\infty)\neq 0. \tag{4.10}$$

Proof. See [63]. □

With (4.10), the new sensitivity function is (after some simplifications)

$$\tilde{S} = \frac{1}{1+PC_{all}} = \frac{1}{1+PC}\left[1-\frac{N}{Y}Q\right] = \frac{1}{1+PC} - \frac{1}{1+PC}\frac{N}{Y}Q \tag{4.11}$$

which is again, affine in Q, and stable if Q is stable.[2]

[2]More specific, $1/(1+PC)$ is stable by assumption; the last term on the right hand side of (4.11) is also stable, as

$$\frac{1}{1+PC}\frac{N}{Y} = \frac{1}{1+\frac{N}{D}\frac{X}{Y}}\frac{N}{Y} = \frac{DN}{DY+NX}$$

where$D$ and N are stable by definition, and $1/(DY+NX)$ is stable as shown below [63]:

Let (U_C, V_C) and (U_P, V_P), both in $\mathcal{S}$, be the elements that comprise the identity (4.9). Then

$$U_C X + V_C Y = 1$$
$$U_P N + V_P D = 1$$

Notice that

$$\begin{aligned}\begin{bmatrix} V_C & U_C \end{bmatrix}\begin{bmatrix} \frac{1}{1+PC} & \frac{P}{1+PC} \\ \frac{C}{1+PC} & \frac{PC}{1+PC}\end{bmatrix}\begin{bmatrix} V_P \\ U_P \end{bmatrix} &= \begin{bmatrix} V_C & U_C \end{bmatrix}\begin{bmatrix} \frac{DY}{NX+DY} & \frac{NY}{NX+DY} \\ \frac{XD}{NX+DY} & \frac{XY}{NX+DY}\end{bmatrix}\begin{bmatrix} V_P \\ U_P \end{bmatrix} \\ &= \frac{1}{NX+DY}\end{aligned} \tag{4.12}$$

All elements on the left hand side of (4.12) are stable, hence $1/(NX+DY)$ is stable.

The most appealing property is the guaranteed stability under YK parameterization. This part is easy to see from the sensitivity function (4.11), and can be proved similar to that in Section 4.1. Next we provide some intuition about the concept of "*all* stabilizing controllers". Let $\tilde{C}$ be any controller that stabilizes the closed loop. Solving

$$\tilde{C} = \frac{X + DQ}{Y - NQ}$$

we get

$$Q = \frac{Y\tilde{C} - X}{N\tilde{C} + D}. \tag{4.13}$$

It is not difficult to show that (4.13) is firstly stable. From (4.11), the sensitivity function is

$$\begin{aligned}\tilde{S} = \frac{1}{1 + P\tilde{C}} &= \frac{1}{1 + PC}\left[1 - \frac{N}{Y}Q\right] \\ &= \frac{1}{1 + \frac{N}{D}\frac{X}{Y}}\left[1 - \frac{N}{Y}\frac{Y\tilde{C} - X}{N\tilde{C} + D}\right] \\ &= \frac{D}{DY + NX}\frac{DY + NX}{D + N\tilde{C}}\end{aligned}$$

Hence we see that by (4.13), there are cancellations of the stable baseline poles that satisfies

$$DY + NX = 0.$$

In addition, new poles from

$$D + N\tilde{C} = 0$$

(recall that the closed loop characteristic polynomial comes from $1 + \frac{N}{D}\tilde{C} = 0$) are introduced, hence the equivalence (at the steady state) with the new controller $\tilde{C}$.

It can now be seen that, in the choice of (4.1), the plant and the baseline controller are parameterized by $P = P/1$ and $C = C/1$, namely, $N = P$, $D = 1$, $X = C$, and $Y = 1$ in (4.10). These are valid coprime factorizations when P and C are stable, since we can choose, e.g., $U = C/(1 + PC) \in \mathcal{S}$ and $V = 1/(1 + PC) \in \mathcal{S}$ to make $UN + VD = 1$.

Remark 4.2. About the inequality $Y(z = \infty) - N(z = \infty)Q(z = \infty) \neq 0$ in (4.10):

- It makes the closed-loop transfer functions proper and rational [63]. Otherwise, if $Y(\infty) - N(\infty)Q(\infty) = 0$, the relative degree of $Y - NQ$ is at least one, which makes $1/(Y - NQ)$ not proper and C_{all} not realizable during implementation. More specific, consider the example where $P = (1.1 + z^{-1})/(1.2 + z^{-1})$ and $C = 1$. The baseline closed loop is clearly stable here. If we let $X = Y = 1$, $N = 1.1 + z^{-1}$, and $D = 1.2 + z^{-1}$, then

$$C_{all}(z) = \frac{1 + (1.2 + z^{-1})\,Q(z)}{1 - (1.1 + z^{-1})\,Q(z)},\ Q(z) = \frac{1}{1.1 + z^{-1}} \tag{4.14}$$

satisfies all other conditions of YK parameterization, except that

$$Y\left(z=\infty\right)-N\left(z=\infty\right)Q\left(z=\infty\right)=0.$$

Clearly, the controller $C_{all}(z)$ in (4.14) is not realizable, as its denominator is zero. Actually, very simple verification shows that, for any

$$Q\left(z\right)=\frac{1+b_{Q1}z^{-1}+b_{Q2}z^{-2}+\dots}{1.1+z^{-1}},\ b_{Qi}\in\mathbb{R}$$

$C_{all}\left(z\right)$ will not be proper.

- $Y(\infty)-N(\infty)Q(\infty)\neq 0$ is however very easy to satisfy for practical problems. For instance, consider the case where $P=z^{-1}$ and $C=0.8$. We can let $N=z^{-1}$, $D=1$, $X=0.8$, $Y=1$, and $Y(\infty)-N(\infty)Q(\infty)=1$. More general:
  - as the relative degree of Y is zero [Fact 4.1], we have $Y\left(\infty\right)\neq 0$;
  - and practical discrete-time plants have delays that make N contain at least one cascaded term of z^{-1}, hence $N\left(\infty\right)=0$;
  - recalling that Q is proper and rational (therefore $Q\left(\infty\right)\neq\infty$), practically we thus will have
  $$Y\left(\infty\right)-N\left(\infty\right)Q\left(\infty\right)=Y(\infty)\neq 0.$$

Remark 4.3. The original Youla-Kucera formulation requires additionally the Bezout identity $XN+YD=1$, which gives instead $S_o=YD$ and $S=\left(Y-NQ\right)D$. This assumption can be dropped without loss of generality, and has been used without proof in a group of literatures including [24] and [73]. We provide a proof below:

Proof. (Drop of the Bezout identity in Remark 4.3) Consider the coprime factorizations $P=N/D$ and $C=X/Y$. By definition, $NX+DY$ is stable. In addition, stability of the baseline closed loop gives that $\left(NX+DY\right)^{-1}$ is stable and proper[3] (due to the stability of $NX+DY$ and its own inverse, $NX+DY$ is called a unit in $\mathcal{S}$). Therefore, $\tilde{X}=X\left(NX+DY\right)^{-1}$ and $\tilde{Y}=Y\left(NX+DY\right)^{-1}$ form a coprime factorization of C.

With (N,D) being a coprime factorization of P, we have

$$N\tilde{X}+D\tilde{Y}=1 \tag{4.15}$$

i.e., the Bezout identity holds for the pairs (N,D) and $(\tilde{X},\tilde{Y})$. Now we can use the standard Youla-Kucera parameterization with $C=\tilde{X}/\tilde{Y}$, $P=N/D$, and (4.15). The parameterization of all stabilizing controllers is given by

$$\frac{\tilde{X}+D\tilde{Q}}{\tilde{Y}-N\tilde{Q}}=\frac{X\left(NX+DY\right)^{-1}+D\tilde{Q}}{Y\left(NX+DY\right)^{-1}-N\tilde{Q}}:\ \tilde{Q}\in\mathcal{S},\ \tilde{Y}\left(\infty\right)-N\left(\infty\right)\tilde{Q}\left(\infty\right)\neq 0.$$

[3]See footnote 2 on page 44.

Note that $Q(NX+DY)^{-1}$ and $\tilde{Q}$ is a one-to-one correspondence in $\mathcal{S}$. Letting $\tilde{Q} = Q(NX+DY)^{-1}$, and after cancellations, we obtain the parameterization

$$C_{all} = \frac{X+DQ}{Y-NQ} : \; Q \in \mathcal{S}$$

□

Theorem 4.1 is formulated under a linear time-invariant control framework. If we make Q time-varying, we have the following useful theorem.

Theorem 4.2. *(Stability of YK parameterization with a time-varying Q filter) Under the assumptions in Theorem 4.1. If Q is a stable time-varying filter, then the closed-loop system remains stable. In particular, switching between stable Q's does not cause instability.*

Proof. See [74, 63]. □

Theorem 4.2 is particularly useful for adaptive control, where we can adapt Q for enhanced servo performance, without introducing instability to the closed loop. Another application is in the field of switched control theory—a subclass of hybrid systems—where a time-invariant plant is controlled by a set of deterministic controllers that switch among each other at the best-suited occasion.

With the conceptual and mathematical preparations, we proceed next to the main result of this Chapter.

4.3 Pseudo Youla-Kucera Parameterization for Discrete-time Systems

From the viewpoint of implementation, a perfect plant model is not possible in practice. In this sense, a practical YK parameterization has to be an approximation, or a robust version, of the ideal case in Section 4.2. In addition, although we have achieved affine parameterization of the sensitivity function, the design of Q is still nontrivial.[4] In this section, we discuss a special choice of the plant and controller parameterization, to form a pseudo Youla-Kucera parameterization. The purpose of this approximation is to take advantage of the flexibility in Section 4.1, and to extend the design simplicity and tuning intuition under practical model imperfections.

Recall in Section 4.1, that for a stable baseline system with $P \in \mathcal{S}$ and $C \in \mathcal{S}$,

$$\tilde{C} = \frac{C+Q}{1-PQ} \tag{4.16}$$

[4]Indeed, YK parameterization has been a well-established concept for decades, in particular in the H_∞- and adaptive-control communities. However, the appreciation of its flexibility and capacity has been growing slowly in actual implementations.

parameterizes all the stabilizing controllers. Candidate Q-filter designs can be found, for example, where a linear combination of some basis transfer functions (e.g., $\sum_{i=0}^{k_Q} \theta_i z^{-i}$ in discrete-time schemes) is used to form Q, and adaptive/H_∞ control is applied to find the scaling coefficients for the combination. Regardless of the tools that are used, to make $\tilde{S} = S_o \times (1 - PQ)$ in (4.2) small at certain frequencies, we require PQ to be close to one in the same region. Consider a block-diagram realization of $\tilde{C}$ in Figure 4.2. In the inner loop marked by $\star$, Q should approximate P^{-1} at the desired frequencies. As $1 - PQ$ becomes small, the overall controller (4.16) will have high gain in the interested frequency regions.

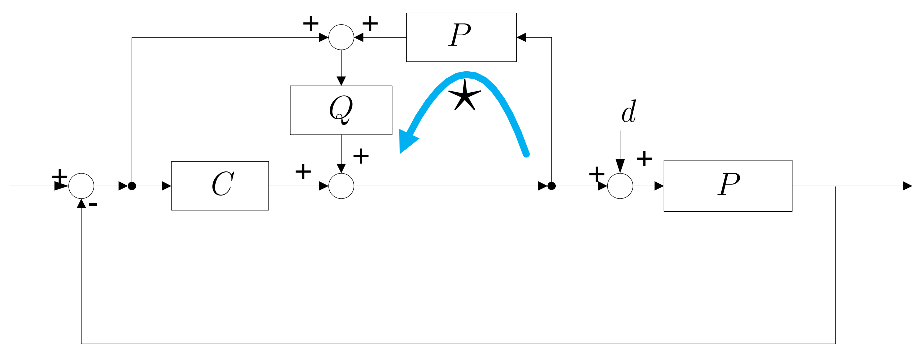


Figure 4.2: A forward-model Youla-Kucera parameterization

Instead of a direct inversion of P by Q, consider a two-step design: first to perform an explicit model inversion P^{-1}, and then to use Q for controlling the amount of loop shaping and disturbance rejection. Since a perfect model of the plant is not available anyway, we allow the adoption of a nominal inversion $\hat{P}^{-1}$. In addition, from Section 2.4, a practical plant has delays and will not be realizable under direct inversion. We therefore use instead $z^{-m}\hat{P}^{-1}(z)$, where m is the relative degree of $P(z)$. For the moment we assume $\hat{P}^{-1}(z)$ is stable. This is one constraint that we can place, if we drop the assumption of perfect model inversion, and is usually not difficult to achieve in a large class of practical systems (we will discuss more about this in Chapter 7).

A block diagram of the aforementioned idea is shown in Figure 4.3. We have now formed a pseudo Youla-Kucera parameterization. Compared to Figure 4.2, this is a discrete-time scheme that have relaxed requirement on the perfect model assumption; and modifications have been made to reduce the dependence of the plant information in the ($\star$) loop.

The transfer function of the overall feedback controller in Figure 4.3 is

$$\tilde{C}(z) = \frac{C(z) + z^{-m}\hat{P}^{-1}(z)Q(z)}{1 - z^{-m}Q(z)} \tag{4.17}$$

In (4.17), high-gain control is directly provided by $1/(1 - z^{-m}Q(z))$, instead of by $1/(1 -$

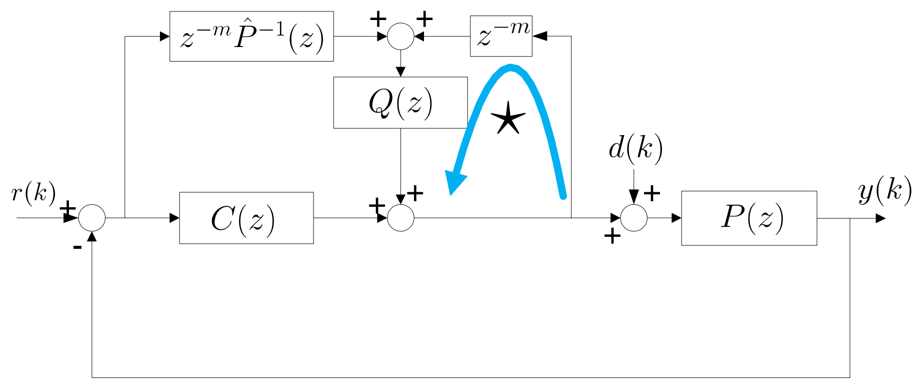


Figure 4.3: A discrete-time inverse-based YK parameterization

PQ) in (4.16). This reduced dependence on P is also reflected in the sensitivity function:

$$S(z) = \frac{1}{1 + P(z)\tilde{C}(z)} = \frac{1 - z^{-m}Q(z)}{1 + P(z)C(z) + z^{-m}Q(z)\left(\hat{P}^{-1}(z)P(z) - 1\right)} \tag{4.18}$$

If $\hat{P}(e^{j\omega}) = P(e^{j\omega})$, namely, at the frequencies where good plant information is available, (4.18) gives

$$S(e^{j\omega}) = \frac{1 - e^{-mj\omega}Q(e^{j\omega})}{1 + P(e^{j\omega})C(e^{j\omega})} \tag{4.19}$$

$$= S_o(e^{j\omega})(1 - e^{-jm\omega}Q(e^{j\omega})) \tag{4.20}$$

In the frequency regions where large model mismatch exists, we will make $|Q(e^{j\omega})|$ small such that the contribution of $z^{-m}Q(z)(\hat{P}^{-1}(z)P(z) - 1)$ in (4.18) is still insignificant to make (4.20) a valid approximation. Of course, a default assumption is that good model information can be obtained at frequencies where enhanced servo is desired. If the system already has large uncertainties at the disturbance frequencies, it is best not to apply large control effort there in the first place.

By the above constructions, similar to an ideal Youla-Kucera parameterization, we have separated the baseline system response $1/(1 + P(z)C(z))$ from $S(z)$ in (4.19), and can now focus on designing the much simplified term $1 - z^{-m}Q(z)$. The use of the inverse model $\hat{P}^{-1}(z)$ has helped to make this added module $1 - z^{-m}Q(z)$ simple and depend little on the dynamics of $P(z)$ (only the plant delay z^{-m} appears here). Additionally, stability and robust stability are easily satisfied. For instance in the nominal case without modeling errors $S(z) = S_o(z)(1 - z^{-m}Q(z))$, which is stable since $Q(z)$ has been designed to be stable. For the robust stability, as long as $z^{-m}Q(z)(\hat{P}^{-1}(z)P(z) - 1)$ is small, we can see that (4.20) is still a valid approximation and $S(z)$ is also expected to be stable. More formally, we have the following theorem:

Theorem 4.3. *(Nominal stability of the pseudo YK parameterization) Consider a stable discrete-time baseline negative feedback loop consisting of the controller $C(z)$, and the plant $P(z)$ whose relative degree is m. If $P(z) = \hat{P}(z)$, and $Q(z)$ and $\hat{P}^{-1}(z)$ are stable in*

$$\tilde{C}(z) = \frac{C(z) + z^{-m}\hat{P}^{-1}(z)Q(z)}{1 - z^{-m}Q(z)} \tag{4.21}$$

then the new feedback loop consisting of $P(z)$ and $\tilde{C}(z)$ has guaranteed stability.

Remark. notice that we are only considering stability and not attempting to parameterize all the stabilizing controllers here.

Proof. The four essential transfer functions for the new feedback loop are

$$\tilde{G}_{yd}(z) = \tilde{S}(z) = \frac{1 - z^{-m}Q(z)}{1 + P(z)C(z)} \tag{4.22}$$

$$\tilde{G}_{yr}(z) = \tilde{T}(z) = \frac{P(z)C(z) + z^{-m}Q(z)}{1 + P(z)C(z)} \tag{4.23}$$

$$\tilde{G}_{yu}(z) = P(z)\tilde{S}(z) = \frac{P(z)(1 - z^{-m}Q(z))}{1 + P(z)C(z)} \tag{4.24}$$

$$\tilde{G}_{ud}(z) = \tilde{C}(z)\tilde{S}(z) = \frac{C + z^{-m}\hat{P}^{-1}(z)Q(z)}{1 + P(z)C(z)} \tag{4.25}$$

Given stability of the baseline transfer functions

$$\frac{1}{1 + P(z)C(z)},\ \frac{P(z)}{1 + P(z)C(z)},\ \frac{C(z)}{1 + P(z)C(z)},\ \frac{P(z)C(z)}{1 + P(z)C(z)},$$

and the stability of $Q(z)$ and $\hat{P}^{-1}(z)$, (4.22) to (4.25) are all stable by observation, hence the validity of Theorem 4.3.

A more direct proof that reveals the actual order of the system and the distribution of closed-loop poles is as follows: let

$$P(z) = \frac{B_P(z)}{A_P(z)},\ \hat{P}(z) = \frac{B_{\hat{P}}(z)}{A_{\hat{P}}(z)},\ C(z) = \frac{B_C(z)}{A_C(z)},\ Q(z) = \frac{B_Q(z)}{A_Q(z)}$$

where $B_G(z)$ and $A_G(z)$ denote, respectively, the numerator and the denominator of a transfer function $G(z)$. We have

$$\begin{aligned}\tilde{C}(z) &= \frac{\frac{B_C(z)}{A_C(z)} + z^{-m}\frac{A_{\hat{P}}(z)}{B_{\hat{P}}(z)}\frac{B_Q(z)}{A_Q(z)}}{1 - z^{-m}\frac{B_Q(z)}{A_Q(z)}} \\ &= \frac{A_Q(z)B_{\hat{P}}(z)B_C(z) + z^{-m}A_C(z)A_{\hat{P}}(z)B_Q(z)}{B_{\hat{P}}(z)A_C(z)\left[A_Q(z) - z^{-m}B_Q(z)\right]}\end{aligned} \tag{4.26}$$

From $1 + P\tilde{C} = 0$, after some algebra, we can obtain the closed-loop characteristic equation:

$$\begin{aligned} &A_Q(z)\, B_{\hat{P}}(z)\left[A_C(z)\, A_P(z) + B_C(z)\, B_P(z)\right] \\ &\qquad + z^{-m} A_C(z)\, B_Q(z)\left[A_{\hat{P}}(z)\, B_P(z) - A_P(z)\, B_{\hat{P}}(z)\right] = 0 \end{aligned} \tag{4.27}$$

If $P(z) = \hat{P}(z)$, (4.27) reduces to

$$A_Q(z)\, B_{\hat{P}}(z)\left[A_C(z)\, A_P(z) + B_C(z)\, B_P(z)\right] = 0 \tag{4.28}$$

Hence the closed-loop poles are composed of: the baseline closed-loop poles, and the poles of $Q(z)$ and $\hat{P}^{-1}(z)$. The stability result of Theorem 4.3 then immediately follows. □

Notice that the theorem applies to unstable plants and unstable baseline controllers as well. If the baseline controller is stable itself, then the above scheme is a special Youla-Kucera parameterization. We have

Lemma 4.1. *(A special case of discrete-time YK parameterization) Consider a stable discrete-time baseline negative feedback loop consisting of the controller $C(z)$, and the plant $P(z)$ whose relative degree is m. If $P^{-1}(z)$ and $C(z)$ are stable, then we can use the coprime factorizations*

$$P(z) = \frac{z^{-m}}{z^{-m} P^{-1}(z)},\ C(z) = \frac{C(z)}{1} \tag{4.29}$$

to form the set of all stabilizing controllers

$$\left\{ C_{all}(z) \in \mathcal{R} :\ C_{all}(z) = \frac{C(z) + z^{-m} P^{-1}(z)\, Q(z)}{1 - z^{-m} Q(z)},\ Q(z) \in \mathcal{S} \right\} \tag{4.30}$$

and the sensitivity function is

$$S(z) = \frac{1 - z^{-m} Q(z)}{1 + P(z)\, C(z)} = S_o(z)\left(1 - z^{-m} Q(z)\right) \tag{4.31}$$

Proof. follows directly by letting $N(z) = z^{-m}$, $D(z) = z^{-m} P^{-1}(z)$, $X(z) = C(z)$, and $Y = 1$ in Theorem 4.1. Notice that the condition that $Y(\infty) - N(\infty)\, Q(\infty) = 1 - z^{-m} Q(z)|_{z=\infty} \neq 0$ is automatically satisfied when $Q(z)$ is rational and causal (and hence $Q(\infty)$ is finite). □

Dropping the parameterization for *all* stabilizing controllers

We emphasize that it is not necessary to be able to express all the stabilizing controllers in practical situations. For instance, a high-pass filter may be able to stabilize a plant. However, this design would seldom be recommended in practice, as the high-frequency performance and robustness would be seriously challenged under practical plant uncertainties and sensor noises. A sub class of all the stabilizing controllers, with promising high-performance feedback potentials, would suffice in practical loop shaping.

This drop of the parameterization for *all* stabilizing controllers is made by allowing the baseline controller $C(z)$ to be not strictly stable.[5] To see which class of stabilizing controllers is excluded, we recall (4.26), which indicates that regardless of the choice of $Q(z)$, $A_C(z)$ remains in the denominator of $\tilde{C}$. If $C(z)$ is unstable, $\tilde{C}(z)$ will thus also be unstable. A particular case that is useful in practice is when $C(z)$ has marginally stable poles. For example, when it contains an integrator, the result then indicates that the integrator action will be invariant under the controller parameterization in Theorem 4.3, and that we will not be able to parameterize strictly stable controllers.

We remark, however, that this does not yield practical disadvantages for the controller design. For instance, an integrator action may be needed although it is marginally stable, e.g. to remove the steady-state bias. However, this does not mean that the baseline controller may be selected without common engineering analysis in practice. Consider the following example where the plant is among the class of systems that are easiest to control

$$P(z) = \hat{P}(z) = z^{-1} P_{\text{mp}}(z)$$

Here $P_{\text{mp}}(z) = B_{P_{\text{mp}}}(z) / A_{P_{\text{mp}}}(z)$ is a minimum-phase transfer function with all poles and zeros strictly inside the unit circle. Assume the relative degree of $P_{\text{mp}}(z)$ is zero. If the baseline design is chosen as

$$C(z) = P_{\text{mp}}^{-1}(z) \frac{-1.8}{1 + z^{-1}}$$

then the closed loop characteristic equation is

$$\begin{aligned} 1 + z^{-1} \frac{B_{P_{\text{mp}}}(z)}{A_{P_{\text{mp}}}(z)} \frac{A_{P_{\text{mp}}}(z)}{B_{P_{\text{mp}}}(z)} \frac{-1.8}{1 + z^{-1}} &= 0 \\ \Leftrightarrow B_{P_{\text{mp}}}(z) A_{P_{\text{mp}}}(z) \left[1 - 0.8 z^{-1}\right] &= 0 \end{aligned}$$

So the closed-loop poles are all inside the unit circle. However if we parameterize according to (4.21), then

$$\tilde{C}(z) = \frac{\frac{-1.8}{1+z^{-1}} P_{\text{mp}}^{-1}(z) + P_{\text{mp}}^{-1}(z) Q(z)}{1 - z^{-m} Q(z)} = \frac{P_{\text{mp}}^{-1}(z)}{1 + z^{-1}} \frac{-1.8 + (1 + z^{-1}) Q(z)}{1 - z^{-m} Q(z)}$$

The controller $\tilde{C}(z)$ always has a pole at $z = -1$. This is not a wise design in motion control, as the open loop will have infinite gains at Nyquist frequency, yielding the closed loop to be extremely sensitive to high-frequency uncertainties.

[5] Of course, unstable controllers are usually less preferred unless absolutely necessary (for example, to stabilize some unstable plants). Hence the influence of the relaxation is minor in practice.

Robust stability

From Figure 4.3, intuitively, $Q(z)$ controls the additional design efforts in the pseudo Youla-Kucera parameterization. At high frequencies where the plant behavior itself can not be well predicted, the magnitude of $Q(z)$ (and hence the add-on control effort) can be kept small such that $S(z)$ remains approximately $1/(1+P(z)C(z))$ in (4.18).

More formally, if the plant is perturbed to be $\tilde{P}(z) = \hat{P}(z)(1+\Delta(z))$,[6] standard robust-stability analysis[7] gives that the closed-loop system is stable if and only if the following hold:

- nominal stability: the closed loop is stable when $\Delta(z) = 0$
- robust stability: $\forall\omega$,
$$\left|\Delta(e^{j\omega})T(e^{j\omega})\right| < 1 \tag{4.32}$$
where $T(e^{j\omega})$ comes from
$$T(z) = 1 - \frac{1 - z^{-m}Q(z)}{1+\hat{P}(z)C(z)} = \frac{\hat{P}(z)C(z) + z^{-m}Q(z)}{1+\hat{P}(z)C(z)} \tag{4.33}$$

Equation (4.32) is not difficult to satisfy. If $Q(e^{j\omega}) = 0$, (4.33) reduces to $T(e^{j\omega}) = T_o(e^{j\omega})$. We thus have $|\Delta(e^{j\omega})T(e^{j\omega})| = |\Delta(e^{j\omega})T_o(e^{j\omega})| < 1$, which is the robust stability condition for the baseline feedback loop. If $e^{-jm\omega}Q(e^{j\omega}) = 1$, then $T(e^{j\omega}) = 1$. At these frequencies we thus require $\Delta(e^{j\omega}) < 1$, i.e., the mismatch between $\tilde{P}(e^{j\omega})$ and $\hat{P}(e^{j\omega})$ has to be less than 100%.

4.4 Comparison

Listed in Table 4.1 are the three parameterization schemes discussed in this chapter. All of them provide the convenience that the closed-loop stability can be easily maintained, and that the sensitivity function can be decoupled to be an affine parameterization of the Q filter. The main differences among them are the complexity of the add-on loop shaping element,

[6]The uncertainty Δ is assumed to be stable and has finite magnitude response.

[7]Review of robust-stability analysis: consider a general (continuous- or discrete-time) feedback system with C, $\tilde{P} = P(1+\Delta)$, the vector from the $(-1,0)$ point to the frequency response $\tilde{P}(\omega)C(\omega)$ is given by
$$1 + \tilde{P}(\omega)C(\omega) = 1 + P(\omega)C(\omega) + P(\omega)C(\omega)\Delta(\omega).$$
Nominal stability gives that the distance $|1+P(\omega)C(\omega)|$ is always positive and $P(\omega)C(\omega)$ has the correct number of encirclements around the $(-1,0)$ point. Robust stability thus is guaranteed if $\left|1+\tilde{P}(\omega)C(\omega)\right|$ is never zero for any Δ at any frequency ω, which is achieved if and only if
$$\left|\frac{P(\omega)C(\omega)}{1+P(\omega)C(\omega)}\Delta(\omega)\right| < 1 \Leftrightarrow |T(\omega)\Delta(\omega)| < 1,\ \forall\omega.$$

and the suitable systems for implementation. The general YK parameterization in the third column does not require special assumptions on the plant, yet is most complicated in the add-on loop shaping element: $1 - NQ/Y$ depends on both the plant $P = N/D$ and the baseline controller $C = X/Y$. The simple YK algorithm in column two works for stable plants and the loop shaping element $1 - PQ$ uses the full information of P. The pseudo Youla-Kucera algorithm was formed aiming to further reduce the dependence of the plant model and improve the design intuition in the add-on loop shaping. These properties are traded with the requirement that $\hat{P}^{-1}$ must be stable (which can be designed to be so).

Table 4.1: Comparison of the three Youla-Kucera schemes

controller structure	simple YK parameterization for stable plant (Figure 4.2)	general YK parameterization (Theorem 4.1)	pseudo YK parameterization (Figure 4.3)
sensitivity function	$S = S_o \times (1 - PQ)$	$S = S_o \times \left(1 - \frac{N}{Y}Q\right)$	$S(z) =$ $S_o(z)\left(1 - z^{-m}Q(z)\right)$ when $\hat{P}(z) = P(z)$
add-on loop shaping element	$1 - PQ$	$1 - \frac{N}{Y}Q$	$1 - z^{-m}Q(z)$
nominal stability requirements	P and Q are stable	Q is stable	$\hat{P}^{-1}(z)$ and $Q(z)$ are stable
applications	continuous- and discrete-time	continuous- and discrete-time	discrete-time

From now on we will be focusing on the pseudo Youla-Kucera parameterization scheme, which we recall in Figure 4.4, with

$$S(z) = \frac{1 - z^{-m}Q(z)}{1 + P(z)C(z) + z^{-m}Q(z)\left(\hat{P}^{-1}(z)P(z) - 1\right)} \tag{4.34}$$

$$\begin{aligned} T(z) &= 1 - S(z) \\ &= \frac{P(z)C(z) + z^{-m}Q(z)\hat{P}^{-1}(z)P(z)}{1 + P(z)C(z) + z^{-m}Q(z)\left(\hat{P}^{-1}(z)P(z) - 1\right)}. \end{aligned} \tag{4.35}$$

Notice that $c(k)$ is the only signal added to the baseline closed loop. If $r(k) = 0$, after the signal processing, $c(k)$ should approximate $-d(k)$ for servo enhancement. We provide next the detailed time-domain design intuitions and the formulation of $Q(z)$ for local loop shaping.

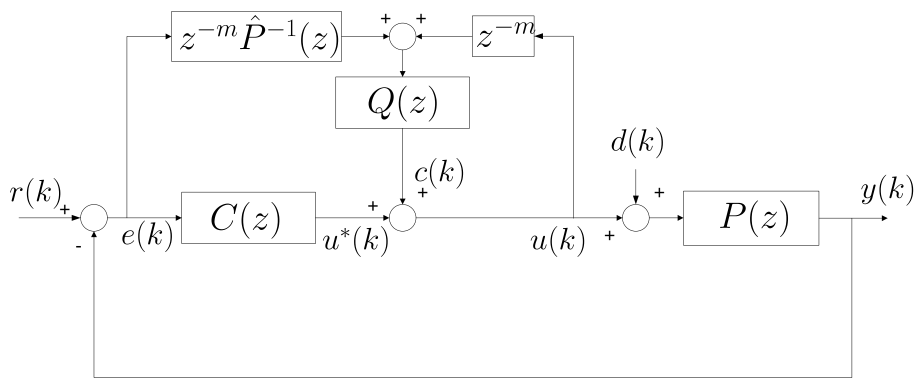


Figure 4.4: Pseudo Youla-Kucera parameterization for local loop shaping

4.5 Time-domain Disturbance-observer Intuition

For either regulation or tracking control, servo design aims at maintaining $e(k)$ small. Performing the block diagram transformations in Figures 4.5 and 4.6, we obtain a unified regulation problem in Figure 4.6, where $-e(k)$ can be regarded as a fictitious output that is regulated in the presence of the equivalent disturbances $d(k)$ and $r(k)$.

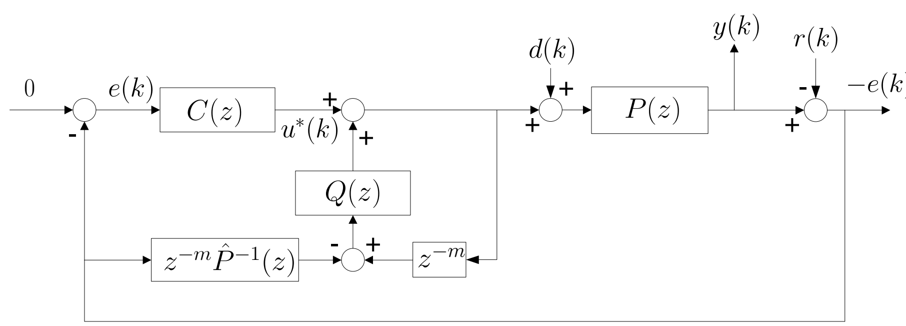


Figure 4.5: Equivalent form of Figure 4.4

In Figure 4.6, consider first the case where $r(k) = 0$ (regulation problem). Since $y(k) = P(q^{-1})(u(k) + d(k))$ (for time-domain operations, we start now to use the pulse transfer function with the delay operator q^{-1}), the output of $z^{-m}\hat{P}^{-1}(z)$ is given by

$$q^{-m}\hat{P}^{-1}(q^{-1})P(q^{-1})(u(k) + d(k)).$$

Notice that $q^{-m}\hat{P}^{-1}(q^{-1}) \approx q^{-m}P^{-1}(q^{-1})$. Through the inverse filtering, the output of $z^{-m}\hat{P}^{-1}(z)$ thus approximately equals $u(k-m) + d(k-m)$. Subtracting now $u(k-m)$, the output of the z^{-m} block, we get $u_Q(k)$, an approximation of $d(k-m)$. Analogously, for the

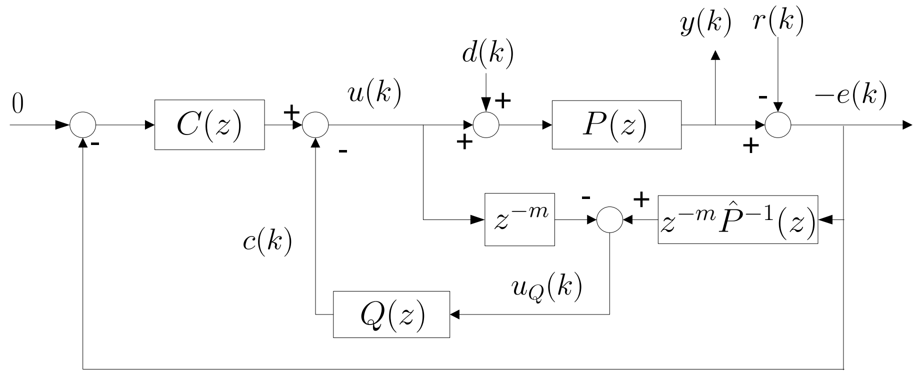


Figure 4.6: Equivalent form of Figure 4.5, from the viewpoint of a disturbance observer

reference-tracking problem, we have

$$\begin{aligned} u_Q(k) &= q^{-m}\hat{P}^{-1}\left(q^{-1}\right)\left(P\left(q^{-1}\right)u(k) - r(k)\right) - q^{-m}u(k) \\ &\approx q^{-m}\left[-\hat{P}^{-1}\left(q^{-1}\right)r(k)\right] \\ &\triangleq d_{\text{eq}}(k-m) \end{aligned}$$

where $d_{\text{eq}}(k) = -\hat{P}^{-1}(q^{-1})\,r(k)$ is an equivalent input disturbance for Figure 4.6.

$Q(z)$ can now be regarded as a signal-selection filter. If we design $Q(z)$ to have the magnitude response similar to that in Figure 4.7, then $c(k) = Q(q^{-1})\,u_Q(k)$ will selectively contain the signal components at the two peak frequencies. If in addition the phase of $Q(z)$

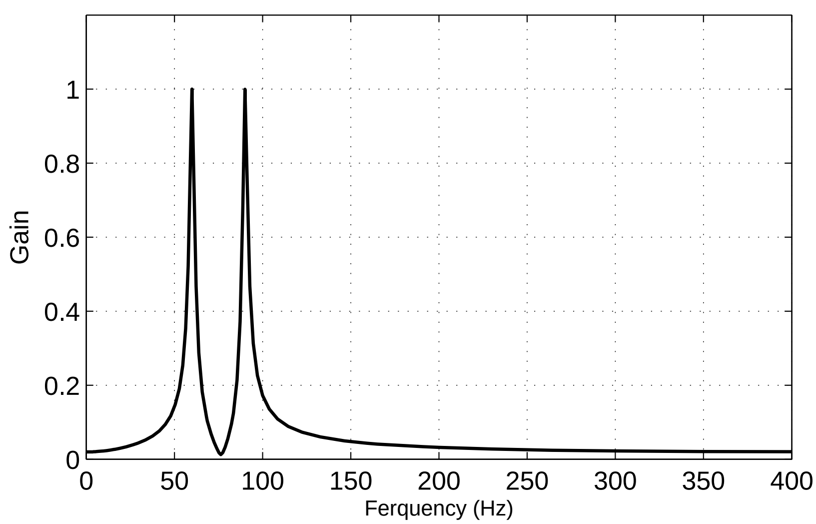


Figure 4.7: Magnitude response of a candidate Q filter

is correctly designed such that the m-step delay in the estimated $d(k-m)$ or $d_{\text{eq}}(k-m)$ is properly compensated, then $c(k)$ can recover the information of $d(k)$ or $d_{\mathrm{eq}}(k)$ at the peak frequencies.

We remark that the shape of $Q(z)$ in Figure 4.7 is central in the proposed design scheme. Uncertainties exist in $P(z)$ no matter how accurately $\hat{P}^{-1}(z)$ is constructed. It is not practical (and is even dangerous) to invert $P(z)$ over the entire frequency region. Keeping the magnitude of $Q(z)$ small except at the interested disturbance frequencies forms a "selective/local" disturbance observer, such that errors due to model mismatches do not pass through $Q(z)$ and get amplified by feedback.

The influence of plant delays: The phase of the delay z^{-m} equals $-m\omega$ (in rad/sec).A smaller value of m gives easier LLS design. Moreover:

- disturbance cancellation is easier at low frequencies than that at high frequencies. For instance, if $m = 1$ and the sampling time T_s equals 1/26400 sec, then at 100 Hz the one-step delay will introduce

$$m \times 100T_s \times 360 = 1.364 \text{ degrees}$$

  of phase mismatch between $d(k)$ and $d(k-m)$. However at 1500 Hz, this number increases to 20.45 degrees.

- the amount of phase mismatch increases linearly with respect to the number of delay steps. To have a quantitative idea of a large value of m, assuming the same sampling time of $T_s = 1/26400$ Hz, we can compute the period of a pure sinusoidal signal at 2640 Hz, which equals

$$\frac{\frac{1}{2640}}{T_s} = 10 \text{ time steps.}$$

  If 2640 Hz is the target disturbance-rejection frequency and m is three, then we have a mismatch of about one third of a period between $d(k)$ and $d(k-m)$.

The special case where $Q(z)$ is a low-pass filter: As disturbance cancellation is easier at low frequencies, a natural candidate for $Q(z)$ is a low-pass filter. Indeed, if the delay is not large then $d(k-m)$ will be close to $d(k)$, and the low-pass feature will avoid amplification of the high-frequency model mismatch. This concept is investigated in the structure of the disturbance observer (DOB) [26]. Next, we briefly analyze discrete-time DOB design, and connect it to the pseudo Youla-Kucera parameterization.

Under the same notations used before, a standard discrete-time DOB is as shown in Figure 4.8. The signals $d(k)$ and $n(k)$ are respectively the lumped input disturbance and the sensor noise. Standard block diagram analysis gives that

$$Y(z) = G_{yu^*}(z)\,U^*(z) + G_{yd}(z)\,D(z) + G_{yn}(z)\,N(z) \tag{4.36}$$

where

$$G_{yu^*}(z) = \frac{\hat{P}(z)\, P(z)}{\hat{P}(z) + z^{-m} Q(z) \left(P(z) - \hat{P}(z) \right)} \tag{4.37}$$

$$G_{yn}(z) = \frac{-z^{-m} P(z)\, Q(z)}{\hat{P}(z) + z^{-m} Q(z) \left(P(z) - \hat{P}(z) \right)} \tag{4.38}$$

$$G_{yd}(z) = \frac{P(z)\, \hat{P}(z) \left(1 - Q(z)\, z^{-m}\right)}{\hat{P}(z) + z^{-m} Q(z) \left(P(z) - \hat{P}(z) \right)} \tag{4.39}$$

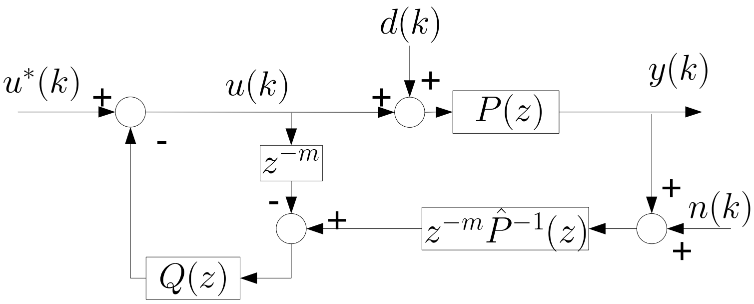


Figure 4.8: Block diagram of a standard digital disturbance observer

With a low-pass $Q(z)$, in the high-frequency region where $Q(e^{j\omega}) \approx 0$, DOB is essentially inactive, therefore $G_{yu^*}(e^{j\omega}) \approx P(e^{j\omega})$, $G_{yd}(z) \approx P(e^{j\omega})$, and $G_{yn}(e^{j\omega}) \approx 0$. In the low-frequency region where $Q(e^{j\omega}) \approx 1$, if the delay is small so that $(1 - e^{-jm\omega})\hat{P}(e^{j\omega}) \approx 0$, then $G_{yd}(e^{j\omega}) \approx 0$ in (4.39), $G_{yn}(e^{j\omega}) \approx -1$ in (4.38), and $G_{yu^*}(e^{j\omega}) \approx \hat{P}(e^{j\omega})$ in (4.37). The disturbance $d(k)$ and the model mismatch between $P(z)$ and $\hat{P}(z)$ are thus attenuated, and the entire DOB loop behaves like the nominal plant $\hat{P}(z)$. This nominal-model-following property makes it convenient to design other feedback or feedforward controllers based on the low-order nominal model $\hat{P}(z)$. For this reason, DOB-based feedback and feedforward servo is usually an "inside-out" design process, where the DOB and the nominal model are selected first in the inner loop, then the feedback controller $C(z)$ is designed in Figure 4.9. In contract, Youla-Kucera-parameterization schemes are "outside-in" approaches, where we build a baseline central controller $C(z)$ first, and then parameterize stabilizing controllers.

Yet when the reference is zero, i.e., in regulation problems, Figures 4.6 and 4.9 can be seen to be equivalent, which reveals that the pseudo Youla-Kucera parameterization scheme also share the nominal model following property for regulation control. Difference occurs in the treatment of reference tracking. If there is no model mismatch between $P(z)$ and $\hat{P}(z)$, we can see that the DOB inner loop in Figure 4.9 keeps the dynamics between $u^*(k)$ and $y(k)$ intact. Indeed, using (4.37), we can get the $r(k)$-to-$e(k)$ transfer function in Figure 4.9:

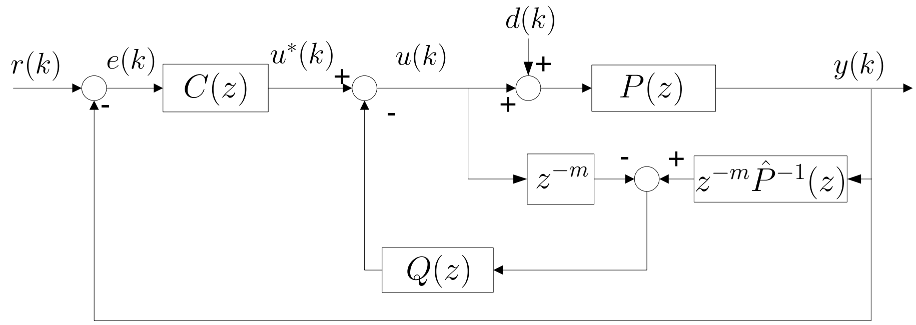


Figure 4.9: Block diagram of a closed-loop system with a disturbance observer

$$G_{er}(z) = \frac{1 + z^{-m}Q(z)\left(P(z)\hat{P}^{-1}(z) - 1\right)}{1 + P(z)C(z) + z^{-m}Q(z)\left(P(z)\hat{P}^{-1}(z) - 1\right)} \approx \frac{1}{1 + P(z)C(z)}$$

while in Figure 4.6

$$G_{er}(z) = \frac{1 - z^{-m}Q(z)}{1 + P(z)C(z) + z^{-m}Q(z)\left(P(z)\hat{P}^{-1}(z) - 1\right)}$$

and tracking performance is influenced by the term $1 - z^{-m}Q(z)$. We thus can still obtain better servo in pseudo YK parameterization.

Of course, as mentioned before, $Q(z)$, the heart of the pseudo Youla-Kucera parameterization, can be a general signal-selection filter other than a low-pass filter. This central design concept is explored for LLS next.

4.6 Overview of Q-filter Design

Recall that in regions where the frequency response $P(e^{j\omega})$ is well modeled by $\hat{P}(e^{j\omega})$, the sensitivity function [see (4.34)] satisfies the decoupled affine parameterization of $Q(z)$:

$$S(z) \approx \frac{1 - z^{-m}Q(z)}{1 + P(z)C(z)} \tag{4.40}$$

$$T(z) = 1 - S(z) \approx \frac{P(z)C(z) + z^{-m}Q(z)}{1 + P(z)C(z)} \tag{4.41}$$

If $e^{-mj\omega}Q(e^{j\omega}) = 1$, then we have $S(e^{j\omega}) = 0$ and $T(e^{j\omega}) = 1$, i.e., perfect disturbance rejection and reference tracking.

At the frequencies where there are large model mismatches, high-performance servo control intrinsically has to be sacrificed for robustness. We will thus make $e^{-mj\omega}Q(e^{j\omega}) \approx 0$, to keep the influence of the uncertainty elements

$$z^{-m}Q(z)\hat{P}^{-1}(z)P(z)$$

and

$$z^{-m}Q(z)[\hat{P}^{-1}(z)P(z)-1]$$

small in (4.34) and (4.35).

Several classes of design problems can be considered.

Low-frequency Servo Enhancement

This is the idea of a standard disturbance observer using a low-pass $Q(z)$. A direct result is that any bias disturbance will be rejected. This is because under the low-pass assumption we have $1-z^{-m}Q(z)|_{z=e^{j0}} = 1-e^{-jm\times 0}Q(e^{j\times 0}) = 0$ at the DC frequency $\omega = 0$. Therefore an integral action is built into the closed-loop controller.

Various researches have been conducted to design such low-pass Q filters [75, 76, 77]. Two main design options are the cut-off frequency (defining the bandwidth) and the high-frequency rolloff (influencing the robustness). A candidate design [78] is to select

$$Q(s) = \frac{3\tau s+1}{(\tau s+1)^3}$$

(τ defines the bandwidth of the filter) and then discretize $Q(s)$ using the bilinear transform

$$Q(z) = Q(s)\,|_{s=\frac{2}{T_s}\frac{1-z^{-1}}{1+z^{-1}}} = \frac{3\tau\frac{2}{T_s}\left(1+z^{-1}\right)^2\left(1-z^{-1}\right)+\left(1+z^{-1}\right)^3}{\left(\tau\frac{2}{T_s}\left(1-z^{-1}\right)+1+z^{-1}\right)^3}, \tag{4.42}$$

where T_s denotes the sampling time.

Figures 4.10 and 4.11 show a simulated example of the single-stage HDD benchmark problem discussed in Section 3.2. The corresponding loop shape design is as shown in Figure 1.7 on page 9. The strong low-frequency enhancement is readily seen from the error spectra. Indeed from Figure 1.7, the bandwidth has been extended to around 1000 Hz, with much stronger low-frequency error reduction. The detailed design process is provided in [17], which also discusses the practical modification of the plant to maximize the cut-off frequency of $Q(z)$.

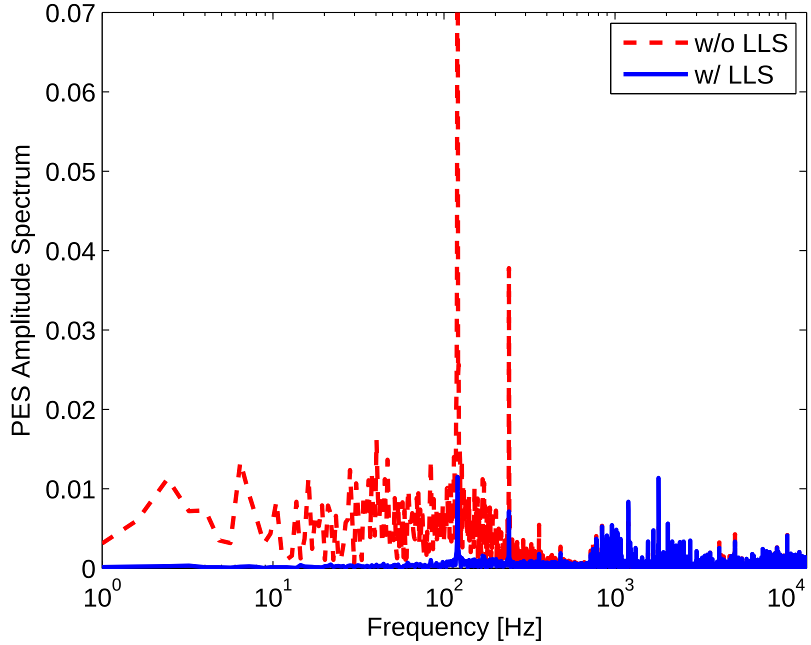


Figure 4.10: Spectra of the position errors under low-frequency LLS enhancement

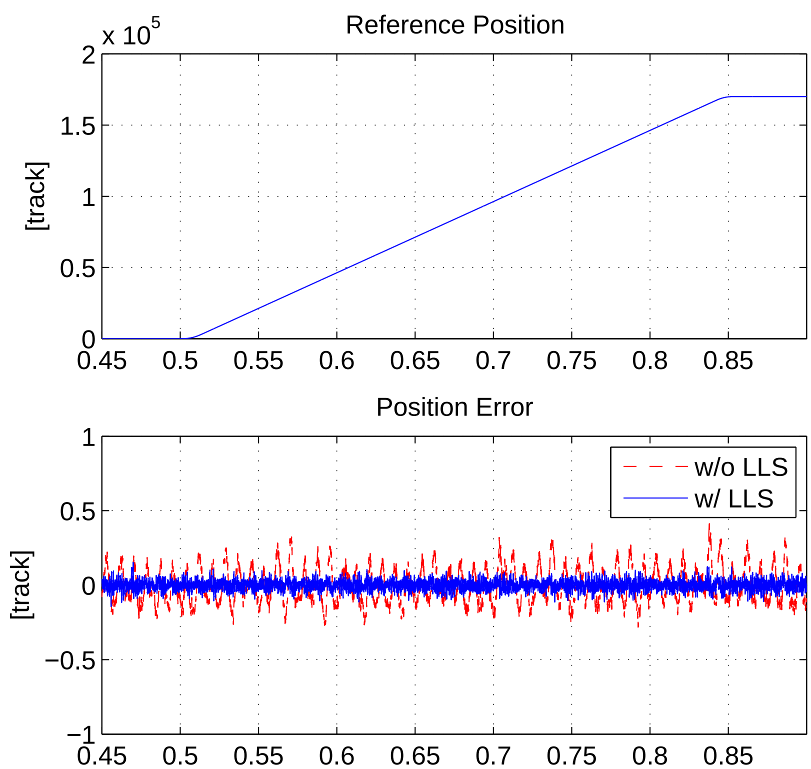


Figure 4.11: Time trace of the position errors under low-frequency LLS enhancement

Narrow-band Disturbance Rejection

Vibrations are frequency-dependent signals by nature. Since the closed-loop bandwidth can not be arbitrarily increased, vibrations at frequencies above the servo bandwidth are fundamentally difficult to handle. Actually, such band-limited disturbances, if strong enough, will significantly limit the servo performance even when their frequencies are below the bandwidth. A candidate design is to use a notch shape for $1 - z^{-m}Q(z)$ such as the one in the bottom plot of Figure 4.12. To demonstrate the flexibility, we introduce five notches in the magnitude of $1 - z^{-m}Q(z)$ here. The first three notches are very close to each other, the other two are separated at higher frequencies.

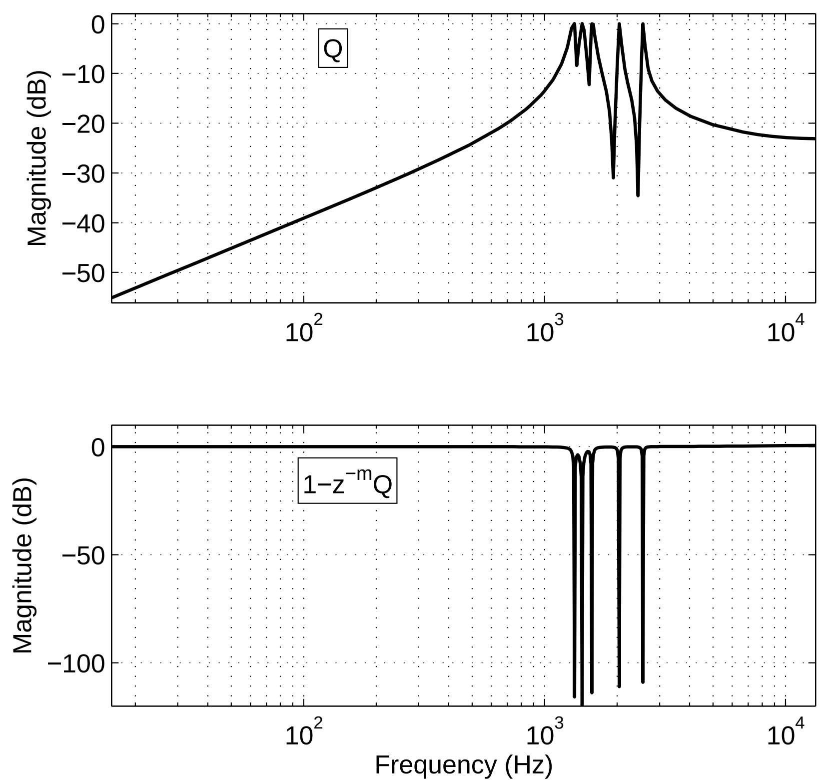


Figure 4.12: A Q-filter example for narrow-band disturbance rejection

Repetitive Control

This is for addressing the problem discussed in Figure 1.6 on page 8. An intuition about repetitive control is that, if the same type of disturbance occurs after a fixed period of time, i.e., $(1-q^{-N})d(k) = 0$ where N is the period of the disturbance, then at the next occurrence of $d(k)$ we can learn and reduce the resulting error, no matter how it behaves within one period of time. In the magnitude response of $1 - z^{-m}Q(z)$ in Figure 4.13, $|1 - e^{-jm\omega}Q(e^{j\omega})|$ has small gains at multiples of the fundamental frequency (120 Hz in this example). Meanwhile at other frequencies $|1 - e^{-jm\omega}Q(e^{j\omega})|$ is approximately unity, yielding no change to $|\tilde{S}(e^{j\omega})|$ in (4.40).

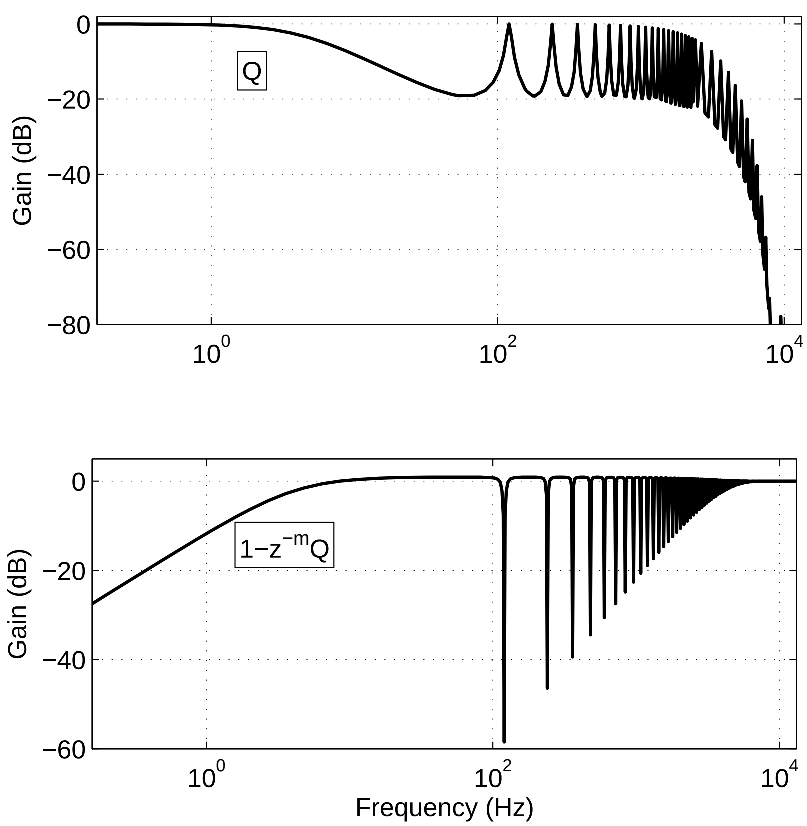


Figure 4.13: A Q-filter example for repetitive control

General Band-limited Vibration Compensation

The loop shaping in Section 4.6 was for narrow-band disturbance rejection. When excitation sources are rich in frequency, a wider attenuation notch shape is needed in $1 - z^{-m}Q(z)$. As we push further the disturbance attenuation, the trade-off amplification due to the waterbed effect will also increase (see an example in Figure 4.14). To achieve Figures 1.3 and 1.4 on page 7, new design considerations have to be made for controlling the waterbed.

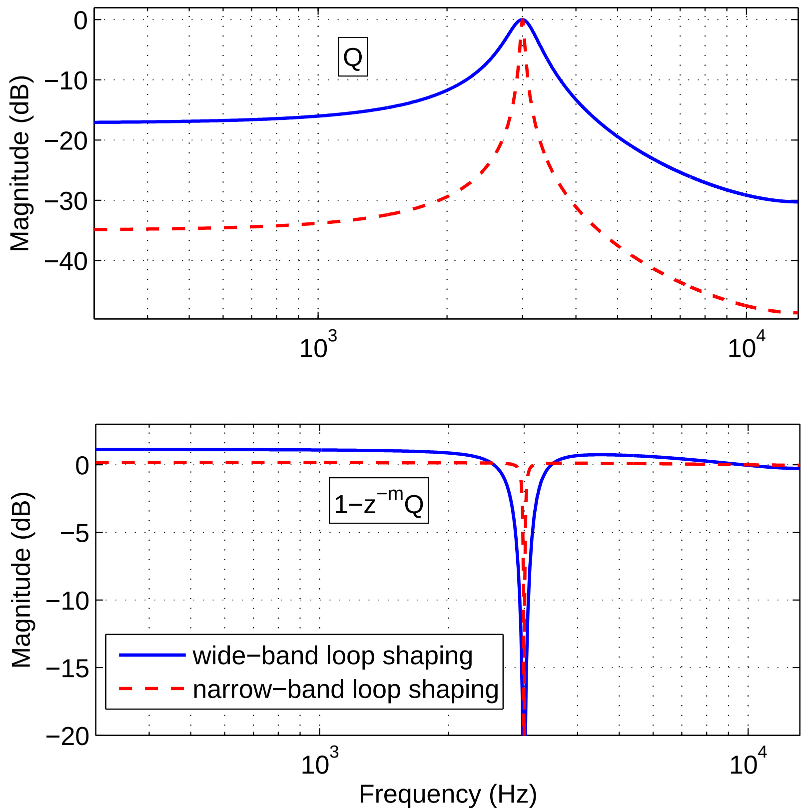


Figure 4.14: Q-filter design for general band-limited error reduction

In the next three chapters, we unfold the design of the aforementioned results, and provide application examples in each class of problems.

4.7 Notes and Additional Discussions

Applications of disturbance observer: As a flexible and powerful add-on element for servo enhancement, DOB has been applied in broad control applications, among which some examples are: hard disk drives [79, 28], optical disk drives [31], linear motors [30], positioning tables [27], robot arms [29], and automotive engines [80].

Pseudo Youla-Kucera parameterization and generalized disturbance observer: Although these two are equivalent in regulation problems, in [3, 4, 5, 6, 7, 10, 16] we have been mainly using using the disturbance-observer terminologies such as narrow-band DOB, and repetitive DOB, to refer to the results of Figures 4.12 and 4.13.

A tighter bound in robust-stability analysis: In Section 4.3, we have perturbed the plant with respect to $\hat{P}(z)$. Figure 4.15 shows all the operations that have been performed to the plant. $\hat{P}^{-1}(z)$ is the pseudo inverse model we proposed when $P^{-1}(z)$ is unstable itself. The difference between these two form the first level of model mismatch. The real plant uncertainty (the second level of model mismatch) is actually placed on $P(z)$. If we consider these two levels of mismatch between $\hat{P}(z)$ and the actual plant, we can obtain a tighter bound of robust stability.

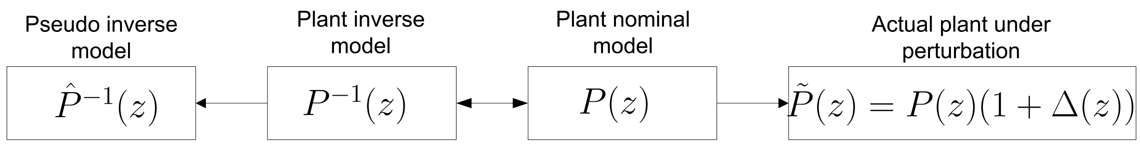


Figure 4.15: Perturbations to the plant in pseudo YK parameterization

- Nominal stability: the closed-loop system is stable when $\tilde{P}(z) = P(z)$, namely, all the roots of (4.27) on page 51 must be inside the unit circle;
- Robust stability: when the plant is perturbed to be $\tilde{P}(z) = P(z)(1+\Delta(z))$ (assume that the uncertainty $\Delta(z)$ is stable and bounded), the following inequality has to hold (recall the footnote 7 on page 53)

$$\left|\Delta(e^{j\omega})T(e^{j\omega})\right| < 1 \tag{4.43}$$

where the nominal complementary sensitivity function $T = P\tilde{C}/(1+P\tilde{C})$ is given by, after substituting in (4.21) and simplification,

$$T(z) = \frac{P(z)C(z) + z^{-m}\hat{P}(z)P(z)Q(z)}{1 + P(z)C(z) + Q(z)\left(\hat{P}^{-1}(z)P(z) - z^{-m}\right)}.$$

Chapter 5

Internal Model Based IIR Q Design for Narrow-band Loop Shaping

In this chapter we discuss the Q filters for narrow-band loop shaping mentioned in Section 4.6. We show that for structured vibrations with clear internal models, there exist configurations that achieve optimal disturbance rejection.

Recalling Figure 4.4 on page 55 and Equation (4.34), we have

$$y(k) = P\left(q^{-1}\right) S\left(q^{-1}\right) d(k) + T\left(q^{-1}\right) r(k)$$

where

$$S\left(q^{-1}\right) = \frac{1 - q^{-m} Q\left(q^{-1}\right)}{1 + P\left(q^{-1}\right) C\left(q^{-1}\right) + q^{-m} Q\left(q^{-1}\right)\left[\hat{P}^{-1}\left(q^{-1}\right) P\left(q^{-1}\right) - 1\right]}$$

$$T\left(q^{-1}\right) = \frac{P\left(q^{-1}\right) C\left(q^{-1}\right) + z^{-m} Q\left(q^{-1}\right) \hat{P}^{-1}\left(q^{-1}\right) P\left(q^{-1}\right)}{1 + P\left(q^{-1}\right) C\left(q^{-1}\right) + q^{-m} Q\left(q^{-1}\right)\left[\hat{P}^{-1}\left(q^{-1}\right) P\left(q^{-1}\right) - 1\right]}.$$

5.1 From FIR to IIR Design

Suppose $d(k) = w(k) + n(k)$ with $w(k)$ and $n(k)$ being respectively the structured disturbance and the additive noise. To regulate $P\left(q^{-1}\right) S\left(q^{-1}\right) w(k)$ to zero, it suffices to design $Q(q^{-1})$ such that

$$\left(1 - q^{-m} Q(q^{-1})\right) w(k) \to 0. \tag{5.1}$$

For instance, if $m = 1$ and $w(k)$ is a sinusoidal signal with the internal model

$$\left(1 - 2\cos\omega_1 q^{-1} + q^{-2}\right) C_1 \sin\left(\omega_1 k + \psi_1\right) = 0,\ \forall k \geq 2 \tag{5.2}$$

then solving

$$1 - z^{-1} Q(z) = 1 - 2\cos\omega_1 z^{-1} + z^{-2} \tag{5.3}$$

yields

$$Q(z) = 2\cos\omega_1 - z^{-1}. \tag{5.4}$$

This is the simplest internal-model-based Q design. Variations of this FIR-filter design have been popular in literature for periodic-disturbance rejection. Recall that we need the magnitude of the Q filter to be small at the non-interested frequencies for stability and robustness. The FIR filter in (5.4) satisfies

$$\begin{aligned} Q\left(e^{j\omega}\right)\big|_{\omega=0} &= 2\cos\omega_1 - 1 \\ Q\left(e^{j\omega}\right)\big|_{\omega=\pi} &= 2\cos\omega_1 - (-1). \end{aligned}$$

Regardless of the value of ω_1, we have

$$|Q(1) - Q(-1)| = 2$$

namely, it is not possible to keep the gains at DC and Nyquist frequency to be small at the same time. Typically $\cos\omega_1$ is positive and $Q(-1)$ is thus larger than one,[1] which indicates that the Q design of (5.4) is sensitive to high-frequency noises and plant uncertainties. To address such issues, in this chapter we will be using IIR Q filters in the structure of

$$Q\left(q^{-1}\right) = \frac{B_Q\left(q^{-1}\right)}{A_Q\left(q^{-1}\right)} \tag{5.5}$$

for enhanced LLS.

5.2 The Internal Models

Consider the multiple narrow-band signal in the general form of

$$w(k) = \sum_{i=1}^{n} C_i \sin(\omega_i k + \psi_i), \tag{5.6}$$

where the frequency $\omega_i = 2\pi\Omega_i T_s$ is in rad/sec (Ω_i is the frequency in Hz, T_s is the sampling time); $C_i\,(\neq 0)$ and ψ_i are respectively the unknown magnitude and phase of each sinusoidal component. For a well-defined problem, we assume that $k \geq 0$ and $\omega_i \neq \omega_j$, $\forall\, i \neq j$.

Extending (5.2) to the case of multiple frequency components, we have

$$\prod_{i=1}^{n} \left(1 - 2\cos(\omega_i)\, q^{-1} + q^{-2}\right) w(k) = 0 \;\forall k \geq 2n. \tag{5.7}$$

[1] Consider a practical example where $\Omega_1 = 2000$ Hz, $T_s = 1/26400$ sec, and $\omega_1 = 2\pi\Omega_1 T_s$, then

$$\begin{aligned} Q(1) &= 0.7777 \\ Q(-1) &= 2.7777 \end{aligned}$$

The cascade form of the internal model is nonlinear in the ω_i's. Introduce new variables a_i's, such that

$$A\left(q^{-1}\right) = \prod_{i=1}^{n}\left(1-2\cos\left(\omega_i\right)q^{-1}+q^{-2}\right) \tag{5.8}$$

$$= 1 + a_1 q^{-1} + \cdots + a_n q^{-n} + \cdots + a_1 q^{-2n+1} + q^{-2n} \tag{5.9}$$

$$= 1 + \sum_{i=1}^{n-1} a_i\left(q^{-i}+q^{-2n+i}\right) + a_n q^{-n} + q^{-2n},$$

where in the second equality we have used the fact that the coefficients of 1 and q^{-2} are the same in $1-2\cos\left(\omega_i\right)q^{-1}+q^{-2}$, resulting in the symmetric coefficient vector

$$\{1, a_1, \ldots, a_n, \ldots, a_1, 1\}.$$

Equations (5.8) and (5.9) are respectively the (nonlinear) cascaded and the (linear) direct forms of the polynomial $A\left(q^{-1}\right)$.

5.3 Multi-Q Approach

Analogous to Section 5.1, to use an IIR $Q(q^{-1}) = B_Q(q^{-1})/A_Q(q^{-1})$, the right hand side of (5.3) must have both numerators and denominators. Recalling (5.7) and (5.9), we see that

$$\left[1-q^{-m}Q\left(q^{-1}\right)\right]w\left(k\right) \to 0$$

is achieved if

$$\boxed{1-q^{-m}Q(q^{-1}) = K(q^{-1})\frac{A(q^{-1})}{A_Q(q^{-1})}.} \tag{5.10}$$

The filter $K(q^{-1})$ is necessary to make the solution causal for a general m. Without $K(q^{-1})$, (5.10) gives

$$Q(q^{-1}) = q^{m}\frac{A_Q(q^{-1})-A(q^{-1})}{A_Q(q^{-1})}$$

where the forward shift operation q^m is not directly realizable.

To get the magnitude response of $1-q^{-m}Q(q^{-1})$ in Figure 4.12 on page 62, $A(q^{-1})/A_Q(q^{-1})$ should have a notch-filter structure. A natural choice is to damp the roots of $A(q^{-1})$ by a scalar $\alpha \in (0,1)$ and let

$$A_Q(q^{-1}) \triangleq \prod_{i=1}^{n}\left(1-2\alpha\cos(\omega_i)q^{-1}+\alpha^2 q^{-2}\right) \tag{5.11}$$

or, in the direct form,

$$\boxed{A_Q\left(q^{-1}\right) = 1 + a_1\alpha q^{-1} + \cdots + a_n\alpha^n q^{-n} + \cdots + a_1\alpha^{2n-1}q^{-2n+1} + \alpha^{2n}q^{-2n}} \tag{5.12}$$

namely, $A_Q\left(q^{-1}\right)$ is nothing but $A(\alpha q^{-1})$, by replacing each q^{-1} in $A\left(q^{-1}\right)$ [see (5.13)] with αq^{-1}:

$$\boxed{A(q^{-1}) = 1 + a_1 q^{-1} + \cdots + a_n q^{-n} + \cdots + a_1 q^{-2n+1} + q^{-2n}} \tag{5.13}$$

The parameter α determines the width of the notch shape. An α that is closer to one gives a sharper notch. Approximately, the width of the notch at -3dB is given by $(1-\alpha^2)/[(\alpha^2+1)(\pi T_s)]$.[2]

A design guide for $K(q^{-1})$ is that it should not introduce much magnitude distortion to the achieved notch shape of $A(q^{-1})/A_Q(q^{-1})$ in (5.10). We discuss next choices of $K(q^{-1})$ for different values of m.

The Case for $m = 0$

For the simplest case where $m = 0$, a scalar value $K(q^{-1}) = k$ provides a realizable solution to (5.10). Recall $Q(q^{-1}) = B_Q(q^{-1})/A_Q(q^{-1}) = B_Q(q^{-1})/A(\alpha q^{-1})$ and (5.13). Equation (5.10) now reduces to

$$A(\alpha q^{-1}) - B_Q(q^{-1}) = kA(q^{-1}),$$

yielding

$$\begin{aligned} B_Q(q^{-1}) = (1-k) + (\alpha - k)\, a_1 q^{-1} + \cdots + (\alpha^n - k)\, a_n q^{-n} + \\ \cdots + (\alpha^{2n-1} - k) a_1 q^{-2n+1} + (\alpha^{2n} - k) q^{-2n}. \end{aligned} \tag{5.14}$$

The parameter k provides an additional freedom to shape the magnitude response of the Q filter. It can be shown (see Appendix A.3) that $k = \alpha^n$ leads to the common factor $1 - \alpha q^{-2}$ in $B_Q(q^{-1})$, which places two zeros to the Q filter at $\pm\sqrt{\alpha}$. This provides balanced magnitude response at low- and high-frequencies, and is the recommended choice for general applications. We thus have

$$\boxed{B_Q(q^{-1}) = \sum_{i=0}^{2n} (\alpha^i - \alpha^n) a_i q^{-i};\ a_i = a_{2n-i},\ a_{2n} = 1.}$$

As an example, when $n = 1$, we have $A_Q\left(q^{-1}\right) = 1 + \alpha a_1 q^{-1} + \alpha^2 q^{-2} = 1 - \alpha 2\cos\left(\omega_1\right) q^{-1} + \alpha^2 q^{-2}$ and

$$Q_{m=0}(q^{-1}) = \frac{(1-\alpha)\left(1 - \alpha q^{-2}\right)}{1 - \alpha \cdot 2\cos(\omega_1) q^{-1} + \alpha^2 q^{-2}}. \tag{5.15}$$

The Case for $m = 1$

Applying analogous analysis as in Section 5.3, we reduce (5.10) to

$$A(\alpha q^{-1}) - q^{-1} B_Q(q^{-1}) = kA(q^{-1}), \tag{5.16}$$

[2] This is an empirical approximation from [81].

the solution of which is

$$B_Q(q^{-1}) = (1-k)\,q + (\alpha - k)\,a_1 + \cdots + (\alpha^n - k)a_n q^{-n+1} + \\ \cdots + (\alpha^{2n-1} - k)a_1 q^{-2n+2} + (\alpha^{2n} - k)q^{-2n+1}.$$

To let the term $(1-k)\,q$ vanish for realizability, we require $k = 1$, which gives

$$\boxed{B_Q(q^{-1}) = \sum_{i=1}^{2n}(\alpha^i - 1)a_i q^{-i+1};\ a_i = a_{2n-i},\ a_{2n} = 1.} \tag{5.17}$$

In this case, if $n = 1$, (5.11) and (5.17) yield $a_1 = -2\cos\omega_1$ and

$$Q_{m=1}(q^{-1}) = \frac{(1-\alpha)\,(2\cos\omega_1 - (1+\alpha)\,q^{-1})}{1 - 2\alpha\cos\omega_1 q^{-1} + \alpha^2 q^{-2}}. \tag{5.18}$$

Evaluating the frequency response at ω_1 and using the identity $2\cos(\omega_1) = e^{j\omega_1} + e^{-j\omega_1}$, we obtain

$$Q_{m=1}\left(e^{j\omega_1}\right) = \frac{-\,(\alpha-1)\,(e^{j\omega_1} + e^{-j\omega_1}) + (\alpha^2 - 1)\,e^{-j\omega_1}}{1 - \alpha\cdot(e^{j\omega_1} + e^{-j\omega_1})\,e^{-j\omega_1} + \alpha^2 e^{-2j\omega_1}} = e^{j\omega_1}. \tag{5.19}$$

Thus, at the center frequency ω_1, the Q filter provides exactly one-step advance to counteract the one-step delay in $1 - q^{-1}Q_{m=1}(q^{-1})$, yielding perfect disturbance cancellation at ω_1.

The Case for an Arbitrary m

For $m > 1$, assigning $K(q^{-1}) = k$ no longer gives a realizable solution. Consider the following IIR design

$$K(q^{-1}) = \sum_{i=0}^{N} k_i \left[\frac{A(q^{-1})}{A(\alpha q^{-1})}\right]^i,\ k_i \in \mathbb{R}. \tag{5.20}$$

Namely, $K(q^{-1})$ is chosen as a linear combination of N filters that influence only the local loop shape (recall that $A(q^{-1})/A(\alpha q^{-1})$ is a notch filter). Take the example of $m = 2$. When $N = 1$,[3] solving (5.10) gives

$$Q(q^{-1}) = q^2\left(1 - k_0\frac{A(q^{-1})}{A(\alpha q^{-1})} - k_1\frac{A(q^{-1})^2}{A(\alpha q^{-1})^2}\right).$$

Partitioning, we obtain

$$Q(q^{-1}) \triangleq q^2\left(1 - \rho_1\frac{A(q^{-1})}{A(\alpha q^{-1})}\right)\left(1 - \rho_2\frac{A(q^{-1})}{A(\alpha q^{-1})}\right). \tag{5.21}$$

[3] $N = 0$ does not give a solvable solution since in that case $K(q^{-1})$ is a scalar again.

The numerator of $Q(q^{-1})$ is given by

$$B_Q(q^{-1}) = q^2(A(\alpha q^{-1}) - \rho_1 A(q^{-1}))(A(\alpha q^{-1}) - \rho_2 A(q^{-1})).$$

From (5.13) and (5.12), we have

$$A(\alpha q^{-1}) - \rho_i A(q^{-1}) = (1-\rho_i) + (\alpha - \rho_i)\, a_1 q^{-1} + \cdots + (\alpha^{2n} - \rho_i) q^{-2n}.$$

To make the q^2 term vanish in $B_Q(q^{-1})$, we must have $1 - \rho_i = 0$ for $i = 1, 2$, yielding $k_0 = 2$, $k_1 = -1$. Therefore, after simplification,

$$Q(q^{-1}) = \left[\frac{\sum_{i=1}^{2n}(\alpha^i - 1)a_i q^{-i+1}}{A(\alpha q^{-1})}\right]^2 ;\ a_i = a_{2n-i},\ a_{2n} = 1 \tag{5.22}$$

$$K(q^{-1}) = 2 - \frac{A(q^{-1})}{A(\alpha q^{-1})}. \tag{5.23}$$

For a general integer m, when $N = m - 1$, applying analogous analysis, we get the following partitioned $Q(q^{-1})$ from (5.20) and (5.10):

$$Q(q^{-1}) = q^m \prod_{i=1}^{m} \left(1 - \rho_i \frac{A(q^{-1})}{A(\alpha q^{-1})}\right).$$

The solution pair is thus obtained when $\rho_i = 1\ \forall i$, and

$$Q(q^{-1}) = \left[\frac{\sum_{i=1}^{2n}(\alpha^i - 1)a_i q^{-i+1}}{A(\alpha q^{-1})}\right]^m \tag{5.24}$$

$$1 - q^{-m}Q(q^{-1}) = 1 - \left(1 - \frac{A(q^{-1})}{A(\alpha q^{-1})}\right)^m \tag{5.25}$$

$$= \frac{A(q^{-1})}{A(\alpha q^{-1})} \sum_{i=1}^{m} \left(\begin{array}{c} m \\ i \end{array}\right) \left[-\frac{A(q^{-1})}{A(\alpha q^{-1})}\right]^{i-1}. \tag{5.26}$$

Here a_i is as defined in (5.22); and from (5.25) to (5.26) we have used the identity

$$(1+x)^m = 1 + \left(\begin{array}{c} m \\ 1 \end{array}\right) x + \cdots + \left(\begin{array}{c} m \\ m \end{array}\right) x^m$$

where $\left(\begin{array}{c} m \\ i \end{array}\right) = \frac{m!}{i!(m-i)!}$ is the binomial coefficient.

It can be observed that the general result obtained here is essentially a cascaded version of the developed Q filter $Q_{m=1}(q^{-1})$. Recall that $Q_{m=1}(q^{-1})$ provides one-step phase advance at the disturbance frequencies $\{\omega_i\}_1^n$, to address the term q^{-1} in $1 - q^{-1}Q_{m=1}(q^{-1})$. For a general m, we see that the cascaded

$$\boxed{Q(q^{-1}) = \left[Q_{m=1}(q^{-1})\right]^m} \tag{5.27}$$

in (5.24) works the same way, as

$$1 - q^{-m} Q(q^{-1}) = 1 - \left[q^{-1} Q_{m=1}(q^{-1}) \right]^m$$

i.e., each $Q_{m=1}(q^{-1})$ block compensates one q^{-1} term, to achieve $1 - e^{-jm\omega} Q(e^{j\omega}) = 0$ when $\omega \in \{\omega_i\}_1^n$. Recall that $Q(q^{-1})$ is a special type of bandpass filter. Cascading multiple $Q_{m=1}(q^{-1})$ together not only provides the compensation for q^{-m}, but also offers an enhanced bandpass frequency response, as $|Q(e^{j\omega})|^m \leq |Q(e^{j\omega})|$ if $m \geq 1$ and $|Q(e^{j\omega})| < 1$ (when ω is outside of the passband).

Remark 5.1. The compensation difficulty gets larger as the value of m increases. This is intuitively quite reasonable, as a larger m corresponds to more plant delays in LLS.

5.4 Direct Approach

In the previous section, we assigned an IIR structure for $K(q^{-1})$ in

$$1 - q^{-m} Q\left(q^{-1}\right) = K\left(q^{-1}\right) \frac{A\left(q^{-1}\right)}{A\left(\alpha q^{-1}\right)}. \tag{5.28}$$

The order of $Q(q^{-1})$ is the sum of the orders of $A(\alpha q^{-1})$ and of $K(q^{-1})$. A lower-order solution can be obtained if we relax $K(q^{-1})$ to be an FIR filter

$$\boxed{K\left(q^{-1}\right) = k_o + k_1 q^{-1} + \cdots + k_{n_K} q^{-n_K}.} \tag{5.29}$$

Letting $Q(q^{-1}) = B_Q(q^{-1}) / A(\alpha q^{-1})$ and reordering (5.28) yield

$$\boxed{K\left(q^{-1}\right) A\left(q^{-1}\right) + q^{-m} B_Q\left(q^{-1}\right) = A\left(\alpha q^{-1}\right).} \tag{5.30}$$

Matching the coefficients of q^{-i} on each side of (5.30), we can obtain $K(q^{-1})$ and $B_Q(q^{-1})$.

More general, a polynomial equation $X(q^{-1}) A(q^{-1}) + Y(q^{-1}) B(q^{-1}) = F(q^{-1})$, with $X(q^{-1})$ and $Y(q^{-1})$ as the unknowns, is a Diophantine equation (Appendix B). Solutions exist as long as the greatest common factor of $A(q^{-1})$ and $B(q^{-1})$ divides $F(q^{-1})$.

Since q^{-m} and $A(q^{-1})$ are coprime (their greatest common factor is 1), (5.30) can be solved if

$$\begin{aligned} \deg\left(A\left(\alpha q^{-1}\right)\right) &\leq \deg\left(B_Q\left(q^{-1}\right)\right) + m \\ \deg\left(B_Q\left(q^{-1}\right)\right) + m &= \deg\left(K\left(q^{-1}\right)\right) + \deg\left(A\left(q^{-1}\right)\right). \end{aligned}$$

As $\deg(A(\alpha q^{-1})) = \deg(A(q^{-1})) = 2n$, the minimum-order solution is obtained when

$$\begin{aligned} \deg\left(K\left(q^{-1}\right)\right) &= m - 1 \\ \deg\left(B_Q\left(q^{-1}\right)\right) &= 2n - 1. \end{aligned} \tag{5.31}$$

Example 5.1. Let $m = 2$, $n = 1$, and

$$A\left(\alpha q^{-1}\right) = 1 + \alpha a q^{-1} + \alpha^2 q^{-2}.$$

The Diophantine equation is thus

$$K\left(q^{-1}\right)\left(1 + a q^{-1} + q^{-2}\right) + q^{-2} B_Q\left(q^{-1}\right) = A\left(\alpha q^{-1}\right). \tag{5.32}$$

Letting

$$K\left(q^{-1}\right) = k_0 + k_1 q^{-1}$$

we get

$$q^{-2} B_Q\left(q^{-1}\right) = 1 - k_0 + \left(\alpha a - a k_0 - k_1\right) q^{-1} + \left(\alpha^2 - k_0 - a k_1\right) q^{-2} - k_1 q^{-3}.$$

To have a realizable $B_Q\left(q^{-1}\right)$, we need

$$1 - k_0 = 0, \;\; \alpha a - a k_0 - k_1 = 0,$$

yielding

$$B_Q\left(q^{-1}\right) = \left(\alpha^2 - k_0 - a k_1\right) - k_1 q^{-1} = \left(\alpha^2 - 1 - a^2\left(\alpha - 1\right)\right) - \left(\alpha - 1\right) a q^{-1}.$$

The final Q filter is thus

$$Q\left(q^{-1}\right) = \frac{\left(\alpha^2 - 1 - a^2\left(\alpha - 1\right)\right) - \left(\alpha - 1\right) a q^{-1}}{1 + \alpha a q^{-1} + \alpha^2 q^{-2}}. \tag{5.33}$$

For the case where $m = 1$, from (5.31), we know that $K\left(q^{-1}\right) = k_0$ is sufficient for the minimum-order solution. The Diophantine equation (5.30) then reduces to (5.16). We have the following detailed Q-filter transfer functions:

- $m = 1$ and $n = 1$:

$$Q\left(q^{-1}\right) = \frac{\left(\alpha - 1\right)\left(a + \left(1 + \alpha\right) q^{-1}\right)}{1 + \alpha a q^{-1} + \alpha^2 q^{-2}} \tag{5.34}$$

- $m = 1$ and $n = 2$:

$$Q\left(q^{-1}\right) = \frac{a_1\left[\left(\alpha - 1\right) + \left(\alpha^3 - 1\right) q^{-2}\right] + \left(\alpha^2 - 1\right)\left[a_2 q^{-1} + \left(\alpha^2 + 1\right) q^{-3}\right]}{1 + a_1 \alpha q^{-1} + a_2 \alpha^2 q^{-2} + a_1 \alpha^3 q^{-3} + \alpha^4 q^{-4}} \tag{5.35}$$

- $m = 1$ and $n = 3$:

$$Q\left(q^{-1}\right) = \frac{B_Q\left(q^{-1}\right)}{A_Q\left(q^{-1}\right)} \tag{5.36}$$

  where

$$A_Q\left(q^{-1}\right) = 1 + a_1 \alpha q^{-1} + a_2 \alpha^2 q^{-2} + a_3 \alpha^3 q^{-3} + a_2 \alpha^4 q^{-4} + a_1 \alpha^5 q^{-5} + \alpha^6 q^{-6}$$

  and

$$\begin{aligned} B_Q\left(q^{-1}\right) = {} & a_1\left[\left(\alpha - 1\right) + \left(\alpha^5 - 1\right) q^{-4}\right] + \\ & a_2\left(\alpha^2 - 1\right)\left[q^{-1} + \left(\alpha^2 + 1\right) q^{-3}\right] + \left(\alpha^3 - 1\right)\left[a_3 q^{-2} + \left(\alpha^3 + 1\right) q^{-5}\right] \end{aligned}$$

Remark 5.2. For implementation of the above filters, we need values of $\{a_i\}_{i=1}^n$, which are related to the frequencies $\{\omega_i = 2\pi T_s \Omega_i\}_{i=1}^n$ (Ω_i is the center frequency in Hz) by

$$\prod_{i=1}^{n} \left(1 - 2\cos(\omega_i) q^{-1} + q^{-2}\right) = 1 + a_1 q^{-1} + \cdots + a_n q^{-n} + \cdots + a_1 q^{-2n+1} + q^{-2n}.$$

Remark 5.3. Small-scale problems can be readily solved by following Example 5.1. For larger-scale equations, the matrix-equation approach for solving general Diophantine equations (Appendix B) can be used. More details will be provided in the next Chapter.

5.5 Case Study and Comparison

We provide a case study and compare the two approaches (multi-Q and direct-Q) about Q-filter design.

The following parameters are assumed:

- sampling time: $T_s = 1/26400$ sec
- plant delay: $m = 2$
- number of narrow bands: $n = 1$
- target frequency: 900 Hz

From Example 5.1 on the preceding page, the Q filter in the direct approach is given by (5.33). Letting $\alpha = 0.9882$, we obtain $Q(z)$ with its frequency response shown in the dashed line in Figure 5.1. Correspondingly in Figure 5.2, the dashed line provides the magnitude response of $1 - z^{-m} Q(z)$. We observe the strong attenuation at 900 Hz and the width of the stop band is about 100 Hz.

For the multi-Q approach, we use (5.27) to get $Q(q^{-1}) = [Q_{m=1}(q^{-1})]^m$ where $Q_{m=1}(q^{-1})$ is from (5.18). Letting $\alpha = 0.9765$ yields the solid lines in Figures 5.1 and 5.2.

Both approaches give the desired shape of $1 - z^{-m} Q(z)$ in Figure 5.2. In the phase response in Figure 5.1, we can see that both filters have non-zero phases at 900 Hz—the center frequency of the pass band. This is for compensating the delay effect z^{-m}. The multi-Q approach provides smaller magnitude response outside the pass band of $Q(z)$, which implies stronger filtering of noises and model mismatches by $Q(z)$.

Associated with the sharper magnitude response, the solid line in the phase response changes more aggressively in Figure 5.1. The root cause of this phenomenon is Bode's Phase Formula (see Section 2.4), which tells that a fast-changing phase response always comes with a rapid magnitude response, if the system is minimum-phase.

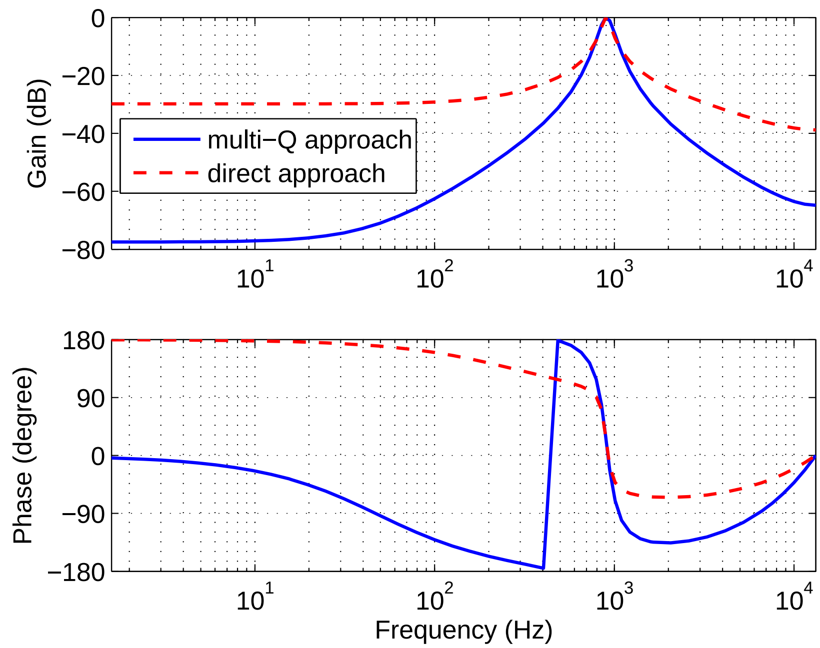


Figure 5.1: Internal-model based Q-design example: $Q(z)$

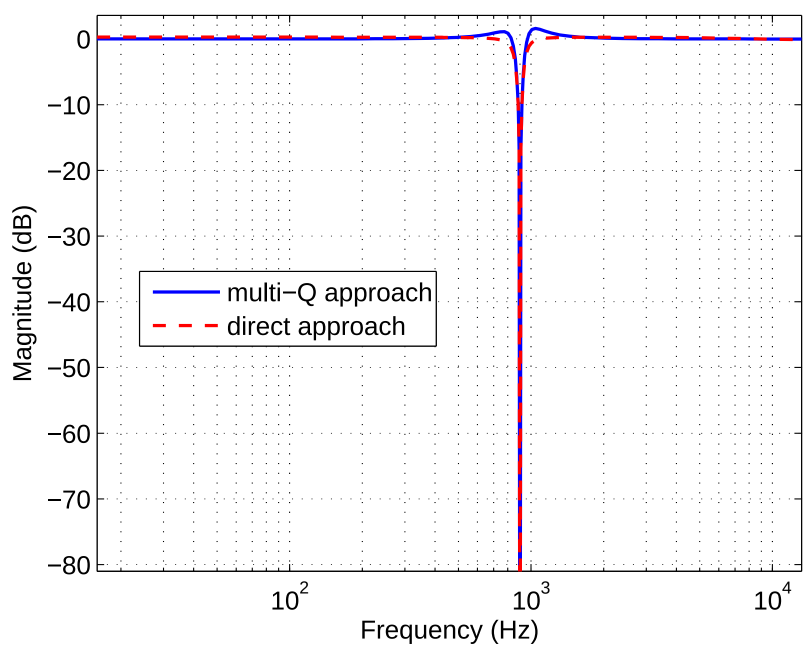


Figure 5.2: Internal-model based Q-design example: $1 - z^{-m}Q(z)$

To see more details, note first that any filter can be decomposed to the product of a minimum-phase filter and an all-pass filter. In the multi-Q approach, we have

$$\begin{aligned} Q(z) &= \left[\frac{(1-\alpha)\left(-a-(1+\alpha)\,z^{-1}\right)}{1+a\alpha z^{-1}+\alpha^2 z^{-2}}\right]^2 \\ &= \left[\frac{(\alpha-1)\left(a+(1+\alpha)\,z^{-1}\right)}{1+a\alpha z^{-1}+\alpha^2 z^{-2}}\right]^2. \end{aligned}$$

The zero at $-(1+\alpha)/a\,(=1.10173)$ is outside the unit circle. Factorization yields

$$\begin{aligned} Q(z) &= \left[\frac{(\alpha-1)\,a\left(1+\frac{(1+\alpha)}{a}z\right)}{1+a\alpha z^{-1}+\alpha^2 z^{-2}}\frac{\left(1+\frac{(1+\alpha)}{a}z^{-1}\right)}{\left(1+\frac{(1+\alpha)}{a}z\right)}\right]^2 \\ &= \left[\frac{(\alpha-1)\,a\left(1+\frac{(1+\alpha)}{a}z\right)z^{-1}}{1+a\alpha z^{-1}+\alpha^2 z^{-2}}\frac{\left(1+\frac{(1+\alpha)}{a}z^{-1}\right)}{z^{-1}\left(1+\frac{(1+\alpha)}{a}z\right)}\right]^2 \\ &= \left[\frac{(\alpha-1)\,a\left(z^{-1}+\frac{(1+\alpha)}{a}\right)}{1+a\alpha z^{-1}+\alpha^2 z^{-2}}\frac{\left(1+\frac{(1+\alpha)}{a}z^{-1}\right)}{z^{-1}\left(1+\frac{(1+\alpha)}{a}z\right)}\right]^2 \end{aligned}$$

where the part

$$\frac{(\alpha-1)\,a\left(z^{-1}+\frac{(1+\alpha)}{a}\right)}{1+a\alpha z^{-1}+\alpha^2 z^{-2}}$$

is of minimum phase and

$$\frac{\left(1+\frac{(1+\alpha)}{a}z^{-1}\right)}{z^{-1}\left(1+\frac{(1+\alpha)}{a}z\right)} \tag{5.37}$$

has unit magnitude response due to the equality

$$1+\frac{(1+\alpha)}{a}e^{-j\omega}=\overline{1+\frac{(1+\alpha)}{a}e^{j\omega}}.$$

The filter (5.37) is thus an all-pass filter. The zero and the pole of (5.37) are, respectively, $-(1+\alpha)/a\,(=1.10173)$ and $-a/(1+\alpha)\,(=0.9829)$. Both are very close to the point $(1,0)$ in the Z domain, and hence do not influence the phase response at high frequencies (see Figure 5.3). Hence, The phase response in Figure 5.1 at around 900 Hz is mainly determined by the minimum-phase portion of the Q filter. As the solid line has a rapid-changing magnitude response, Bode's Phase Formula (see Theorem 2.4 on page 20) thus explains the sharper slope of the solid line in the phase response.

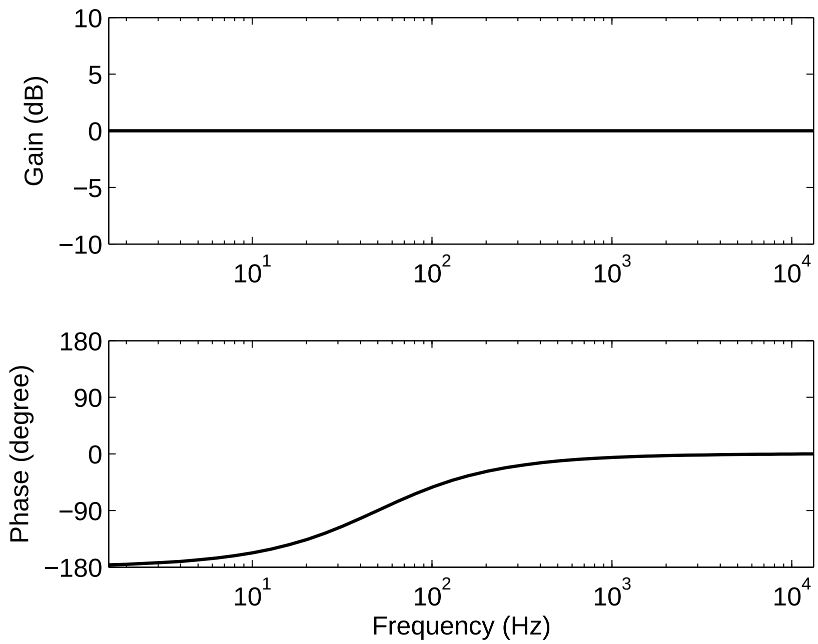


Figure 5.3: Frequency response of the all-pass component in the Q filter

Similar analysis can be applied to the Q filter from the direct approach

$$Q(z) = (\alpha - 1)\frac{(\alpha + 1 - a^2) - az^{-1}}{1 + \alpha a z^{-1} + \alpha^2 z^{-2}}$$

whose zero is at 1.0673—also close to the point $(1, 0)$.

The design approach in this chapter is intrinsically suitable for multiple narrow-band local loop shaping. Figure 5.4 shows the use of the direct-Q approach to obtain the previously mentioned example of a five-band Q filter design. Such complex bandpass $Q(z)$ is challenging to obtain using conventional summation formulation $Q(z) = \sum_{i=1}^{n} Q_i(z)$, especially when the frequencies of different bands are very close to each other.

Finally, examining Figure 5.5, the enlarged version of Figure 5.2, we see that $1 - z^{-m}Q(z)$ has slightly different behaviors outside the narrow band at 900 Hz: in the direct approach, the magnitude is almost strictly 1 (0 dB) for the majority of the frequency region, and slightly more amplified near 800 and 1000 Hz [about 1.185 (1.5dB) in the peak value]; in the multi-Q approach, $1 - z^{-m}Q(z)$ has flatter magnitude response but slightly less robustness against noise and uncertainties (recall Figure 5.1). For narrow-band loop shaping, the impact of the difference is quite small. When the desired attenuation region is wider, we need to be more careful about the waterbed effect. This is the topic of the next chapter.

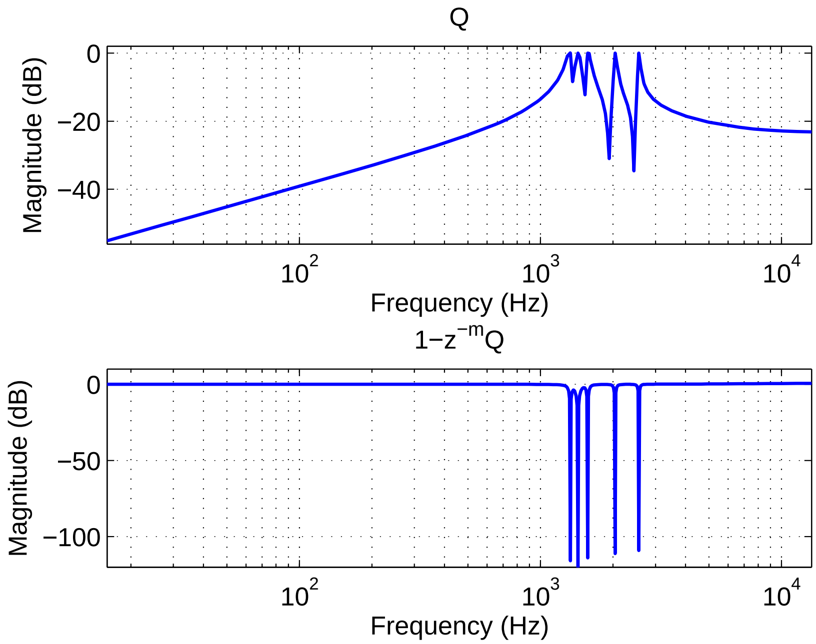


Figure 5.4: Multiple narrow-band Q-design example

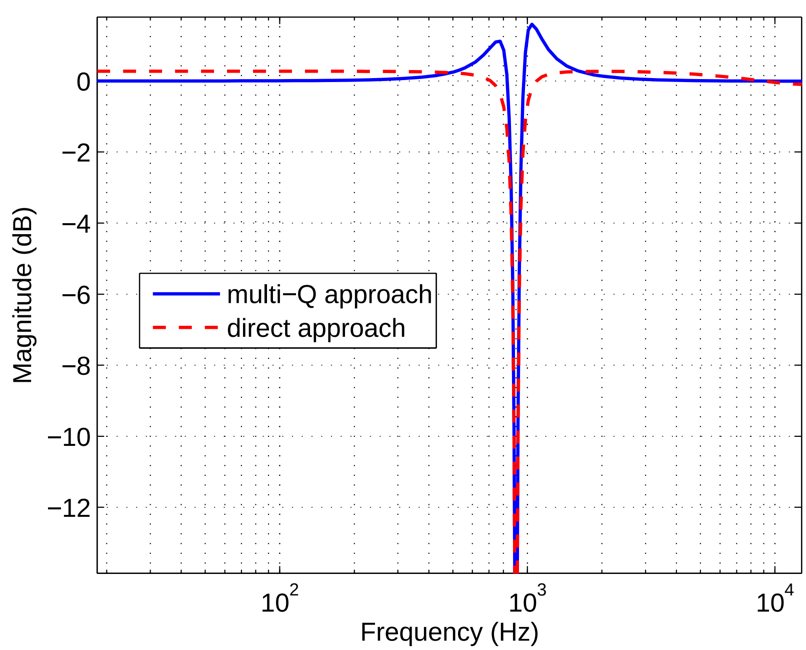


Figure 5.5: Internal-model based Q-design example: $1 - z^{-m}Q(z)$ (enlarged view)

5.6 Application: Vibration Rejection on an Active Suspension

The Pseudo Youla-Kucera parameterization with the multi-Q approach is applied to the benchmark on active suspension described in Section 3.3. In this system, the plant delay is two. Hence two $Q_{m=1}(z)$ blocks are needed in (5.24). Recall the frequency response of the plant in Figure 5.6. The system has not only nonminimum-phase zeros but also sharp resonant and anti-resonant modes, which make the control problem intrinsically difficult. As a benchmark, different types of vibration signals, in the frequency region between 50 Hz and 95 Hz, are applied to the system to test the performance, robustness, and complexity, of the algorithm. The benchmark compares the open- and closed-loop residual errors among all participants. As the baseline controller differs in each research group and the open loop is robustly stable, we set our baseline controller to have very small gains ($C_{\text{baseline}} \approx 0$), so that the performance comparison reflects directly the effect of the add-on YK parameterization.

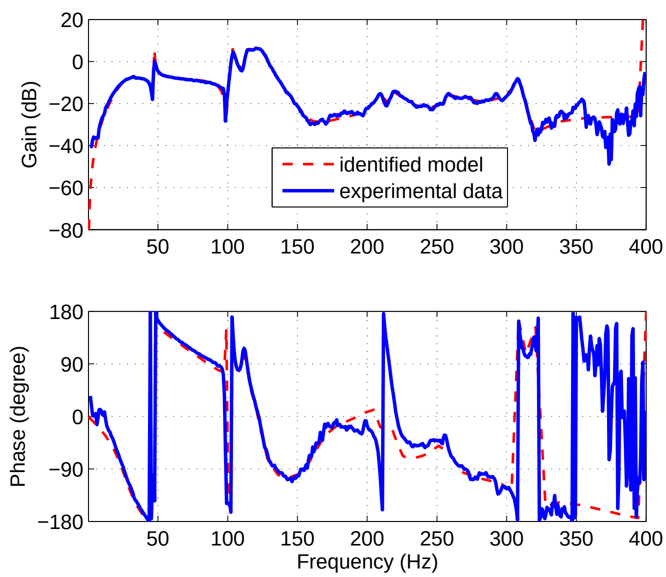
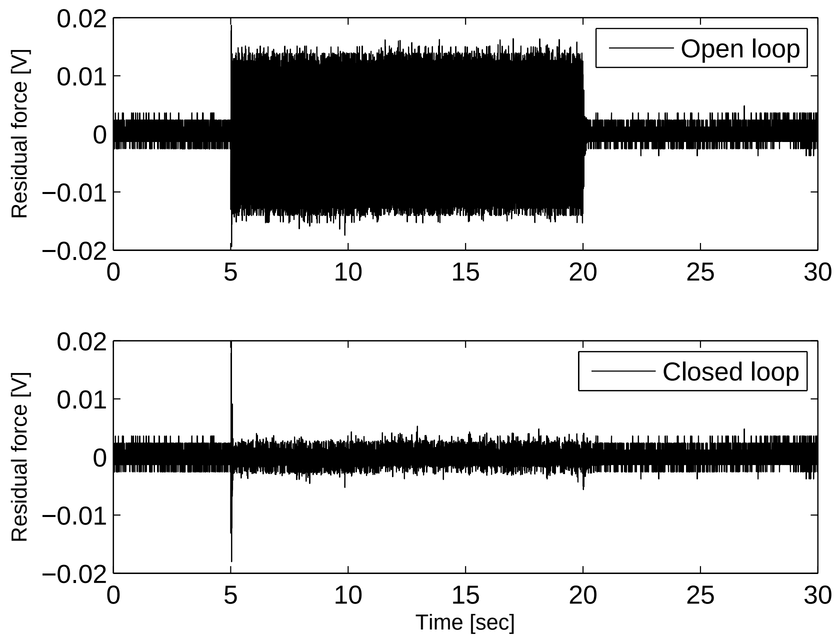


Figure 5.6: Frequency response of the plant in the active suspension

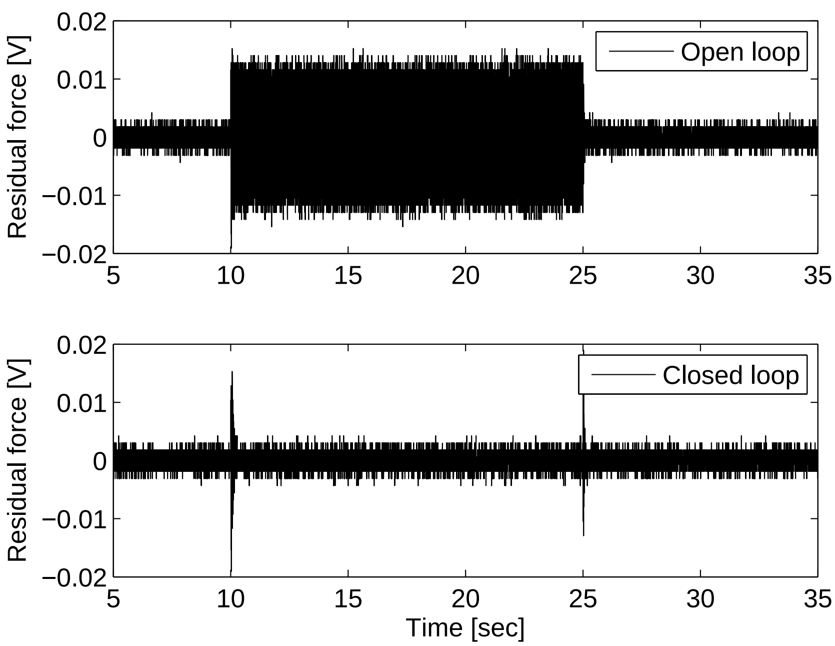
Three levels of evaluations are conducted:

Level 1: rejection of one narrow-band vibration

Figures 5.7 and 5.8 present the time trace and the steady-state error spectra when the system is subjected to a 70 Hz disturbance. In the time trace, the disturbance have been significantly attenuated such that vibration-induced errors are visually invisible under the compensation scheme. Indeed in Figure 5.8, the spectral peak at 70 Hz has reduced from around -42 dB



(a) Simulation: vibration injected at the fifth second



(b) Experiment: vibration injected at the tenth second

Figure 5.7: Time-domain results of rejecting a 70 Hz vibration

Table 5.1: Simulation and experimental results of rejecting vibrations with unknown constant frequencies (level 1)

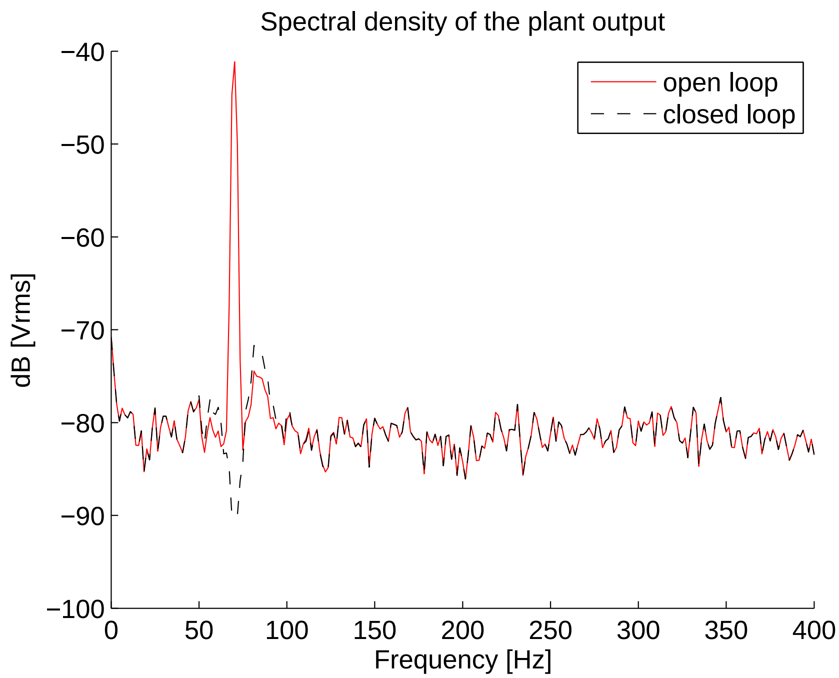
	freq. (Hz)	**global attenuation** (dB)	**disturbance attenuation** (dB)	**maximum amplification** (dB)		**freq.** (Hz)	**global attenuation** (dB)	**disturbance attenuation** (dB)	**maximum amplification** (dB)
Simulation	50	34.55	51.18	3.10	**Experiment**	50	35.80	46.61	7.58
	55	34.40	56.55	4.04		55	35.43	51.38	10.91
	60	34.40	52.22	3.33		60	35.51	52.10	8.89
	65	34.43	50.93	3.08		65	33.54	53.89	7.52
	70	34.44	55.44	2.79		70	31.42	48.37	8.02
	75	34.77	54.76	2.96		75	31.05	49.01	7.85
	80	34.99	46.99	3.74		80	31.44	49.04	9.29
	85	34.62	45.87	4.88		85	30.23	45.70	6.63
	90	32.53	45.83	4.25		90	29.40	42.62	8.20
	95	25.39	48.33	4.72		95	26.42	31.58	6.64

to -90 dB, indicating a disturbance attenuation of about 48 dB and a full removal of the 70-Hz peak. These results are directly reflected in Figure 5.9, which presents the magnitude responses of the sensitivity functions (note the sharp notch at 70 Hz and small amplifications at other regions).[4] There is a permanent disturbance at around 50 Hz and some other small spectral peaks at low frequencies in experiments. They did not appear during simulation but are however not amplified during experiments.

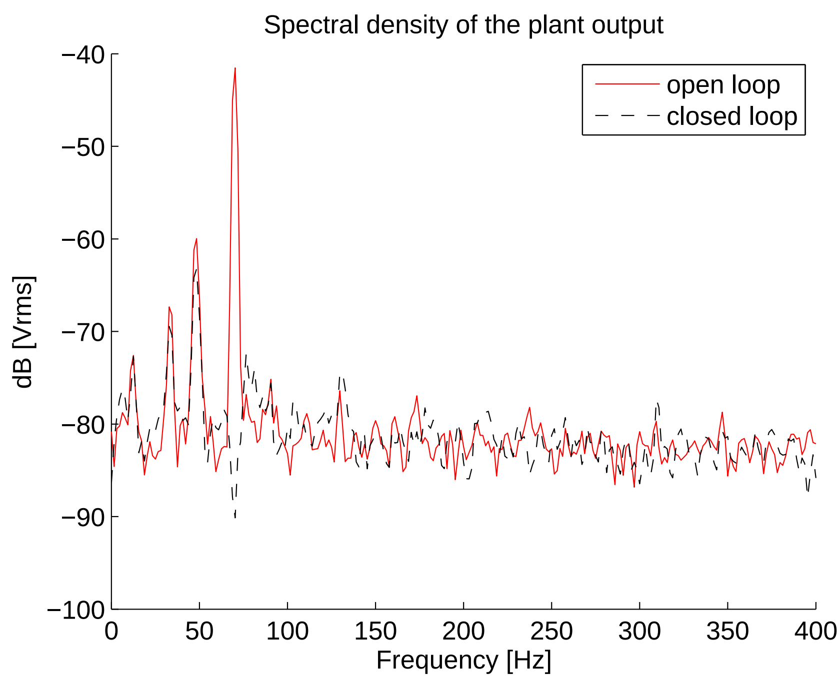
The algorithm is additionally tested at frequencies uniformly sampled between 50 Hz and 90 Hz. Table 5.1 summarizes the overall 2-norm reduction of the errors (global attenuation), the disturbance attenuation at the vibration frequencies, and the maximum amplification in the error spectrum. Similar to Figures 5.7 and 5.8, in all tests, the narrow-band disturbance has been greatly attenuated within a short period of time.[5]

[4] For demonstration of the pure disturbance-rejection performance, we used a very weak baseline controller here. This is feasible as the plant is open-loop stable.

[5] As the system has two pairs of strong resonant and anti-resonant modes neat 50 Hz and 100 Hz (recall Figure 3.15), we intentionally reduced the depth of attenuation at frequencies below 52 Hz and above 92 Hz in the experiments. This is done by selecting α to be closer to 1, and hence a sharper notch shape for $1-z^{-m}Q(z)$. The simulation results do not have this modification and reflects the best possible performance.



(a) Simulation



(b) Experiment

Figure 5.8: Error spectra of rejecting a 70 Hz vibration

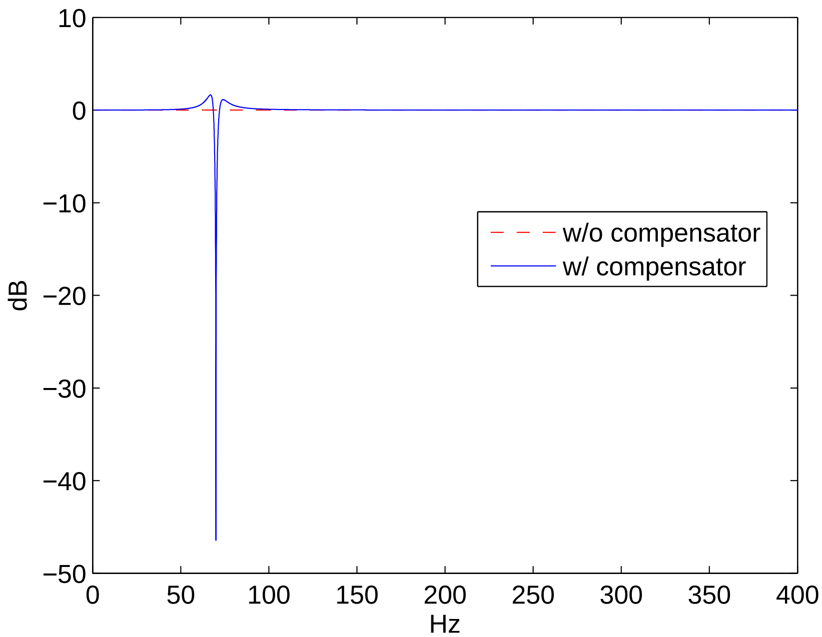


Figure 5.9: Magnitude response of the sensitivity function with $C_{\text{baseline}} \approx 0$

Level 2 and level 3: rejection of two and three narrow-band vibrations

The simulation and experimental results for level-2 and -3 tests are summarized respectively in Tables 5.2 and 5.3. Sample results are shown in Figures {5.10, 5.11, 5.12}. The complexity of the problem has been much increased, especially in the level-3 test. It appears the strong vibrations have excited other system modes at the beginning of all experiments (see the small side peaks in Figure 5.12). Yet we observe consistent and strong error reduction. In particular the waterbed effect has been very well controlled: with strong attenuation at the disturbance frequencies, the spectral amplification has been maintained very small.[6] Actually, the maximum amplification of our approach is the smallest among the benchmark participants [25].

The benchmark has set up several evaluation quantities about overall performance, robustness, and complexity. The proposed algorithm achieved 100% in the benchmark specification index for transient performance; 100%, 100%, and 99.78% respectively in the steady-state simulation performance of level 1, level 2, and level 3; and ranked one, three, and two, respectively in the experimental results of different levels. The recorded task execution time, which measures the algorithm complexity, is also small among the benchmark participants. Detailed summaries and discussions are provided in [25, 7].

[6]Compared to the simulation result, actual experiments includes various other random disturbances, which account for the larger values in the maximum spectral amplification.

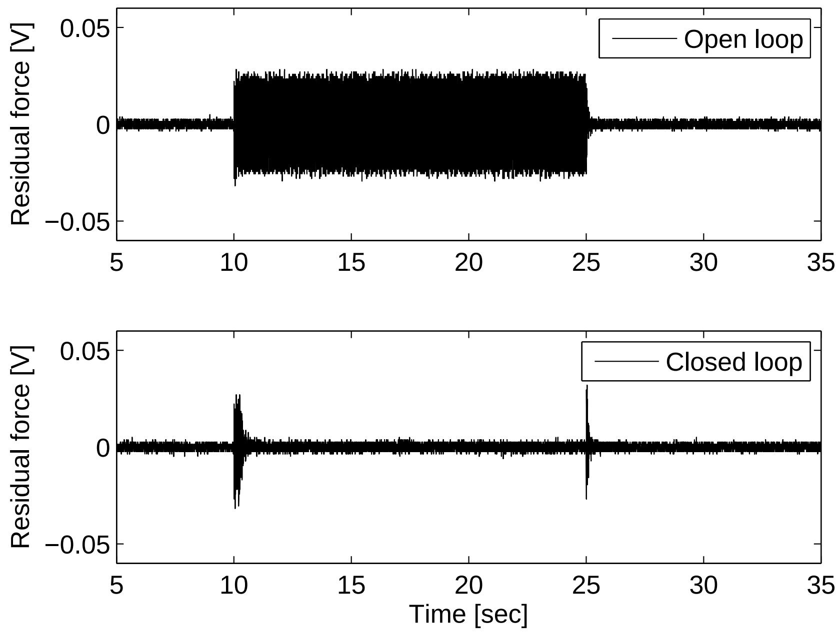


Figure 5.10: Time-domain experimental result of rejecting two narrow-band vibrations

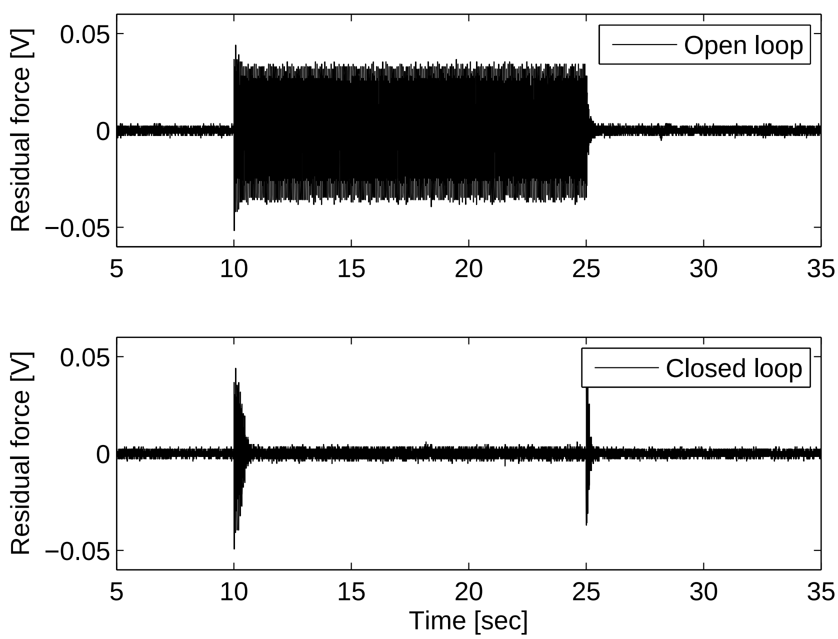


Figure 5.11: Time-domain experimental result of rejecting three narrow-band vibrations

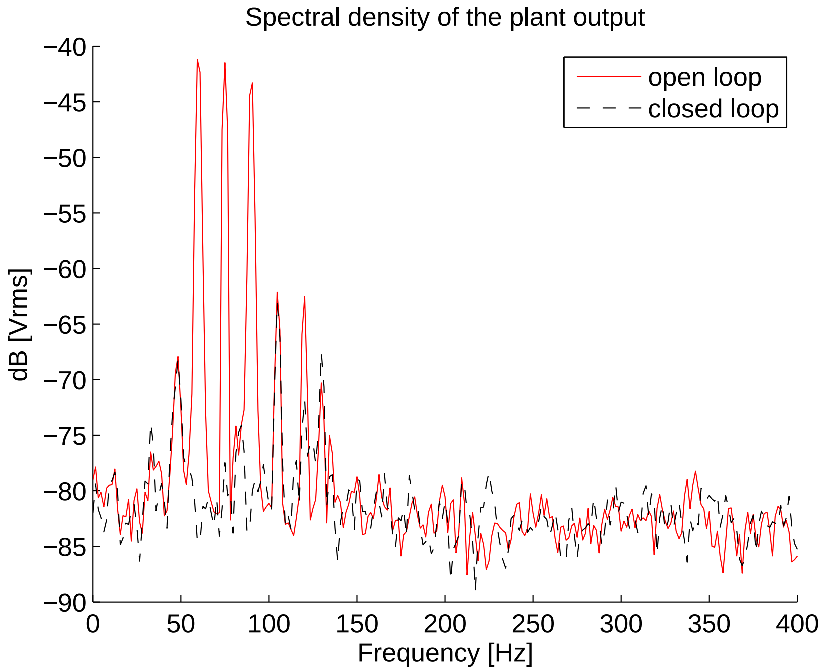


Figure 5.12: Frequency-domain experimental result of rejecting three narrow-band vibrations

Table 5.2: Simulation and experimental results of rejecting vibrations with unknown constant frequencies (level 2)

	frequency (Hz)	**global attenuation** (dB)	**disturbance attenuation** (dB)-(dB)	**maximum amplification** (dB)
Simulation	50,70	39.87	(43.87)(49.81)	5.17
	55,75	40.00	(51.22)(50.18)	3.83
	60,80	40.35	(47.51)(42.28)	5.49
	65,85	40.38	(46.66)(42.15)	6.69
	70,90	39.66	(50.24)(41.23)	6.06
	75,95	37.33	(49.84)(43.08)	4.87
Experiment	50,70	37.56	(42.95)(45.04)	10.77
	55,75	38.56	(47.11)(44.71)	7.98
	60,80	39.83	(41.91)(39.05)	8.10
	65,85	35.31	(50.39)(38.52)	11.27
	70,90	37.05	(44.28)(37.33)	7.47
	75,95	35.31	(46.31)(33.15)	9.04

Table 5.3: Simulation and experimental results of rejecting vibrations with unknown constant frequencies (level 3)

	frequency (Hz)	**global attenuation** (dB)	**disturbance attenuation** (dB)-(dB)-(dB)	**maximum amplification** (dB)
Sim.	50,65,80	43.91	(42.63)(39.54)(40.76)	6.57
	55,70,85	43.93	(47.93)(44.80)(40.36)	4.91
	60,75,90	43.48	(45.57)(47.12)(40.29)	4.88
	65,80,95	42.00	(44.58)(39.94)(41.39)	5.11
Exp.	50,65,80	41.97	(38.48)(45.66)(42.86)	7.54
	55,70,85	39.59	(44.79)(44.41)(37.54)	9.46
	60,75,90	38.31	(42.65)(41.75)(35.95)	8.27
	65,80,95	39.01	(43.70)(37.90)(33.14)	8.26

5.7 Application: Unnatural-torque Compensation in Active Steering

The Q filter design using the direct approach is evaluated on the active steering system described in Section 3.4 on page 36.

In one level of the baseline electric power steering (EPS) system, we have a velocity feedback control loop to control the EPS motor speed in Figure 5.13. Figure 5.14 shows the frequency response of the plant. The system has a 4ms input delay, yielding $m = 4$ in the Pseudo YK parameterization and Q-filter design.

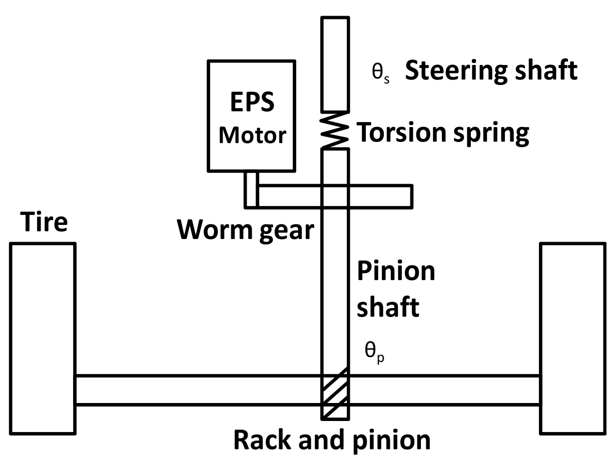


Figure 5.13: Hardware configuration of the EPS components

The top plot of Figure 5.15 demonstrates the motor velocity during a variable-speed steering test, where we observe large tracking errors between the reference velocity and the

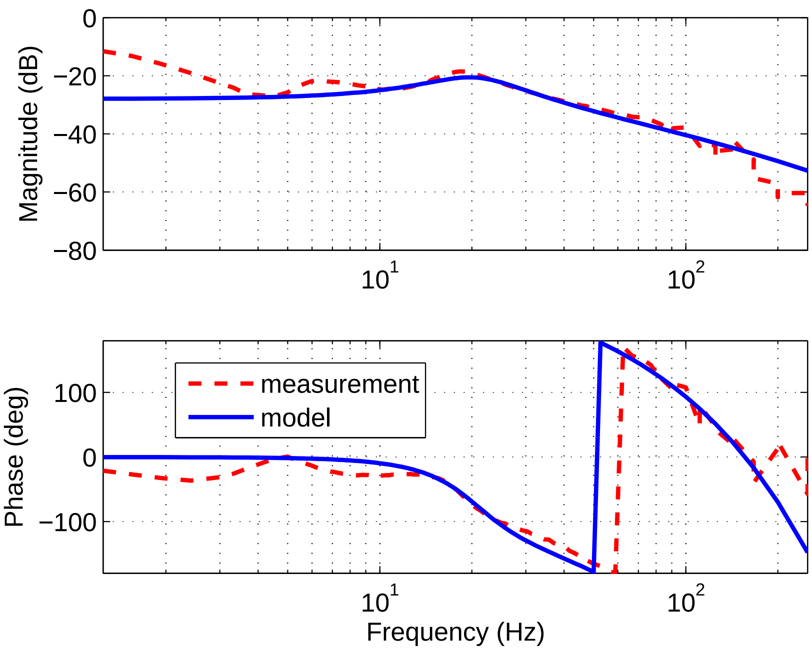


Figure 5.14: Open-loop frequency response from the EPS controller output to the pinion speed: dashed–measurement; solid–$\hat{P}(z)$

actual motor speed. Further investigation shows that there are strong narrow-band vibrations due to imperfect motor rotations. The dashed line in Figure 5.16 presents the spectrum of the tracking errors when we apply a constant-velocity steering. The strong spectral peak at 15 Hz contributes greatly to the tracking errors.

Applying the proposed direct Q-filter design yields the solid line in Figure 5.16. Visual comparison indicates that the algorithm has removed the original spectral peak at 15 Hz. Computing the standard deviations for the tracking errors, we obtain a 74.77% error reduction on the 3σ value (from 0.20826 rad/s to 0.052549 rad/s).

The vibration frequency actually is time-varying. In the constant-speed tests, alternating the motor speed ω_m reveals that the vibration frequency f_v changes according to $f_v = \gamma_n \omega_m$, where γ_n depends on the number of magnetic pairs in the motor. Using this result, an adaptive/speed-dependent Q filter is constructed for the variable-speed test in Figure 5.15. We see from the bottom plot, that the algorithm provides significant performance enhancement, both visually and quantitatively (3σ reduces from 0.19402 to 0.102).

With the improvements in the low-level control, the full EPS system with variable gear ratio gained much better overall response. Figure 5.17 shows the actual steering torque under a variable speed steering test. Compared to Figure 3.20, the case without the proposed vibration compensation, the steering-wheel angle is better maintained at a constant value, and the errors in the steering torque is also greatly reduced.

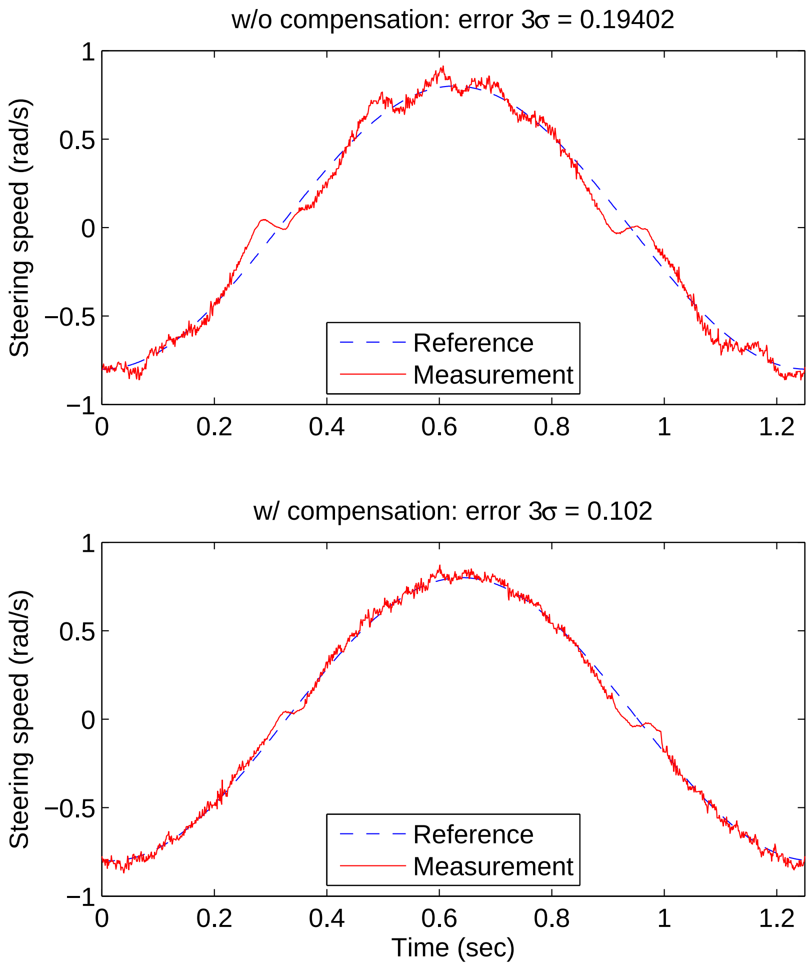


Figure 5.15: Time traces of the EPS tracking result during variable-speed steering

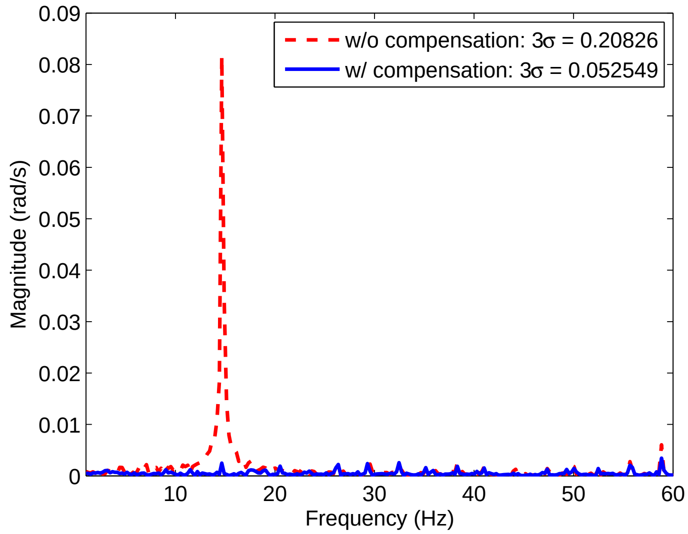


Figure 5.16: Spectra of the tracking errors during constant-speed steering

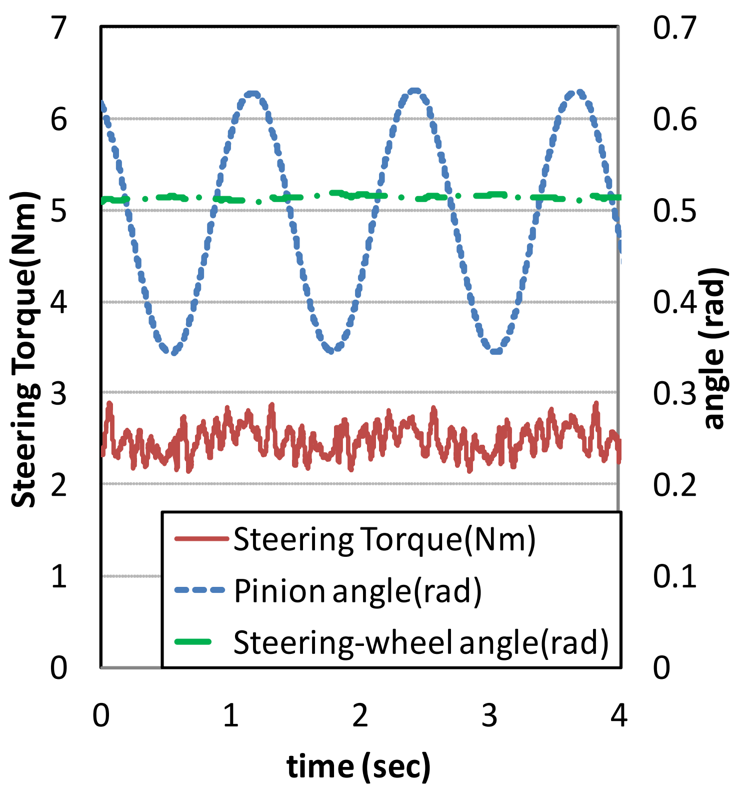


Figure 5.17: Steering torque with unnatural-torque compensation

Chapter 6

Advanced Q-filter Design: "Shaping the Waterbed"

We have discussed two approaches in designing

$$1 - z^{-m}Q(z) = \frac{A(z)}{A_Q(z)}K(z) \tag{6.1}$$

so that local loop shaping can be achieved via

$$S\left(e^{j\omega}\right) \approx S_o\left(e^{j\omega}\right)\left(1 - e^{-mj\omega}Q\left(e^{j\omega}\right)\right). \tag{6.2}$$

The dashed line in Figure 6.1 presents an example of the solved $Q(z)$ and $1 - z^{-m}Q(z)$ using the direct approach, with $\alpha = 0.993$, $T_s = 1/26400$ sec, $\Omega_0 = 3000$ Hz, and $m = 2$. $1 - z^{-m}Q(z)$ is approximately unity except at the desired attenuation frequency 3000 Hz, where we have

$$1 - e^{-mj\omega_0}Q(e^{j\omega_0}) = 0,\ \omega_0 = 2\pi\Omega_0 T_s.$$

This is because

$$A\left(e^{j\omega_0}\right) = 1 - 2\cos\left(\omega_0\right)e^{-j\omega_0} + e^{-2j\omega_0} = 0$$

from the internal model principle. Thus, $S(e^{j\omega_0}) = 0$ and disturbances at 3000 Hz get perfectly attenuated. Meanwhile, the magnitude of $Q(z)$ reduces from 0dB at 3000 Hz quickly down to -35dB (0.0178 in absolute value) in the low-frequency region, and -50dB (0.0032) in the high-frequency region. These bandpass properties reduce the influence of the model uncertainties, and make (6.2) a valid approximation.

With the same center-frequency configuration, the solid line in Figure 6.1 shows the solved $Q(z)$ and $1 - z^{-m}Q(z)$ for $\alpha = 0.945$. The -3dB bandwidth of the pass band for this $Q(z)$ is approximately 475 Hz. As we increase the servo-enhancement region, some amplifications [less than 1.2dB (= 1.1482)] occur in the magnitude response of $1 - z^{-m}Q\left(z\right)$. In simple situations where one wide notch or multiple narrow-band notches are sufficient for the control objective, the small amplification is an acceptable trade off. For more complicated cases, such waterbed-type phenomenon needs more careful treatments. In this chapter, we first prove the inevitability of the amplification, then provide ways to control it.

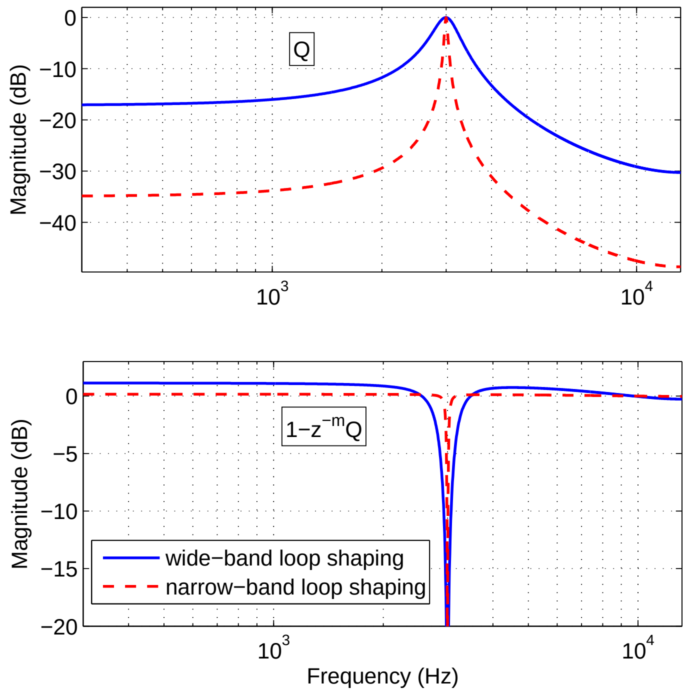


Figure 6.1: Q-filter design by direct approach: from narrow band to wide band

6.1 The Fundamental Limitation

The fact that $1 - z^{-m}Q(z)$ always has magnitudes higher than one, is due to several results in fundamental complex analysis about analytic functions, harmonic functions, and Poisson Integrals.

Theorem 6.1. *Let*

$$1 - z^{-m}Q(z) = \frac{A(z)}{A_Q(z)}K(z)$$

where $Q(z)$ is a proper and rational transfer function in z; $K(z)$ is stable; and

$$A(z) = \prod_{i=1}^{n}\left(1 - 2\cos(\omega_i)z^{-1} + z^{-2}\right)$$

$$A_Q(z) = \prod_{i=1}^{n}\left(1 - 2\alpha\cos(\omega_i)z^{-1} + \alpha^2 z^{-2}\right).$$

If $K(z)$ has l unstable zeros $\{\gamma_i\}_{i=1}^{l}$, then

$$\frac{1}{\pi}\int_0^{\pi} \ln\left|1 - e^{-mj\omega} Q\left(e^{j\omega}\right)\right| d\omega = \sum_{i=1}^{l} \ln|\gamma_i| \tag{6.3}$$

As a special case, when $K(z)$ is a minimum-phase transfer function, then

$$\frac{1}{\pi}\int_0^{\pi} \ln\left|1 - e^{-mj\omega} Q\left(e^{j\omega}\right)\right| d\omega = 0 \tag{6.4}$$

Proof. see Appendix A.4. □

Theorem 6.1 explains the reason for the shape of $1-z^{-m}Q(z)$ in Figure 6.1. An important remark for Figure 6.1 is that, although the waterbed effect is inevitable, we can control it and achieve smooth spread of the magnitude increase for $1-z^{-m}Q(z)$, to avoid introducing new spectral peaks in the residual error.

Corollary 6.1. *(Sensitivity Integral for Add-on Loop Shaping in Pseudo YK Parameterization) Under perfect-model assumption in pseudo Youla-Kucera parameterization, we have*

$$\frac{1}{\pi}\int_0^{\pi} \ln\left|S\left(e^{j\omega}\right)\right| d\omega = \frac{1}{\pi}\int_0^{\pi} \ln\left|S_0\left(e^{j\omega}\right)\right| d\omega$$

if $1-z^{-m}Q(z)$ does not have unstable zeros. Otherwise

$$\frac{1}{\pi}\int_0^{\pi} \ln\left|S\left(e^{j\omega}\right)\right| d\omega = \frac{1}{\pi}\int_0^{\pi} \ln\left|S_0\left(e^{j\omega}\right)\right| d\omega + \sum_{i=1}^{l} \ln|\gamma_i|$$

where $\{\gamma_i\}_{i=1}^{l}$ are the unstable zeros of $1-z^{-m}Q(z)$.

Proof. From (4.20),

$$S\left(e^{j\omega}\right) = S\left(e^{j\omega}\right)\left(1 - e^{-mj\omega} Q\left(e^{j\omega}\right)\right)$$

if there is no mismatch between $P\left(e^{j\omega}\right)$ and $\hat{P}\left(e^{j\omega}\right)$. Hence

$$\frac{1}{\pi}\int_0^{\pi} \ln\left|S\left(e^{j\omega}\right)\right| d\omega = \frac{1}{\pi}\int_0^{\pi} \ln\left|S_0\left(e^{j\omega}\right)\right| d\omega + \frac{1}{\pi}\int_0^{\pi} \ln\left|1 - e^{-mj\omega} Q\left(e^{j\omega}\right)\right| d\omega$$

The result follows immediately from Theorem 6.1. □

6.2 Gain Scheduling

The simplest approach to reduce the amplification is to design first the regular Q filter in (6.1) and then scale down its magnitude via

$$Q(z) \longleftarrow gQ(z), \; g \leq 1. \tag{6.5}$$

By making $Q(e^{j\omega})$ smaller overall, the value of $1 - e^{-jm\omega}Q(e^{j\omega})$ is closer to unity, especially outside the pass band of $Q(z)$. Of course, this is accomplished by the lost of perfect-error-reduction property. Yet this simple gain-scheduling scheme offers quite effective loop shapes as shown in Figure 6.2.

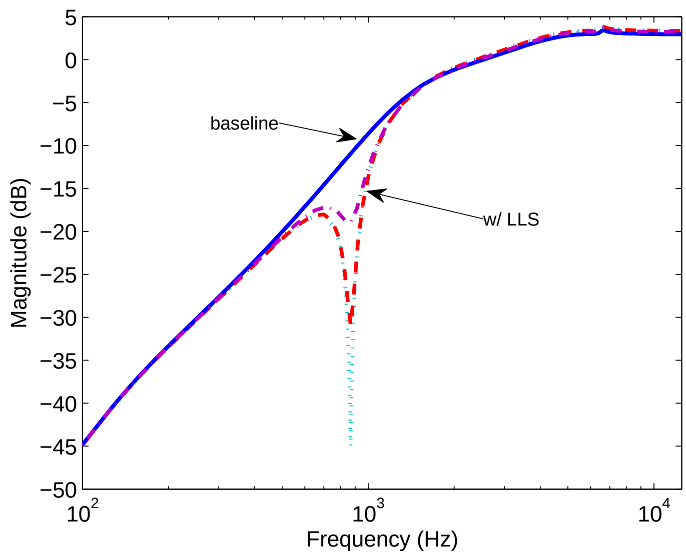


Figure 6.2: Example wide-band local loop shaping results

Next, we discuss additional methods to flexibly control the waterbed when designing the Q filter. These can be combined with (6.5) when the simple gain-scheduling scheme is insufficient itself to provide a stable and high-performance closed loop.

6.3 Add-on Pole and Zero Placement

The Effect of Fixed Zeros

In Chapter 5, we chose the IIR $Q(q^{-1})$ with a customized denominator $A(\alpha q^{-1})$, but have not placed specific structural designs for the numerator $B_Q(q^{-1})$. Actually $B_Q(q^{-1})$ is the unknown to be solved in

$$K\left(q^{-1}\right)A\left(q^{-1}\right)+q^{-m}B_Q\left(q^{-1}\right)=A\left(\alpha q^{-1}\right) \tag{6.6}$$

and its nontrivial frequency response is solely determined by the algebraic equation (6.6).

We can add a fixed part $B_0(q^{-1})$ such that

$$B_Q(q^{-1})=B_0(q^{-1})B_Q^{'}(q^{-1}). \tag{6.7}$$

Designing $B_0(q^{-1})=1-q^{-1}$ for example, will add a scaled differentiator in $Q\left(q^{-1}\right)$, yielding $Q(e^{j\omega})|_{\omega=0}=0$, i.e., zero magnitude at DC frequency. Figure 6.3 presents the effect of such a design: in addition to a gain scaling of $gQ\left(z\right)$, the enhanced small gain at low frequencies clearly makes the magnitude response of $1-z^{-m}Q\left(z\right)$ smaller in the highlighted dark region.

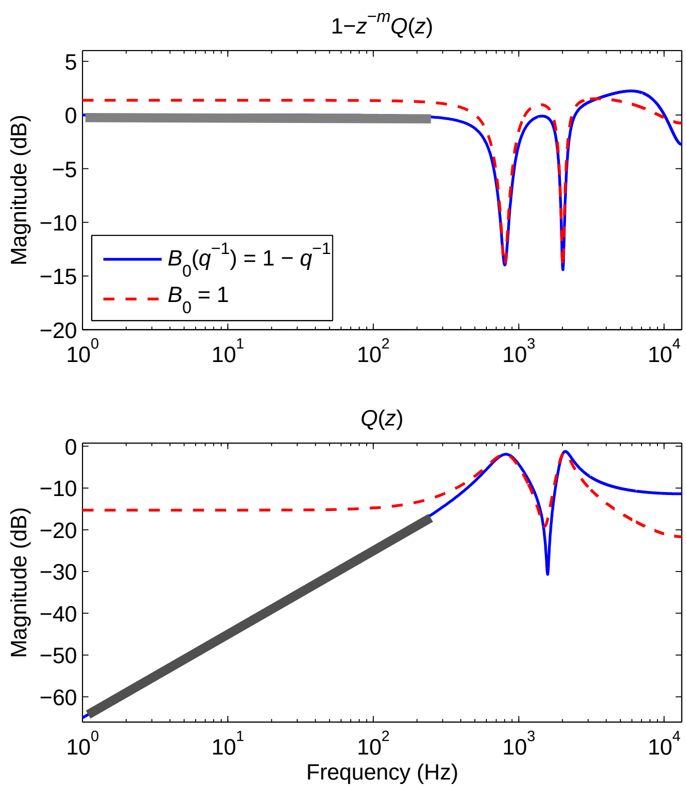


Figure 6.3: Effect of a fixed zero at low frequency

More generally, introducing fixed zero near $q = -1/q = 1$ in the Z plane will provide enhanced small gains for $Q(q^{-1})$ in the high/low-frequency region. Extending this idea, we can essentially place magnitude constraints at arbitrary desired frequencies, by letting

$$B_0(q^{-1}) = 1 - 2\beta \cos \omega_p q^{-1} + \beta^2 q^{-2} \tag{6.8}$$

in (6.7), which places the fixed zeros $\beta e^{\pm j\omega_p}$ to penalize the magnitude of the Q filter near ω_p. Figures 6.4 and 6.5 demonstrate two examples of such additional frequency based constraints. The effect of $B_0(q^{-1})$ is immediately seen in both $Q(q^{-1})$ and $1 - q^{-m}Q(q^{-1})$.

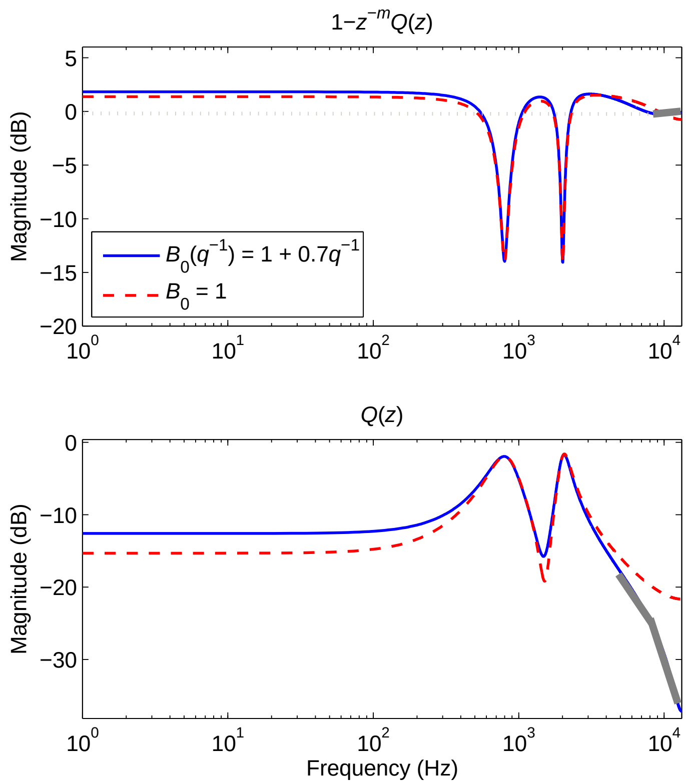


Figure 6.4: Effect of a fixed zero at high frequency

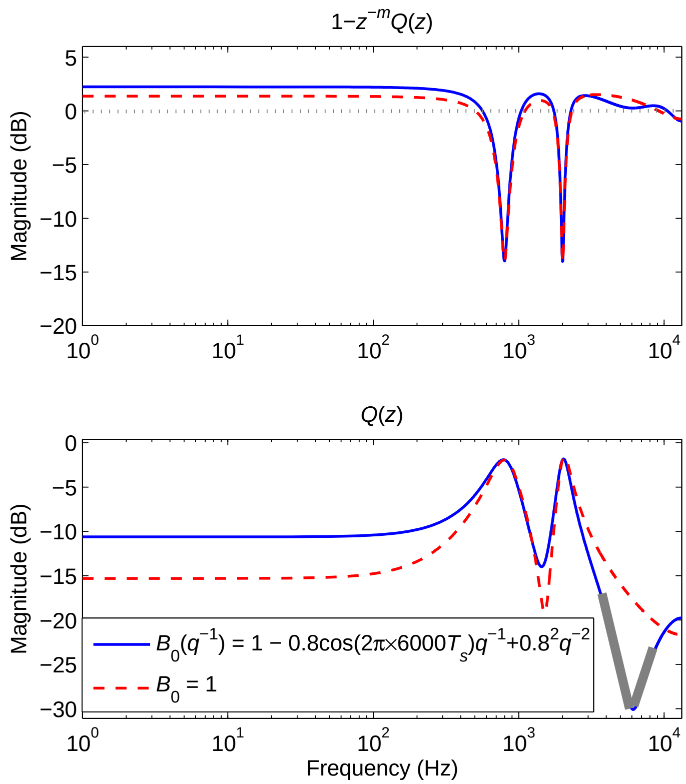


Figure 6.5: Effect of a fixed zero at a specific frequency

Table 6.1 summarizes the effects of different configurations for the fixed term $B_0(q^{-1})$. For band-limited loop shaping, it is natural to place magnitude constraints at both low and high frequencies. In this case, the modules in Table 6.1 can be combined, e.g., as shown in Figure 6.6, to form the overall enhancement in Figure 6.7.

Table 6.1: Effects of placing fixed zeros to $Q(q^{-1})$

$B_0(q^{-1})$	zeros	small $\lvert Q(e^{j\omega})\rvert$
$1+q^{-1}$	-1	around Nyquist freq.
$1+\rho q^{-1},\ \rho\in[0.5,1]$	$-\rho$	at high freq.
$1-\rho q^{-1},\ \rho\in[0.5,1]$	ρ	at low freq.
$1-2\beta\cos\omega_p q^{-1}+\beta^2q^{-2}$ $\beta\in(0,1]$	$\beta e^{\pm j\omega_p}$	around ω_p
$1-q^{-1}$	1	at low freq.

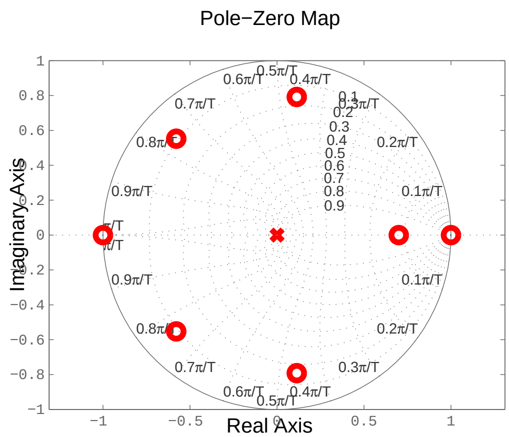


Figure 6.6: Location of a combination of fixed zeros for $Q\left(q^{-1}\right)$

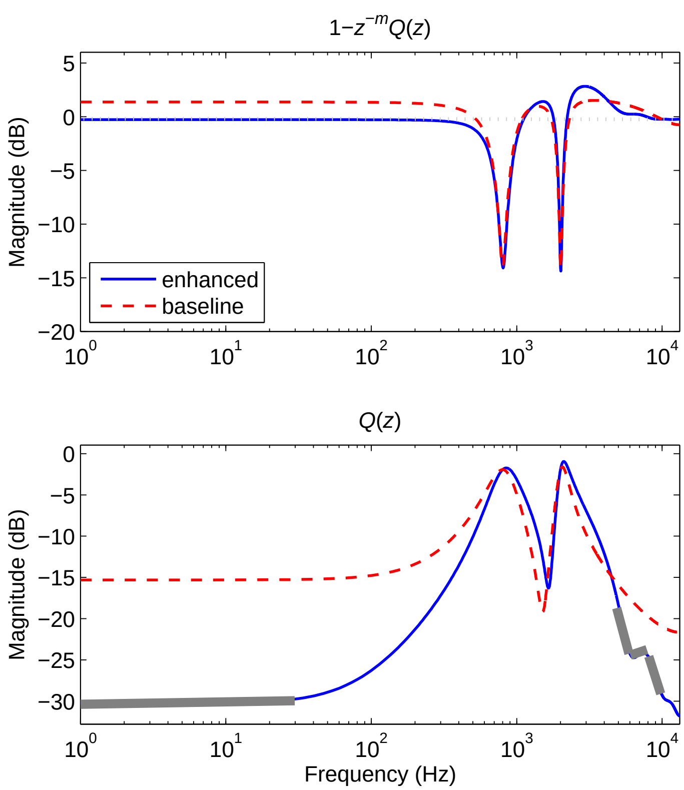


Figure 6.7: Effect of combined zeros at different frequencies

The Effects of Cascaded IIR Filters

By (6.7) we essentially have cascaded the FIR filter $B_0(q^{-1})$ to $Q(q^{-1})$. $B_0(q^{-1})$ has been designed to control the magnitude response of $Q(q^{-1})$ at some specific frequency regions. IIR design can potentially create additional flexibility compared to FIR filters. From the frequency-response perspective, cascading two bandpass filters with the same center frequency provides a new bandpass $Q(q^{-1})$, which can have reduced magnitudes at all frequencies outside the passband. This is inspired by the multi-Q approach in the previous chapter, and suggests us to assign an IIR bandpass $B_0(q^{-1})$ in

$$Q(q^{-1}) = Q_0(q^{-1})B_0(q^{-1}),$$

where $Q_0(q^{-1})$ is a regular solution from the direct or multi-Q approaches in Chapter 5. Note that $Q_0(q^{-1})$ is not a conventional bandpass filter in the sense that

$$Q_0(e^{j\omega_i}) = e^{jm\omega_i}$$

at its center frequency ω_i [see (5.19) and (5.27)]. $B_0(q^{-1})$ hence needs to satisfy $B_0(e^{j\omega_i}) = 1$ to preserve the property $Q_0(e^{j\omega_i})B_0(e^{j\omega_i}) = Q_0(e^{j\omega_i}) = e^{jm\omega_i}$. A standard bandpass filter will satisfy this requirement. Recall that $A(q^{-1})/A(\alpha q^{-1})$ is a notch filter. One candidate $B_0(q^{-1})$ is

$$B_0\left(q^{-1}\right) = 1 - \eta\frac{A(q^{-1})}{A(\alpha q^{-1})},\ \eta \in (0, 1],$$

as unity minus the notch shape $A(q^{-1})/A(\alpha q^{-1})$ generates a bandpass shape.

Figure 6.8 presents the $Q(q^{-1})$ and $1 - q^{-m}Q(q^{-1})$ solved from the discussed algorithms in this section. The solid lines are the direct-Q solution from Section 5.4; the dashed lines are respectively the FIR and the IIR enhancement from this section. We observe from the magnitude responses of $1-q^{-m}Q(q^{-1})$, that all three methods create the required attenuation at around 3000 Hz. Also, the additional magnitude constraints on $Q(q^{-1})$ are effectively reflected in the bottom plot of Figure 6.8: in the dashed-line $Q(q^{-1})$, the design of $B_0(q^{-1}) = 1 + 0.7q^{-1}$ places a zero $q = -0.7$ near the Nyquist frequency ($e^{\pi} = -1$), yielding the small gain in the high-frequency region compared to the solid-line $Q(q^{-1})$; in the dotted-line $Q(q^{-1})$, by cascading the bandpass filter $B_0(q^{-1})$ $(= 1 - \alpha A(q^{-1})/A(\alpha q^{-1}))$, we have reduced the magnitude of $Q(q^{-1})$ at both low and high frequencies.

In general, it is preferred to evenly spread the amplifications, so the solid or the dashed lines may usually be preferred from the performance perspective. Yet if large model uncertainty exists which enforces $Q(q^{-1})$ to have small magnitudes at high and/or low frequencies, the dotted line may be considered over the other designs. Nonetheless, the maximum amplification among all designs is around 1.6dB (1.2023) while the attenuation is quite large in a wide frequency region.

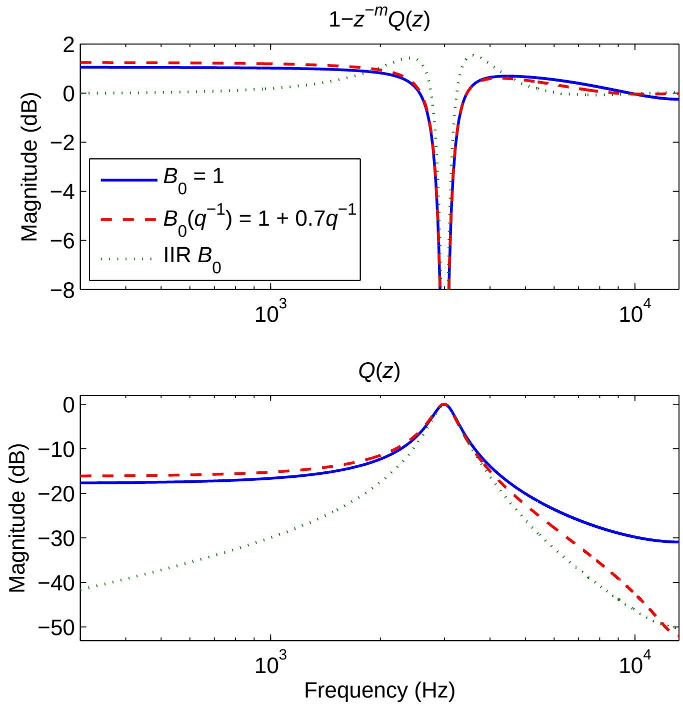


Figure 6.8: Comparison of the magnitude responses in three designs of wide-band Q filter

6.4 Design Based on Convex Optimization

In this section, we formulate the Q-design problem in an optimization framework. Using convex optimization, we are able to design $Q(z)$ with arbitrary magnitude (upper) bounds, and at the same time minimize the disturbance amplification in the closed-loop system.

Consider the following construction:

$$1 - z^{-m}Q(z) = F_{nf}(z)K(z), \tag{6.9}$$

$$K(z) = k_1 + k_2 z^{-1} + \dots k_{n_k+1} z^{-n_k}. \tag{6.10}$$

Here $F_{nf}(z)$ is a notch filter that provides the desired low gains (in a range of frequencies) to the sensitivity function; $K(z)$ is introduced for causality of $Q(z)$ and provides additional optimal properties to $Q(z)$.

With the flexibility in optimization, we can make the relaxation that instead of taking the specific damped-pole form

$$F_{nf}(z) = \frac{1 + a_1 z^{-1} + \cdots + a_n z^{-n} + \cdots + a_1 z^{-2n+1} + z^{-2n}}{1 + a_1 \alpha z^{-1} + \cdots + a_n \alpha^n z^{-n} + \cdots + a_1 \alpha^{2n-1} z^{-2n+1} + \alpha^{2n} z^{-2n}},$$

the notch filter can have a general structure

$$F_{nf}(z) = \frac{B_{nf}(z)}{A_{nf}(z)} \tag{6.11}$$

with

$$\begin{aligned} B_{nf}(z) &= b_1 + b_2 z^{-1} + \cdots + b_{n_b+1} z^{-n_b} \\ A_{nf}(z) &= a_1 + a_2 z^{-1} + \cdots + a_{n_a+1} z^{-n_a}. \end{aligned}$$

Causality Constraint

Solving (6.9) and (6.11) gives

$$Q(z) = z^m \frac{A_{nf}(z) - B_{nf}(z)K(z)}{A_{nf}(z)} =: z^m \frac{X(z)}{A_{nf}(z)} \tag{6.12}$$

$$X(z) = A_{nf}(z) - B_{nf}(z)K(z) \tag{6.13}$$

Similar as before, since z^m is not causal, the coefficients of z^{-i} need to be zero for $i = 0, 1, \dots, m-1$ in $X(z)$ to have a realizable $Q(z)$.[1] Expanding the convolution $B_{nf}(z)K(z)$ and grouping the coefficients in $A_{nf}(z) - B_{nf}(z)K(z)$, we can obtain the causality condition in the following matrix form:

$$\underbrace{\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix}}_{\tilde{A}} - \underbrace{\begin{bmatrix} b_1 & 0 & \dots & 0 & 0_{1,n_k+1-m} \\ b_2 & b_1 & 0 & \vdots & 0_{1,n_k+1-m} \\ \vdots & \ddots & \ddots & 0 & 0_{1,n_k+1-m} \\ b_m & \dots & b_2 & b_1 & 0_{1,n_k+1-m} \end{bmatrix}}_{\tilde{B}} \begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ \vdots \\ k_{n_k+1} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{6.14}$$

Here $a_i = 0$ if $i > n_a$ and $b_i = 0$ if $i > n_b$. Notice the columns of $\tilde{B}$ span $\mathbb{R}^m$ if $b_1 \neq 0$. Therefore, the value of $n_k + 1$ has to be no smaller than m for a solution to exist for a general $\tilde{A}$.

As an example, if $m = 4$ and

$$F_{nf}(z) = \frac{1 - 2\beta\cos\omega_0 z^{-1} + \beta^2 z^{-2}}{1 - 2\alpha\cos\omega_0 z^{-1} + \alpha^2 z^{-2}}$$

we have $n_a = 2$, $n_b = 2$, $n_k = 3$, and

$$\begin{bmatrix} 1 \\ -2\alpha\cos\omega_0 \\ \alpha^2 \\ 0 \end{bmatrix} - \begin{bmatrix} 1 & 0 & 0 & 0 \\ -2\beta\cos\omega_0 & 1 & 0 & 0 \\ \beta^2 & -2\beta\cos\omega_0 & 1 & 0 \\ 0 & \beta^2 & -2\beta\cos\omega_0 & 1 \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \\ k_3 \\ k_4 \end{bmatrix} = 0$$

[1] If $m = 0$, causality is automatically satisfied.

which has the solution

$$\begin{aligned}
k_1 &= 1 \\
k_2 &= -2\cos\omega_0\left(\alpha - \beta\right) \\
k_3 &= \alpha^2 - \beta^2 + 2\beta\cos\omega_0 k_2 \\
k_4 &= -\beta^2 k_2 + 2\beta\cos\omega_0 k_3.
\end{aligned}$$

and hence

$$Q(z) = \frac{-\beta^2 k_3 + 2\beta\cos\omega_0 k_4 - \beta^2 k_4 z^{-1}}{1 - 2\alpha\cos\omega_0 z^{-1} + \alpha^2 z^{-2}}.$$

Remark 6.1. In the above example, although the order of $K\left(z\right)$ is four, the solved $Q\left(z\right)$ is still second-order.

Optimal Performance

As has been shown in the last section, if $n_k + 1 = m$, the m equations in (6.14) define a unique solution for $K(z)$. Additionally, n_k can be set to be larger than $m-1$ so as to allow more design freedom in $Q(z)$.

First, we can minimize the infinity norm of $1 - z^{-m}Q(z)$ (maximum magnitude in frequency response), which will in turn minimize the error amplification in the sensitivity function $S\left(z\right) \approx S_o\left(z\right)\left(1 - z^{-m}Q\left(z\right)\right)$. This can be achieved by $\min\left|\left|K(z)\right|\right|_\infty$ in (6.9). Second, to keep the system robustly stable, the magnitude of $Q(z)$ should be small at frequencies outside its passbands, especially at high frequencies where large model uncertainty exists. These two objectives can be formulated into the following optimization problem:

$$\min : \left|\left|K(z)\right|\right|_\infty \tag{6.15}$$

$$\text{s.t.: causality constraint} \tag{6.16}$$

$$\left|Q(e^{j\omega})\right| \leq \delta\left(\omega\right) \;\; \omega = \omega_1, \omega_2 \ldots \tag{6.17}$$

where $\delta\left(\omega\right)$ is the user-defined magnitude upper bound at frequency ω.

The causality constraint (6.14) is a set of linear equations [explicitly $n_k > m-1$ is assumed in this case, since if $n_k = m-1$, then the solution of $K(z)$ is unique from (6.14)], and tractable in the optimization formulation.

Equations (6.15) and (6.17) however are not yet convex.

For (6.15), the H_∞ performance objective can be easily translated to a linear matrix inequality (LMI) by applying the bounded-real lemma (see, e.g., [82, 83]). Let A, B, C and D be the state-space matrices of $K(z)$. Then

$$\min : \left|\left|K(z)\right|\right|_\infty$$

is equivalent to

$$
\begin{aligned}
\min_{\gamma, M} \; &: \gamma \\
\text{s.t.} & \begin{bmatrix} A^T M A - M & A^T M B & C^T \\ B^T M A & B^T M B - \gamma I & D^T \\ C & D & -\gamma I \end{bmatrix} \preceq 0 \\
& M \succ 0 \\
& \gamma > 0.
\end{aligned}
$$

For (6.17), by applying convex-optimization techniques, we can transform the magnitude constraint to a set of convex quadratic constraints. To do so, notice first that

$$Q(z) = z^m \left(1 - F_{nf}(z^{-1}) K(z^{-1})\right). \tag{6.18}$$

Constraining $|Q(e^{j\omega})| \leq \delta(\omega)$ is equivalent to requiring $\left|Q(e^{j\omega})\right|^2 \leq \delta^2(\omega)$. The z^m term has unity gain, and can be dropped in the magnitude design. Notice also, that the frequency response of $K(e^{j\omega})$ is

$$K(e^{j\omega}) = \sum_{n=1}^{n_k+1} k_n e^{-j\omega(n-1)} = \phi_r^T(\omega)\,\theta - j\phi_i^T(\omega)\,\theta$$

where

$$
\begin{aligned}
\phi_r^T(\omega) &= \begin{bmatrix} 1 & \cos(\omega) & \ldots & \cos(n_k\omega) \end{bmatrix} \\
\phi_i^T(\omega) &= \begin{bmatrix} 0 & \sin(\omega) & \ldots & \sin(n_k\omega) \end{bmatrix} \\
\theta^T &= \begin{bmatrix} k_1 & k_2 & \ldots & k_{n_k+1} \end{bmatrix}.
\end{aligned}
$$

Denote $F_{nf}(e^{j\omega}) = F_r(\omega) - jF_i(\omega)$. It can now be computed that

$$F_{nf}(e^{j\omega}) K(e^{j\omega}) = \psi_r^T(\omega)\,\theta - j\psi_i^T(\omega)\,\theta,$$

where

$$
\begin{aligned}
\psi_r^T(\omega) &= F_r(\omega)\,\phi_r^T(\omega) - F_i(\omega)\,\phi_i^T(\omega) \\
\psi_i^T(\omega) &= F_r(\omega)\,\phi_i^T(\omega) + F_i(\omega)\,\phi_r^T(\omega).
\end{aligned}
$$

By computing the magnitude square of $1 - F_{nf}(e^{j\omega})K(e^{j\omega})$, we finally obtain the following quadratic constraint for (6.18):

$$\left|Q(e^{j\omega})\right|^2 = \theta^T \left[\psi_r(\omega)\,\psi_r^T(\omega) + \psi_i(\omega)\,\psi_i^T(\omega)\right]\theta - 2\psi_r^T(\omega)\,\theta + 1 \leq \delta^2(\omega).$$

Notice that $\psi_r(\omega)\,\psi_r^T(\omega) + \psi_i(\omega)\,\psi_i^T(\omega)$ is positive semi-definite. The above constraint is thus convex in θ.

Summarizing, we obtain the following equivalent form of Equations (6.15)-(6.17):

$$\begin{aligned}
&\min_{\theta,M,\gamma} : \gamma \\
&\text{s.t.: } \tilde{B}\theta = \tilde{A} \\
&\begin{bmatrix} A^TMA - M & A^TMB & C^T \\ B^TMA & B^TMB - \gamma I & D^T \\ C & D & -\gamma I \end{bmatrix} \preceq 0,\ M \succ 0,\ \gamma > 0 \\
&\theta^T \left[\psi_r(\omega)\psi_r^T(\omega) + \psi_i(\omega)\psi_i^T(\omega)\right]\theta - 2\psi_r^T(\omega)\theta + 1 \le \delta^2(\omega),\ \omega = \omega_1, \omega_2, \ldots
\end{aligned} \tag{6.19}$$

$$\begin{aligned}
&\min_{\theta,M,\gamma} :\gamma \\
&\text{s.t.: } \tilde{B}\theta = \tilde{A} \\
&\begin{bmatrix} A^TMA - M & A^TMB & C^T \\ B^TMA & B^TMB - \gamma I & D^T \\ C & D & -\gamma I \end{bmatrix} \preceq 0 \\
&M \succ 0 \\
&\gamma > 0 \\
&\theta^T \left[\psi_r(\omega)\psi_r^T(\omega) + \psi_i(\omega)\psi_i^T(\omega)\right]\theta \\
&\quad - 2\psi_r^T(\omega)\theta + 1 \le \delta^2(\omega),\ \omega = \omega_1, \omega_2, \ldots
\end{aligned}$$

where $\tilde{A}$ and $\tilde{B}$ are from (6.14).

As a final note, to make (6.19) linear also in the decision variable θ, we choose the controllable canonical form of $K(z)$:

$$\begin{aligned}
A &= \begin{bmatrix} 0_{n_k-1,1} & I_{n_k-1} \\ 0 & 0_{1,n_k-1} \end{bmatrix},\ B = \begin{bmatrix} 0_{n_k-1,1} \\ 1 \end{bmatrix} \\
C &= [k_2, \ldots, k_{n_k+1}],\ D = k_1.
\end{aligned}$$

The entire optimization problem is now convex, and can be efficiently solved using the interior-point method (see, e.g., [84]) in modern optimization.

Optimization Result

As an example, let $m = 2$ and

$$F_{nf}(z) = \frac{1 - 2\beta\cos\omega_0 z^{-1} + \beta^2 z^{-2}}{1 - 2\alpha\cos\omega_0 z^{-1} + \alpha^2 z^{-2}}$$

where $\omega_0 = 2\pi\Omega_{Hz}T_s$. Since $m = 2$, the order of $K(z)$ should be no smaller than one. Consider the case of $\Omega_{Hz} = 2000$, $T_s = 0.04$ ms, $\alpha = 0.9421$, and $\beta = 0.999$. Figure

6.9 demonstrates the magnitude responses of $Q(z)$ and $1 - z^{-m}Q(z)$ using the proposed optimization. Also plotted are the results if we ignore z^{-m} and directly assign a standard bandpass filter $Q(z) = 1 - F_{nf}(z)$. Notice first, from the second plot, that a standard bandpass filter does not create the desired magnitude response in $1 - z^{-m}Q(z)$. Instead there are undesired large gains at around 2100 Hz. On the contrary, the causal and optimal designs correctly introduce a sharp notch at the desired frequency. Second, compared to the causal solution, the optimal solution enforces reduced gains in $Q(z)$. In this example, the magnitude of $Q(z)$ is constrained to be no larger than -40dB at {0, 6000, 7500, 10000, 12500 (Nyquist frequency)} Hz.

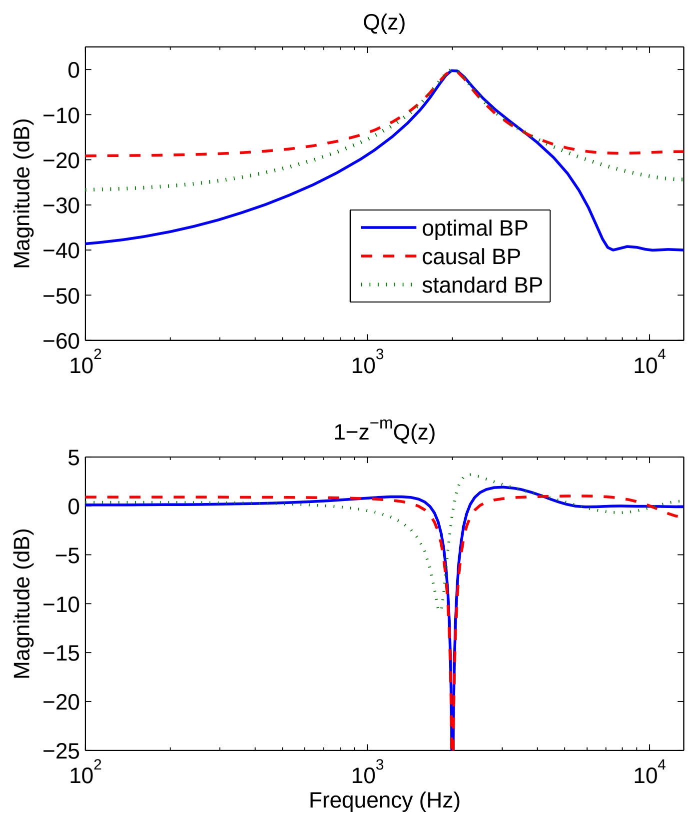


Figure 6.9: Q-design example based on convex optimization: BP denotes "bandpass"; the standard BP filter is obtained from $Q(z) = 1 - F_{nf}(z)$; the causal and optimal BP designs are from Section 6.4

Chapter 7

Stable Selective Model Inversion

Inverse-based design has numerous practical applications. Designing inverse models itself is an important control problem that has received much research attention (see, e.g., [85, 86, 87, 88]).

We have been assuming a stable $\hat{P}^{-1}(z)$ in the formulation of the pseudo Youla-Kucera parameterization. If $P(z)$ is a minimum-phase system, $P^{-1}(z)$ can directly be used.[1] For a practical sampled-data system, non-minimum phase zeros may occur in $P(z)$. In [89], it was proved that, as long as the sampling time is sufficiently short, all continuous systems with relative degree larger than or equal to two will have nonminimum-phase zeros in their zero-order-hold equivalent. Fractional delays also introduce unstable zeros [89, 90]. In particular, real unstable zeros will appear at high frequencies after fast sampling.

One fundamental reason for the feasibility of *stably* inverting (at least partially) the plant dynamics, is that we can (and usually need to) accept imperfections in practical inverse design. To begin with, high-frequency uncertainties always exist in actual systems. Perfect model inversion is at first place impractical. Also, practical feedback design has an effective servo bandwidth, above which the control efforts are always suggested to be kept small [61] and perfect model inversion is hence not necessary. As long as nonminimum-phase zeros do not occur at the desired servo-enhancement regions, a stable model inversion should be feasible.

7.1 Unstable-zero Modulation

We provide first a simple ad-hoc method to get some intuitions of the solution.

It is natural for motion-control systems to have integrator-type plant dynamics [91], as motors commonly take force/torque as the input and generate (angular) position or velocity as the output. For discrete-time models of such systems, high-frequency unstable zeros usually occur on the real axis on the left half side of the complex plane. Such unstable zeros are less challenging compared to those at low frequencies. Intuitively, if the zeros of the

[1] For instance, piezoelectric actuators have flat nominal magnitude response, and is minimum-phase.

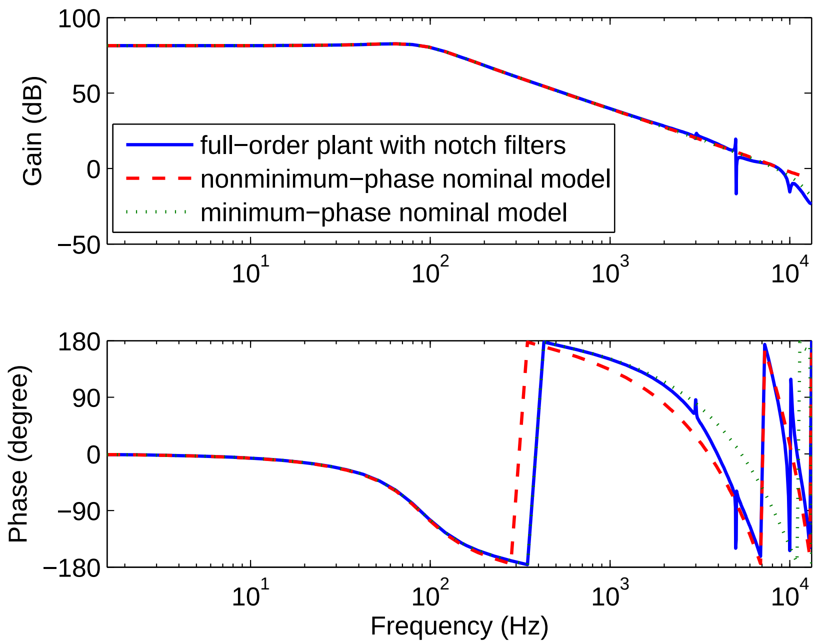


Figure 7.1: Nominal model inversion by zero modulation

continuous-time plant are all on the left half plane, approximate discrete-time stable system inversion should not be impossible.

Consider

$$P_d(z) = z^{-3}\frac{1.447663z^2 + 3.684538z + 0.183621}{z^2 - 1.978354z + 0.978808} \tag{7.1}$$

which is the zero-order-hold (sampled at 26400 Hz) equivalent of the continuous-time transfer function

$$P(s) = e^{-10^{-5}s}\frac{3.74488 \times 10^9}{s^2 + 565.487s + 3.19775 \times 10^5} \tag{7.2}$$

with two steps of additional delays. Equation (7.2) is an initial guess (by modeling the physics of the motor and actuator) of the nominal model of the 14-order HDD system in Figure 3.9 on page 32. Several notch filters have been designed to attenuate the resonances at high frequencies. The solid line in Figure 7.1 shows the frequency response of the actual plant with notch filters. The dashed line is the response of (7.1), which overall matches well with the actual system.

One zero of (7.1) is unstable, at around -2.5, as shown in Figure 7.2. Factorizing yields

$$P_d(z) = z^{-m}\frac{B^*(z)}{A(z)}\left(1 - z^{-1}\gamma\right)$$

where γ is the zero outside the unit circle.

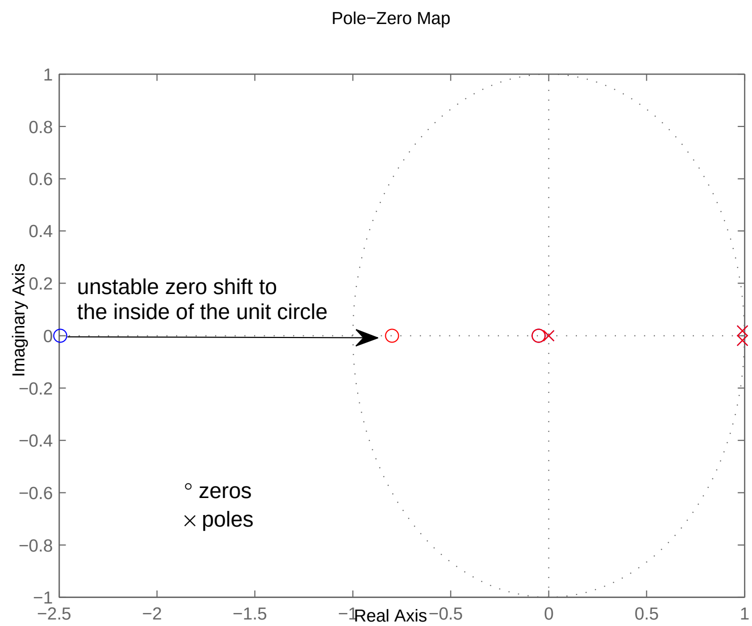


Figure 7.2: Pole-zero plots of the nominal plant models

To show the impact of the term $1 - z^{-1}\gamma$ on the frequency response, we note first, that the servo bandwidth for the system (single-stage HDD) is usually around 1300 Hz while the sampling frequency is 26400 Hz. Consider $e^{j\omega} - \gamma$, the contribution of the unstable zero in the frequency response. At 1300 Hz, $\omega = 2\pi \times 1300/26400 \approx 0.309$ rad $= 17.72$ deg. Hence, between 0 Hz and 1300 Hz (which is the main "performance region"), $e^{j\omega}$ varies only in a small arc on the unit circle, yielding very mild changes to the value of the vector $e^{j\omega} - \gamma$. Replacing the unstable zero at γ with a stable one near $(-1, 0)$ as shown in Figure 7.2, and after normalization,[2] we obtain

$$\hat{P}(z) = z^{-3} \frac{z^2 + 0.850852z + 0.040681}{0.355831z^2 - 0.703959z + 0.348290}$$

whose frequency response is shown in the dotted (green) line in Figure 7.1. We observe that $\hat{P}(z)$ matches well with the actual plant dynamics in the solid line. Actually, below 3000 Hz, which is sufficiently high for major servo-enhancement schemes, the dotted line appears to be a better fit compared to the dashed (red) line. For simple systems with few zeros, such unstable-zero modulation is seen to maintain the essential plant information.

The next example provides the modeling of the top stage in the wafer scanner discussed

[2] The replacement of the unstable zero will change the low-frequency gains. Normalization is carried out to counteract this gain change.

in Section 3.1 on page 26. From physics, the nominal plant model is

$$P(s) = \frac{1}{ms^2 + bs} = \frac{1}{0.2556s^2 + 0.279s}$$

whose zero-order-hold equivalent (sampled at 2500 Hz) is

$$\mathrm{ZOH}(P) = \frac{3.12943 \times 10^{-7}z + 3.12897 \times 10^{-7}}{z^2 - 1.99956z + 0.99956} \approx \frac{3.129 \times 10^{-7}(z+1)}{(z-1)^2}.$$

Comparing the frequency response with the experimental data, we found that there is one additional input delay, yielding

$$P(z) = \frac{3.129 \times 10^{-7}(z+1)}{z(z-1)^2}. \tag{7.3}$$

The transfer function in (7.3) has a zero on the unit circle. Shifting it to be a strictly minimum-phase zero, and normalizing the gain such that the magnitude at 100 Hz is perfectly preserved, we have

$$\hat{P}(z) = \frac{3.4766 \times 10^{-7}(z+0.8)}{z(z-1)^2} \tag{7.4}$$

or, equivalently

$$\hat{P}(z) = z^{-2}\frac{3.4766 \times 10^{-7}(1+0.8z^{-1})}{(1-z^{-1})^2}.$$

Figure 7.3 provides the measured frequency response and the nominal fitting by (7.4). Figure 7.4 shows the frequency responses of (7.3) and (7.4). We observe that the difference in Figure 7.4 is quite small and the model matching at low frequencies is quite accurate.

The automation of the above modulation process, and the extension to general systems with possibly more poles and zeros, are precisely the focus of the next section.

7.2 H_∞-based Optimal Design

From the above discussions, we now formally propose to perform inverse design at selective frequency regions. Denote $\hat{P}^{-1}(z)$ as the *nominal* stable inverse for $P(z)$. At frequencies where there are no nonminimum-phase zeros and no large model uncertainties (usually low and middle frequencies), we enforce correct model matching between $\hat{P}(z)$ and $P(z)$; otherwise (commonly at high frequencies), we construct constrains such that $\hat{P}^{-1}(z)$ has a limited magnitude response.

We use an H_∞ formulation to achieve this selective model inversion.

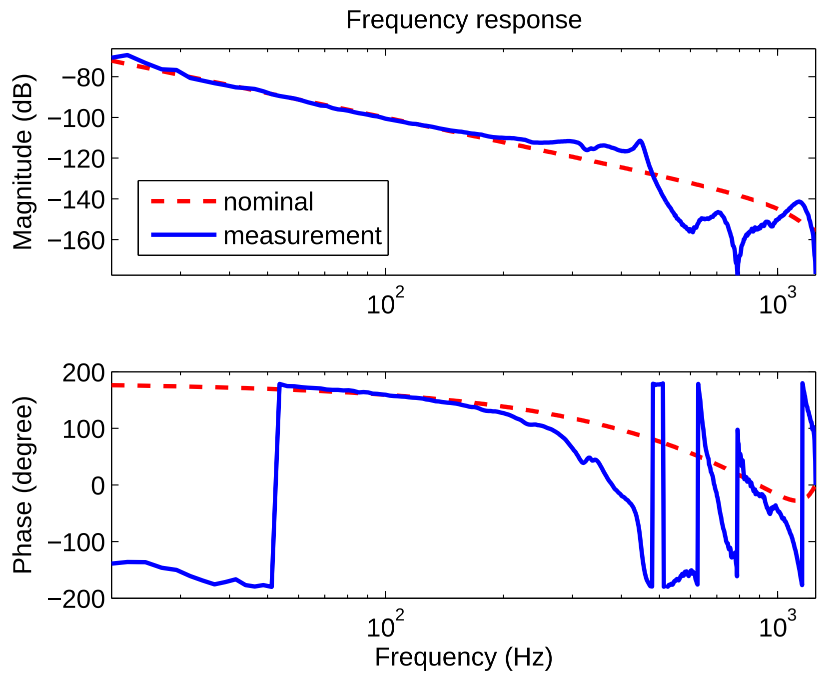


Figure 7.3: Frequency responses of the reticle stage in the wafer scanner: measurement and nominal model

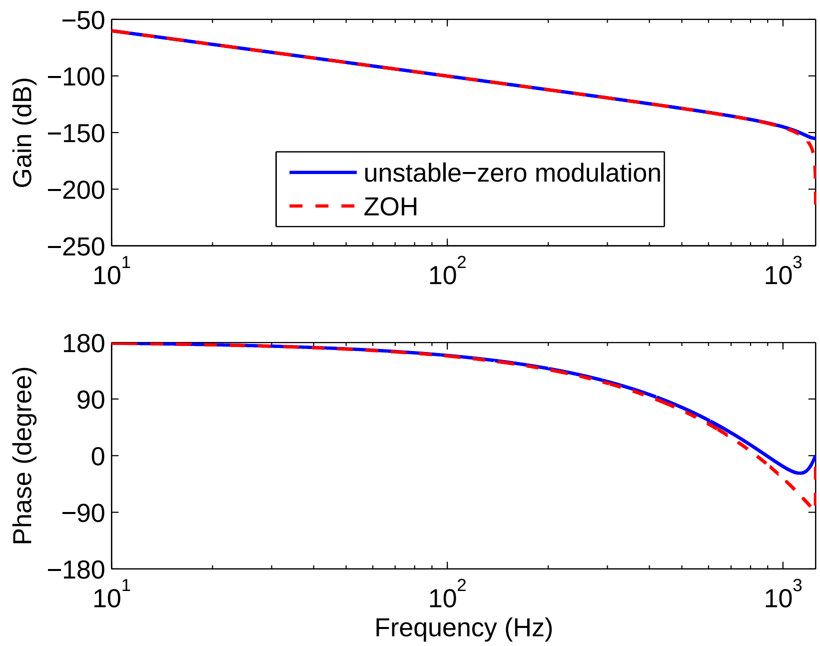


Figure 7.4: Comparison of nominal models for the reticle stage

Let $\mathcal{S}$ denote the set of all stable discrete-time rational transfer functions. We search among $\mathcal{S}$ to find

$$M(z) = z^{-m}\hat{P}^{-1}(z) \tag{7.5}$$

such that the following three criteria are satisfied:

(i) $M(z)$ is realizable/proper. As it is commonly desired to have the minimal amount of delays in $M(z)$, m in (7.5) can take the value of the relative degree of $\hat{P}^{-1}(z)$;

(ii) *model matching* to achieve

$$\min ||W_1(z)\left(M(z)P(z) - z^{-m}\right)||_\infty.$$

Namely, we minimize the maximum magnitude of the model mismatch

$$M(z)P(z) - z^{-m},$$

weighted by the filter $W_1(z)$. The ideal solution, if $P^{-1}(z)$ is stable, is simply $M(z) = z^{-m}P^{-1}(z)$. The weighting function determines the interested region where we would like to have good model accuracy.

(iii) *gain constraint*: as the inverse filter is used for signal processing in feedforward and feedback configurations, we should be careful to avoid noise amplification at high frequencies. Consider

$$\min ||W_2(z)M(z)P(z)||_\infty$$

where the magnitude of $M(z)P(z)$ is scaled by the weight $W_2(z)$. The optimal solution for this part alone would be that $M(z) = 0$, i.e., $M(z)$ will not amplify any of its input components.

Combining the three design goals, we get

$$\min_{M(z)\in\mathcal{S}} \left\| \begin{bmatrix} W_1(z)\left(M(z)P(z) - z^{-m}\right) \\ W_2(z)M(z)P(z) \end{bmatrix} \right\|_\infty. \tag{7.6}$$

The optimization in (7.6) finds the optimal inverse that preserves accurate model information in the frequency region specified by $W_1(z)$, and in the meantime penalizes excessive high gains of $M(z)$ at frequencies where $W_2(z)$ has high magnitudes. As mentioned before, typically $W_1(z)$ is a low-pass filter and $W_2(z)$ is a high-pass filter.

By the formulation of the problem, (7.6) falls into the framework of H_∞ control, and can be efficiently solved in the robust control toolbox in MATLAB. The solution exists as long as $P(z)$, $W_1(z)$, and $W_2(z)$ are stable. The order of $M(z)$ will be the sum of the orders of $W_1(z)$, $W_2(z)$, z^{-m} and $P(z)$. After solving (7.6), standard model-reduction techniques can be applied to obtain a lower-order solution of $M(z)$.

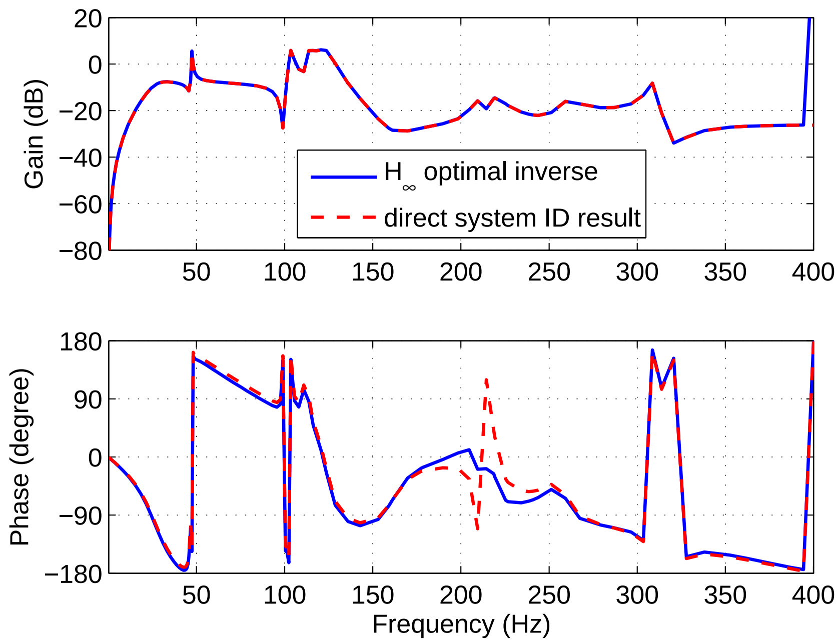


Figure 7.5: Frequency response of the plant and its minimum-phase approximation

As an example, we apply the above formulation to the active suspension system in Section 3.3 on page 34. The dashed line in Figure 7.5 is the frequency response of $P(z)$ from standard system identification. The solid line is the response of the solved $\hat{P}(z)$, where we have first obtained $M(z) = z^{-m}\hat{P}^{-1}(z)$ from (7.6), performed model reduction to reduce the order of $M(z)$ to 23, and then let

$$\hat{P}(z) = z^{-m}M^{-1}(z).$$

We can see that the optimal solution matches well with the actual plant dynamics, and moreover, $\hat{P}^{-1}(z)$ is stable although $P(z)$ itself is of nonminimum phase.

Remark. We intentionally used a high-order (and challenging) example in Figure 7.5. For many motion control problems, the plants have much simplified dynamics by the nature of hardware design.

Chapter 8

Enhanced Repetitive Control

8.1 Introduction

In this chapter, we provide one particular application of the pseudo Youla-Kucera parameterization.

Repetitive control (RC) is a well-known servo design tool for systems that are subjected to periodic disturbances/references. It implements an internal model [92] $1/(1-z^{-N})$ (N is the period of the disturbance/reference), or $1/(1-e^{-T_p s})$ in the continuous-time case (T_p denotes the period), into a feedback system, such that errors in the previous repetition can be used to improve the current regulation/tracking control. Distinguished by its high performance as well as the simple design and implementation criteria, ever since its introduction [93, 94, 95], RC has attracted a great amount of research efforts [96, 97, 94, 95]. Its versatility has been tested in various practical applications, including but not limited to: track-following in magnetic and optical disk drives [98, 99, 100, 101], robot arm control [102], and regulation control in vehicles [103]. For more complete lists of applications, readers can refer to the survey papers [96, 97].

The configuration of the internal model and its interaction with the feedback system vary in literature. The continuous-time RC design mainly applies a series or parallel plug-in configuration [99, 104, 94, 105, 106, 107]. The prototype RC [95, 108] applies the Zero-Phase-Error-Tracking [85] idea and directly cascades a robust version of $1/(1-z^{-N})$ into the open-loop transfer function. Additionally there are plug-in configurations of discrete RC design, among which [109, 110] applied optimization techniques with an extended high-order internal model.

Ultimately, a generalized version of $1-z^{-N}$ or $1-e^{-T_p s}$ is absorbed into the denominator of the overall feedback controller, therefore creating high-gain control at the repetitive frequencies (frequencies of the roots of $1-z^{-N}=0$ or $1-e^{-T_p s}=0$).[1] From Bode's Integral Theorem, enhanced servo performance at certain frequencies commonly results in deterio-

[1]i.e., kF_s/N Hz for $1-z^{-N}=0$ and k/T_p Hz for $1-e^{-T_p s}=0$. Here $k=0,1,2,...$ and F_s is the sampling frequency. F_s/N and $1/T_p$ are called the fundamental frequencies.

rated loop shapes at other frequencies. This fundamental limitation, reflected in repetitive control, is the comb-like magnitude response in the closed-loop sensitivity function, along with undesired gain amplifications at frequencies other than the comb centers (see some examples in [99, 104, 106, 107, 109, 110]). The problem is more significant if there are large non-periodic components in the disturbance (e.g., in hard disk drive systems [70]).

Relaxing the previous performance limitations, in this chapter we discuss a new structural RC design with improved loop-shaping properties. Instead of using the full information of the previous errors, we use the pseudo YK parameterization to extract only the repetitive errors in feedback control. In the frequency domain, this corresponds to a series-parallel implementation of the internal model, with direct control of the comb-like loop shape, leading to greatly reduced gain amplifications at the non-repetitive frequencies. An additional benefit of the reduced gain amplification is that the proposed design shows increased ability to reject repetitive errors at high frequencies. Finally, we discuss the control of the transient response for obtaining a smoother and accelerated transient.

8.2 Repetitive Loop Shaping

Recall the pseudo Youla-Kucera parameterization structure in Figure 8.1. From the class of stabilizing controllers, we seek for one advantageous $Q(z)$ to obtain enhanced repetitive control (ERC). The plug-in repetitive signal generator provides the compensation signal $c(k)$ in Figure 8.1. In the case of regulation control, $r(k) = 0$; we aim to have $c(k)$ cancel the *periodic components* in $d(k)$. In the tracking-control case, $c(k)$ functions to reduce the tracking error between $y(k)$ and $r(k)$.

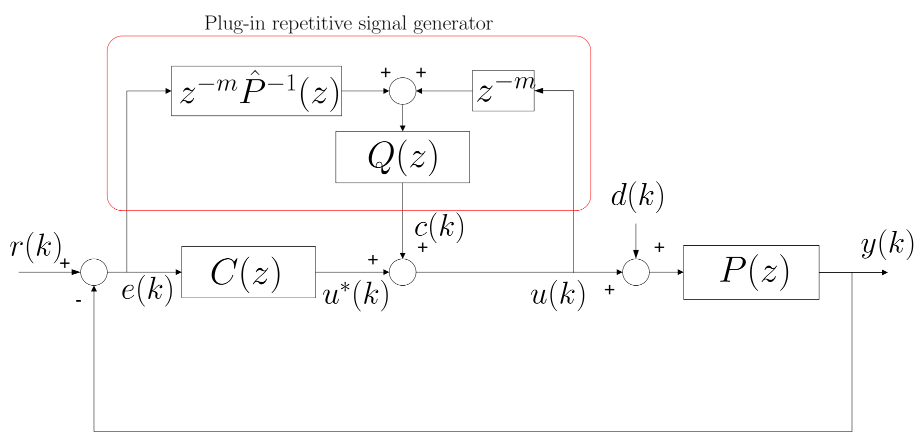


Figure 8.1: Block diagram of the proposed repetitive control scheme

Remark 8.1. If the $z^{-m}\hat{P}^{-1}(z)$ block is removed from Figure 8.1 and $Q(z)$ is set to $z^{-(N-m)}$, the open-loop transfer function becomes $P(z)\frac{1}{1-z^{-N}}C(z)$, and the proposed compensator reduces to an ideal-case plug-in repetitive controller that is similar to prior literature.

From Chapter 4 we have

$$S(z) = \frac{1 - z^{-m}Q(z)}{1 + P(z)\,C(z) + z^{-m}Q(z)\,(\hat{P}^{-1}(z)\,P(z) - 1)} \tag{8.1}$$

and the closed-loop transfer functions from $d(k)$ and $r(k)$ to $e(k)$ are respectively given by

$$G_{ed}(z) = -P(z)S(z) \tag{8.2}$$

$$G_{er}(z) = S(z). \tag{8.3}$$

Ultimately, we wish to obtain a Q filter that achieves Figure 8.2.

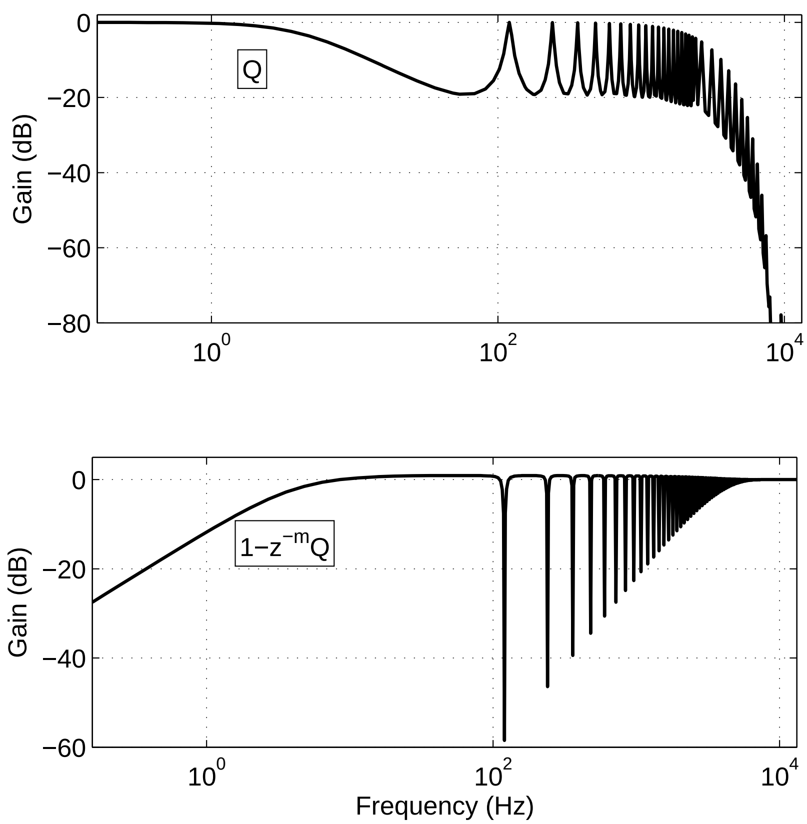


Figure 8.2: Intuition for Q-filter design in ERC

In this way, the selective unity gain in $Q(z)$ provides the desired attenuation in $1 - z^{-m}Q(z)$, and the high-frequency low gain of $Q(z)$ keeps the influence of the uncertainty term $z^{-m}Q(z)\,(\hat{P}^{-1}(z)\,P(z) - 1)$ small in (8.1), so that

$$S(z) \approx \frac{1 - z^{-m}Q(z)}{1 + P(z)C(z)} \tag{8.4}$$

and the actual sensitivity function has the characteristics shown in Figure 8.3.

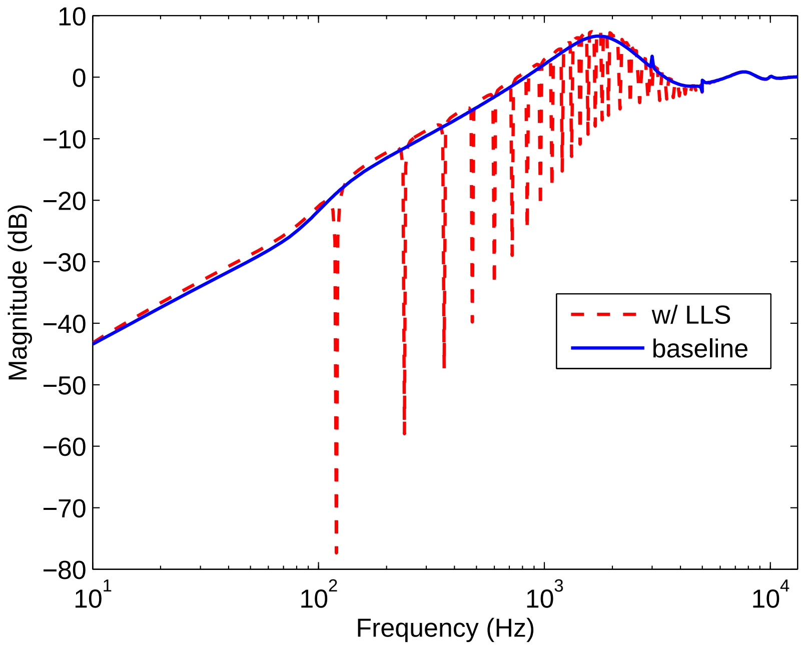


Figure 8.3: Magnitude response of the sensitivity function in ERC

8.3 Ideal-case Q design

To derive the detailed mathematics, assume the disturbance contains only repetitive components that asymptotically satisfy the internal model

$$(1 - q^{-N})d(k) = 0 \tag{8.5}$$

or, in the tracking-control case,

$$(1 - q^{-N})r(k) = 0. \tag{8.6}$$

From (8.2) and (8.3), to reject $d(k)$ or track $r(k)$, it suffices to have $S(z)d(k)$ and $S(z)r(k)$ converge asymptotically to zero. By combining (8.1), (8.5) and (8.6), one may notice that this sufficient condition is achieved if $1 - z^{-m}Q(z)$ contains the term $1 - z^{-N}$.

Assigning $Q(z) = z^{-(N-m)}$ is one way which gives a scheme similar to conventional RC. We adopt our usual IIR design concept, and propose

$$Q(z) \triangleq \frac{B_Q(z)}{A_Q(z)}$$

with

$$A_Q(z) - z^{-m}B_Q(z) = 1 - z^{-N}. \tag{8.7}$$

Designing

$$A_Q(z) = 1 - \alpha^N z^{-N} \tag{8.8}$$

and solving (8.7) yield

$$B_Q(z) = (1 - \alpha^N)z^{-(N-m)} \tag{8.9}$$

$$1 - z^{-m}Q(z) = \frac{1 - z^{-N}}{1 - \alpha^N z^{-N}}. \tag{8.10}$$

Hence we have achieved to include $1 - z^{-N}$ in the numerator of $1 - z^{-m}Q(z)$, with an additional tunable module $1-\alpha^N z^{-N}$. Here $\alpha(\in [0,1])$ is the ratio between magnitudes of the poles and the zeros of $1 - z^{-m}Q(z)$.[2] If $\alpha = 0$, $Q(z)$ becomes an FIR filter ($Q(z) = z^{-N+m}$) and ERC generates a loop shape that is similar to prior publications. This will be discussed in more details in Section 8.8. On the other hand, $\alpha = 1$ cuts off the repetitive compensation. When $\alpha \in [0, 1)$, the loop shape can be flexibly designed. For instance, let $N = 10$, $m = 1$, and assume a sampling frequency of 26400 Hz. Increasing α from 0 to 0.99 yields the magnitude responses in Figure 8.4. From the top plot, we observe that as α increases towards 1 (while still satisfying $\alpha \in [0, 1)$), $1 - z^{-m}Q(z)$ has a sharper comb-like magnitude response and a smaller H_∞ norm. Correspondingly in the bottom plot, $Q(z)$ behaves as a sharper spectral-selection filter to preserve only the repetitive components. Specifically, if $\alpha = 0$, $Q(z)$ has a magnitude response valued always at 1, and both the repetitive and the non-repetitive error components are directly used for feedback compensation; in the mean time, the maximum magnitude of $1 - z^{-m}Q(z)$ equals $\left\|1 - z^{-m}z^{-(N-m)}\right\|_\infty = 2$, i.e., disturbances at the corresponding frequencies get amplified by 100%. One can observe that the design of (8.8) and the introduction of α have provided an additional degree of freedom for repetitive loop shaping, enabling the improvement in Figure 8.4, from the solid lines to the dotted lines. Additionally we have the following theorem:

Theorem 8.1. *When $P(z) = \hat{P}(z)$, conventional RC amplifies non-repetitive disturbances by 100% in the worst case. The worst-case amplification is $\left(2/(1 + \alpha^N) - 1\right) \times 100\%$ in the proposed scheme. The maximum amplification occurs to the disturbance components at the frequencies $(2k + 1) / (2T_s N)$ Hz, $k = 0, 1, \ldots$.*

[2]For this analytical reason, we used α^N instead of defining $A_Q(z) = 1 - \beta z^{-N}$ in (8.8), to avoid the appearance of (numerically more fragile) $\sqrt[N]{\cdot}$ in our discussion. Yet for practical implementation, $\beta \triangleq \alpha^N$ can directly be used without the need of computing α^N online.

Proof. The maximum disturbance amplification corresponds to the maximum magnitude response of $1 - z^{-m}Q(z)$ in (8.1). For (8.10), the *squared* magnitude response is

$$\left|Q\left(e^{j\omega}\right)\right|^2 = Q\left(e^{-j\omega}\right)Q\left(e^{j\omega}\right) \tag{8.11}$$

$$= \frac{1-e^{-j\omega N}}{1-\alpha^N e^{-j\omega N}} \times \frac{1-e^{j\omega N}}{1-\alpha^N e^{j\omega N}} = \frac{1-\cos\left(\omega N\right)}{\frac{1+\alpha^{2N}}{2} - \alpha^N \cos\left(\omega N\right)} \tag{8.12}$$

where $\omega = 2\pi\Omega_{Hz}T_s$.

Noting that $\cos\left(\omega N\right) \in [-1, 1]$, we need only consider the behavior of the function

$$f\left(x\right) = \frac{1-x}{\frac{1+\alpha^{2N}}{2} - \alpha^N x},\ x \in \left[-1, 1\right].$$

The derivative of $f\left(x\right)$ is

$$f^{'}\left(x\right) = \frac{-\frac{1}{2}(1-\alpha^N)^2}{\left(\frac{1+\alpha^{2N}}{2} - \alpha^N x\right)^2},\ x \in \left[-1, 1\right].$$

It is straightforward to see that $f^{'}\left(x\right)$ monotonically decreases as x increases from -1 to 1. Thus, $\min\left\{f\left(x\right)\right\}$ and $\max\left\{f\left(x\right)\right\}$ are attained respectively at $x = 1$ and $x = -1$, with $\min\left\{f\left(x\right)\right\} = 0$ and

$$||1 - z^{-m}Q(z)||_\infty^2 = \max\left\{f\left(x\right)\right\} = \left(\frac{2}{1+\alpha^N}\right)^2. \tag{8.13}$$

Taking the square root of (8.13) gives the maximum amplification gain. Solving $x = \cos\left(2\pi\Omega_{Hz}T_sN\right) = -1$ gives that the maximum occurs at the frequencies $\Omega_{Hz} = (2k + 1)/(2T_sN)$, $k = 0, 1, \ldots$. In the special case of $\alpha = 0$ (conventional RC)

$$||1 - z^{-m}Q(z)||_\infty = \sqrt{\max\left\{f\left(x\right)\right\}|_{\alpha=0}} = 2$$

The proof is done by computing the relative amplification $(||1 - z^{-m}Q||_\infty - 1) \times 100\%$. $\square$

8.4 Robustness and Implementation of $Q(z)$

It is necessary to incorporate a low-pass filter in $Q(z)$ to make the influence of the uncertainty term $z^{-m}Q(z)\left[P(z)\hat{P}^{-1}(z) - 1\right]$ small in (8.1). In the context of repetitive control, it is additionally possible (and recommended) to apply a zero-phase low-pass filter. One simple and flexible construction is as follows. Define first the following zero-phase low-pass filter as a base structure

$$q_0(z, z^{-1}) = \frac{(1+z^{-1})^{n_0}(1+z)^{n_0}}{4^{n_0}} \tag{8.14}$$

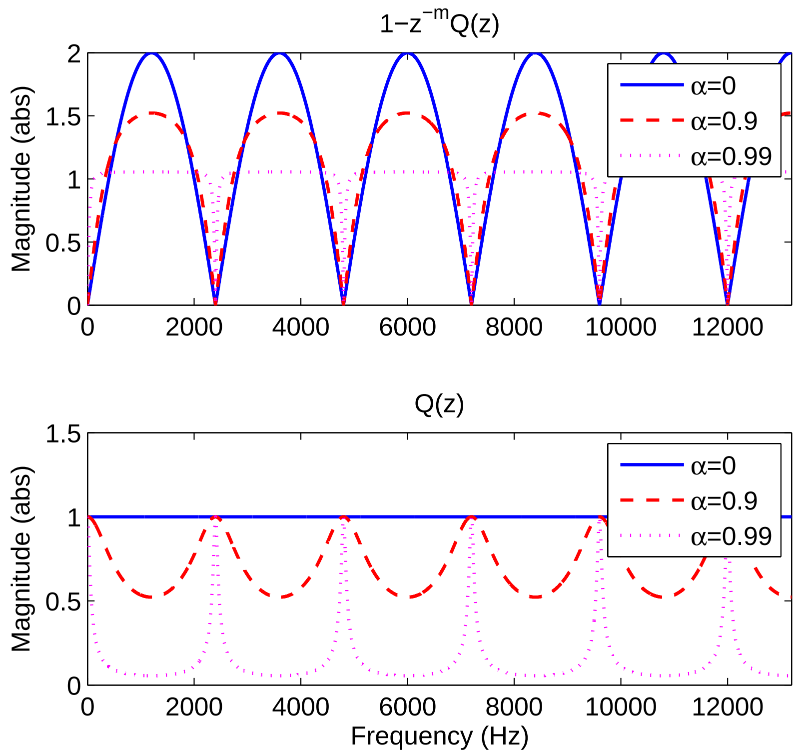


Figure 8.4: Magnitude responses of $1 - z^{-m}Q(z)$ and $Q(z)$ with different values of α

where $2n_0$ is the number of placed zeros at the Nyquist frequency. To have additional freedom on the cut-off frequency, we can add extra zero-phase pairs given by

$$q_i(z, z^{-1}) = q_i(z^{-1})q_i(z), \tag{8.15}$$

$$q_i(z) = \frac{1 - 2\cos(\omega_i T_s)z^{-1} + z^{-2}}{2 - 2\cos(\omega_i T_s)}. \tag{8.16}$$

Here i is the index number; ω_i is in rad/sec; $q_i(z^{-1})$ is obtained by replacing every z in $q_i(z)$ by z^{-1}. The filter $q_i(z, z^{-1})$ places four zeros at $e^{\pm\omega_i T_s}$ to remove the frequency components at ω_i's, and is normalized by $(2-2\cos(\omega_i T_s))^2$ to have a unity DC gain. The zero-phase property is preserved since the frequency responses of $q_i(z^{-1})$ and $q_i(z)$ are complex conjugates of each other.

Defining

$$q(z, z^{-1}) = \prod_{j=0}^{n_q} q_j(z, z^{-1}),$$

where n_q is the number of zero-phase pairs in (8.15), we can now construct the practical version of $Q(z)$:

$$\boxed{Q(z) = \frac{(1-\alpha^N)z^{-(N-m-n_q)}}{1-\alpha^N z^{-N}} z^{-n_q} q(z, z^{-1})} \tag{8.17}$$

where n_q is the highest order of z in $q(z, z^{-1})$ (so that $z^{-n_q}q(z, z^{-1})$ is realizable). It can be noted that $Q(z)$ is causal as long as $N - m - n_q \geq 0$. Figure 8.5 presents one realization of (8.17), with N memory elements for the repetitive signal generator. The bandwidth of the low-pass filter $q(z, z^{-1})$ can roughly be tuned by comparing the magnitude responses of $\left[P(z)\hat{P}^{-1}(z) - 1\right] Q(z)$ and $1+P(z)C(z)$. Note that $Q\,(z)$ is clearly stable in itself. From the stability condition in general pseudo Youla-Kucera parameterization, the closed-loop system has guaranteed stability if $\hat{P}\,(z) = P\,(z)$ and $\hat{P}\,(z)^{-1}$ is stable.

When the plant is perturbed to be $\tilde{P}(z) = P(z)(1+\Delta(z))$ (assume the uncertainty $\Delta(z)$ is stable and has a bounded H_∞ norm), we have the robust-stability condition:

$$\|\Delta(z)T(z)\|_\infty < 1 \tag{8.18}$$

where $T\,(z)$ is the nominal complementary sensitivity function (see Section 4.3 on page 53).

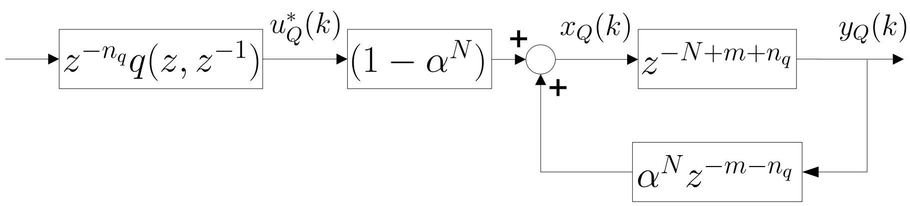


Figure 8.5: Implementation of the Q filter

8.5 Transient Response and Algorithm Implementation

With the plug-in compensator, a new feedback system is formed. The plug-in repetitive controller may be turned on or off depending on the presence of repetitive disturbances. Although the two closed loops are designed to be asymptotically stable, switching between the two stabilizing controllers in general does not yield smooth response [111]. This aspect should be carefully recognized for plug-in repetitive control, as N—the period of the repetitive disturbance/reference—can be large in practice. We note that Figure 8.5 has the following state-space realization

$$x_Q(k) = (1-\alpha^N)u_Q^*(k) + \alpha^N x_Q(k-N)$$
$$y_Q(k) = x_Q(k-N+m+n_q)$$

where $u_Q^*(k),\ y_Q(k) \in \mathbb{R}$, and $x_Q(k) \in \mathbb{R}^N$. As N may be large, α^N can be quite small. When $x_Q(k)$ is initialized to zero, the first $N-m-n_q$ values of $y_Q(k)$ equal zero. Starting from the time instant $N-m-n_q+1$, $y_Q(N-m-n_q+i) = (1-\alpha^N)u_Q^*(i)$ for $i \in [1, N]$. At this first period of actual compensation, depending on the baseline closed-loop dynamics, the impulse of $u_Q^*(k)$ can create high-amplitude transient response in the error signal. Additionally, all the information in $u_Q^*(k)$, including the non-repetitive components, are fed back by the compensation signal $c(k)$ in Figure 4.6, yielding mismatched cancellation for the non-periodic errors.

To reduce the possible overshoot and amplification of non-repetitive components, we can apply a time-varying α for transient improvement. It is proposed to initialize α at 1, and gradually reduce it to a designed value α_{end} (from steady-state analysis), following the decay rule

$$\alpha(k+1) = \alpha_{end} - (\alpha_{end} - \alpha(k))\alpha_{rate}, \tag{8.19}$$

with $\alpha(0) = 1$ and the decay rate $\alpha_{rate} \in (0, 1)$. Notice that when $\alpha = 1$, the Q filter is essentially turned off in (8.17). By the above construction, at the first period of compensation, u_Q^* is gradually (weighted by $1 - \alpha^N$) released to y_Q.

As for the settling time of the Q filter, the transient duration is determined by the pole location of the filter. Let n_t denote the number of periods for the impulse response of $Q(z)$ to reduce to less than 36.8% ($\approx e^{-1}$) of its peak value. From (8.17), this time constant is determined by $(\alpha^N)^{n_t} = e^{-1}$, i.e.,

$$n_t = \frac{-1}{\log \alpha^N}. \tag{8.20}$$

Here we allow non-integer value of n_t (e.g., $n_t = 0.5$ means that it takes half the time of a period to settle).

From (8.20), the smaller the term α^N, the shorter the settling time. In the case that $\alpha = 0$, $\lim_{\alpha \to 0}(n_t) = 0$.

Combining the above discussion with that of Figure 8.4, we can obtain in Table 8.1, the influence of α on various closed-loop properties. Notice the two conflicting objectives of maintaining (a) short transient duration and (b) small transient overshoot as well as good steady-state performance. Initializing α at 1 keeps the transient smooth and gradually reducing it afterward helps to accelerate the transient. Yet to maintain the steady-state performance, the final value of α may be required to be not too small. A slightly more complicated design of α is to first reduce it from 1 to a middle value α_{mid} and then increase it to a final α_{end} (see Figure 8.11).

In summary, the following design procedures are suggested for implementing the proposed algorithm:

1. analyze the plant; obtain z^{-m} and $\hat{P}^{-1}(z)$.

2. *design for steady-state performance according to Sections* 8.3 *and 8.4*: obtain (8.17), check the frequency responses of $1 - z^{-m}Q(z)$ and $S(z)$ [from (8.10) and (8.1) respectively]; compute the maximum amplification from Theorem 8.1. Here it is suggested to

start with an α that is close to unity (this gives smaller amplification of non-repetitive errors), and alter the value if stability or the desired performance metric is not reached.

3. *transient improvement*: compute (8.20) and simulate the time-domain closed-loop response–if large overshoot occurs, consider the time-varying α and initialize it at 1 as discussed in this section; if transient is excessively long, choose an intermediate value for α that is smaller than its steady-state value; keep the final value of α the same as the one designed in step 2).

Table 8.1: Influence of α on the transient and the steady-state performance

value of $\alpha(\in [0,1])$	steady-state performance	transient overshoot	transient duration
large	small amplification of non-repetitive components	small	long
small	converse of the above	possibly large	short

8.6 Application: Repeatable-runout Rejection on a Hard Disk Drive

This section provides a design example in the track-following control of the single-stage HDD system discussed in Section 3.2 on page 30. In this regulation-control example, the period of the repeatable disturbance is thus $N = 220$, at a fundamental frequency of $7200/60 = 120$ Hz. The baseline controller is a PID controller with several notch filters. The resulting baseline feedback system has a 5.45-dB gain margin, a 38.2-deg phase margin, and a 1.19-kHz open-loop servo bandwidth.

In the ERC design, we model $P(z)$ to contain the plant as well as the notch filters. The frequency responses of $P(z)$ and $\hat{P}(z)$ ($m = 2$ in this example) have been shown in Figure 7.1 on page 106. Since modeling errors appear after around 2 kHz, the zero-phase low-pass filter in Section 8.4 is designed to have a cut-off frequency of 2025 Hz, with $n_0 = 1$ in (8.14); $\omega_1 = 2\pi \times 122000$ rad/sec and $\omega_2 = 2\pi \times 8400$ rad/sec in (8.16). In view of the large value of N, α is designed to be 0.999 to achieve good steady-state performance. Correspondingly, α^N becomes 0.8024. α^N is directly implemented instead of α.

The magnitude responses of $Q(z)$ and $1-z^{-m}Q(z)$ have been shown in Figure 8.2. Notice the repetitive spectral-selection property (at multiples of the fundamental frequency 120 Hz) in $Q(z)$. This indicates that ERC only "observes" the periodic components and filters out the non-repetitive noise in the disturbance estimation.[3] For robustness, the zero-phase low-pass filter keeps the Q-filter gain small at high frequencies, yielding the gradual reduction of compensation capacity at high frequencies in $1 - z^{-m}Q(z)$. The magnitude responses of the actual closed-loop sensitivity functions are shown in Figure 8.3. We can see that the designed loop shape in $1 - z^{-m}Q(z)$ is successfully transformed to the closed-loop system, and that the loop shape at the non-repetitive frequencies is preserved in Figure 8.3. The loop-shaping results can be compared with those in [107, 109, 106].

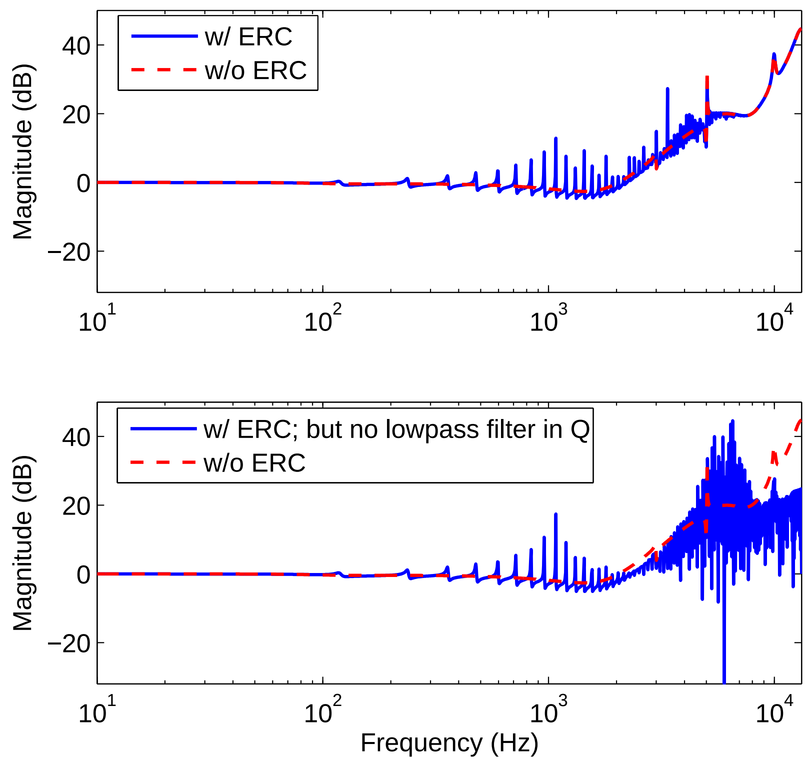


Figure 8.6: Magnitude responses of $1/T(z)$ for ERC robust stability

Figure 8.6 shows the magnitude responses of $1/T(z)$, the inverse of the complementary sensitivity function. From (8.18), in order to preserve the robust stability, magnitude of the plant uncertainty has to be lower than that of $1/T(z)$ at all frequencies. From the top plot, we observe that the introduction of ERC largely preserves the robust stability bounds (compared to the baseline closed-loop system), especially in the high-frequency region. The minimal value of the solid line is -4.7dB (0.582 in absolute value) at 1327 Hz, i.e., the plant should not have an uncertainty that is larger than 58.2% at this frequency. The necessity of

[3]Note that a constant disturbance is also repetitive, and observed by $Q(z)$.

the zero-phase low-pass filter $q(z, z^{-1})$ is evident from the bottom plot. Without $q(z, z^{-1})$, three percent (-31dB) of model uncertainty at 6000 Hz will drive the system unstable.

Simulation is conducted by applying a full set of practical disturbances that includes the disk-flutter disturbance, the sensor noise, the repeatable runout (RRO), and the input force disturbance. Figure 8.7 presents the spectra of the position error signals (PES) (in the steady state) with and without ERC. One can remark that the repetitive errors below 2000 Hz are successfully removed,[4] and that amplification of other errors is visually not distinguishable. As a performance metric in HDD industry, the 3σ (σ denotes the standard deviation) value of the PES reduces from 10.77% Track Pitch (TP) to 9.30% TP, indicating a 13.6 percent improvement.

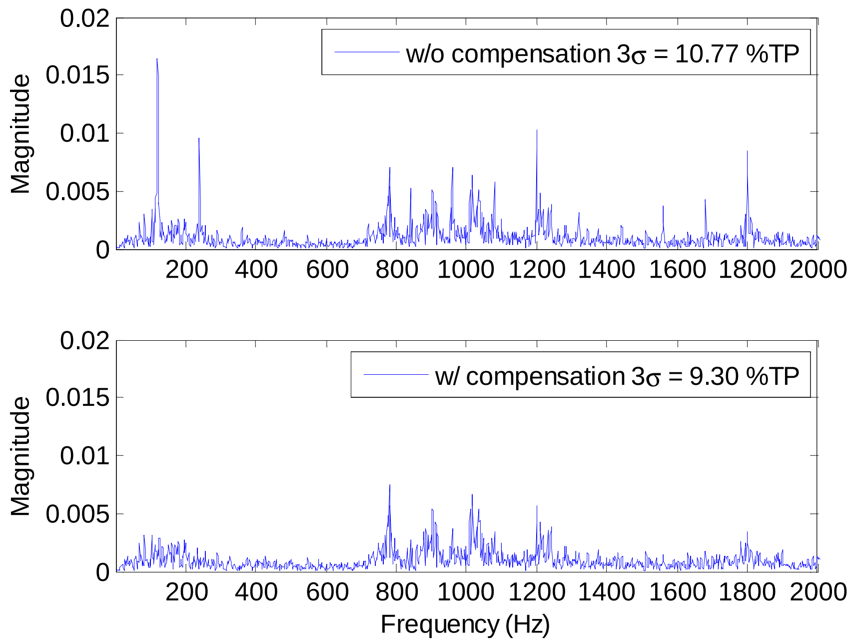


Figure 8.7: Spectra of PES with and without ERC

The bottom plot of Figure 8.8 shows the PES spectrum with ERC and $\alpha = 0$, which corresponds to previous RC schemes. It is observed that the repetitive disturbance components are also significantly reduced. However, due to the amplification of the non-periodic components (see the amplified peaks compared to Figure 8.7, and also the enlarged view in the top plot of Figure 8.8), the overall 3σ value does not improve but is instead amplified, as can be predicted from the steady-state loop-shaping analysis in Figure 8.4. In addition, to avoid excessive high-frequency disturbance amplification, the bandwidth of the zero-phase low-pass filter in $Q(z)$ has to be reduced to 1585 Hz. In this environment that consists of not

[4]The multiple spectral peaks between 800 Hz and 1300 Hz are due to the non-repetitive disk-flutter disturbances.

only repetitive but also a significant amount of non-repetitive disturbances, a conventional RC has experienced difficulty improving the overall regulation performance.

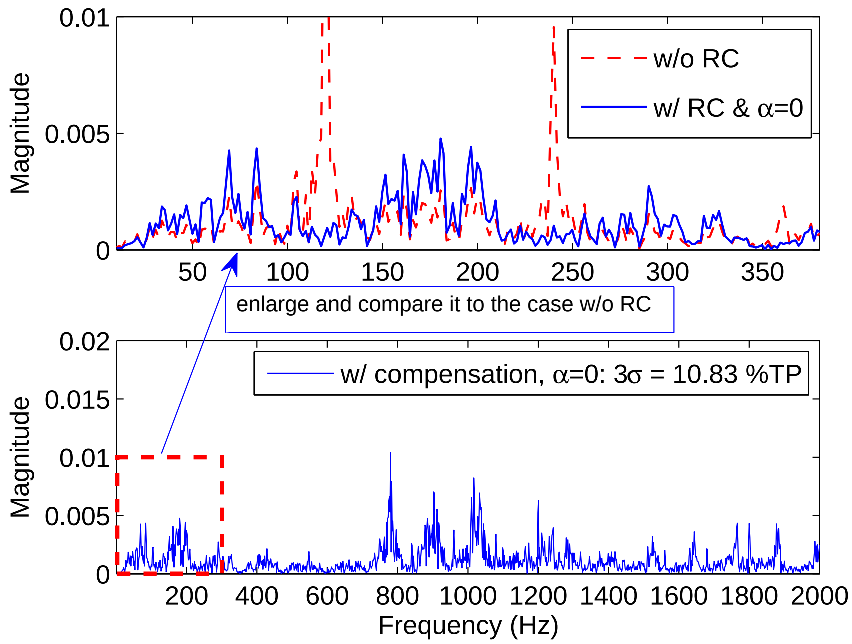


Figure 8.8: PES spectrum in ERC with an FIR Q

To investigate further the transient performance, we provide next the simulation results using an additional disturbance profile that is richer in repetitive components. Figures 8.9 and 8.10 demonstrate time traces of PES using different configurations of α in $Q(z)$. In all cases, the baseline feedback loop has been running for 3 revolutions before ERC is turned on. In Figure 8.9, α maintains at 0.999 in the top plot throughout the simulation, and is configured to exponentially decay from 1 to 0.999, at the rate of 0.9/sample in the bottom plot. We observe that the dynamic switching algorithm provides a much smoother transient response with no visually distinguishable overshoots.

In the top plot of Figure 8.10, the final value of the time-varying α is chosen as 0.99. Compared to the bottom plot of Figure 8.9, we can see that a smaller α yields shorter transient response, as predicted by the analysis in Section 8.5. More specifically, the time constants [defined by (8.20)] for $\alpha = 0.999$ and 0.99 are respectively 4.5432 and 0.4523 revolutions. One can observe from Figure 8.9 and the top plot of Figure 8.10, that the transient durations are indeed about 4.5 and 0.5 revolutions, in agreement with what have just been computed from (8.20). Note that $\alpha = 0.99$ yields worse disturbance rejection results at the steady state. This is supported by the analysis in Section 8.3. One way to balance the performance is to let α first reduce quickly from 1 to 0.99, and then gradually

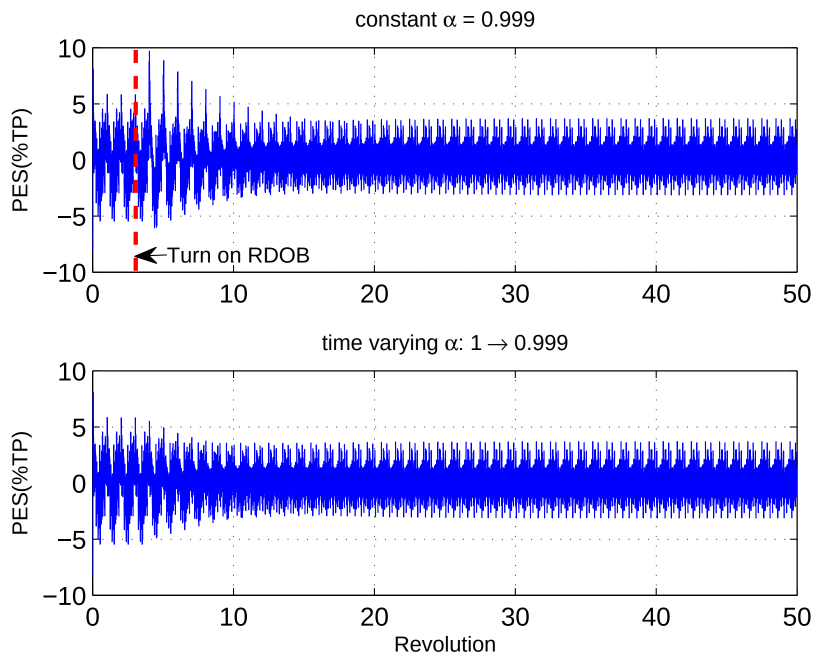


Figure 8.9: Comparison of the transient responses with and without the time-varying α

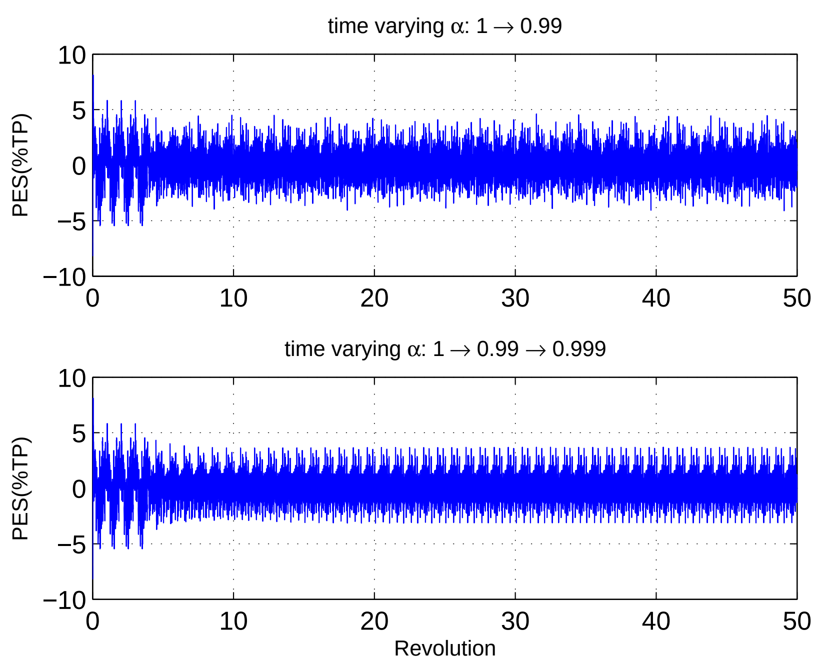


Figure 8.10: Comparison of the transient responses w.r.t. different configurations of α

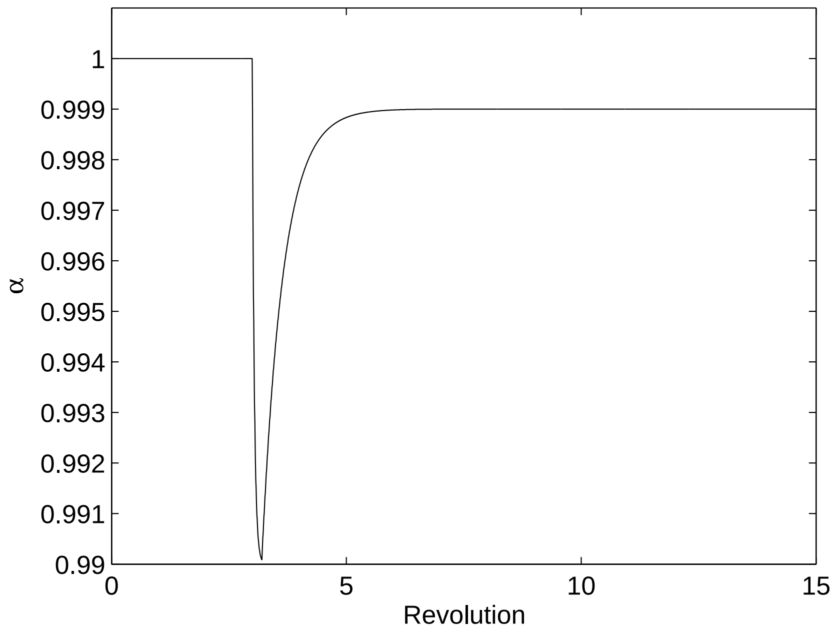


Figure 8.11: Time traces of α to achieve the bottom plot in Figure 8.10

increase to the final value 0.999, as shown in Figure 8.11. The bottom plot of Figure 8.10 depicts the achieved PES time trace using such a configuration.

8.7 Application: Repetitive Tracking and Regulation on a Wafer Scanner

Besides regulation control, the proposed algorithm has also been implemented in tracking control on the wafer-scanner testbed described in Section 3.1 on page 26.

The top stage is used for verification of the proposed algorithm. The system has a nominal model

$$\hat{P}(z) = z^{-2}\frac{3.4766 \times 10^{-7}(1 + 0.8z^{-1})}{(1 - z^{-1})^2}$$

with a baseline PID controller. The applied reference trajectory is shown in the solid line in Figure 8.12. The dashed line in Figure 8.12 shows the tracking result when we apply only the baseline feedback controller. For the resulting tracking errors, the 3σ value is 3.814×10^{-4} m.

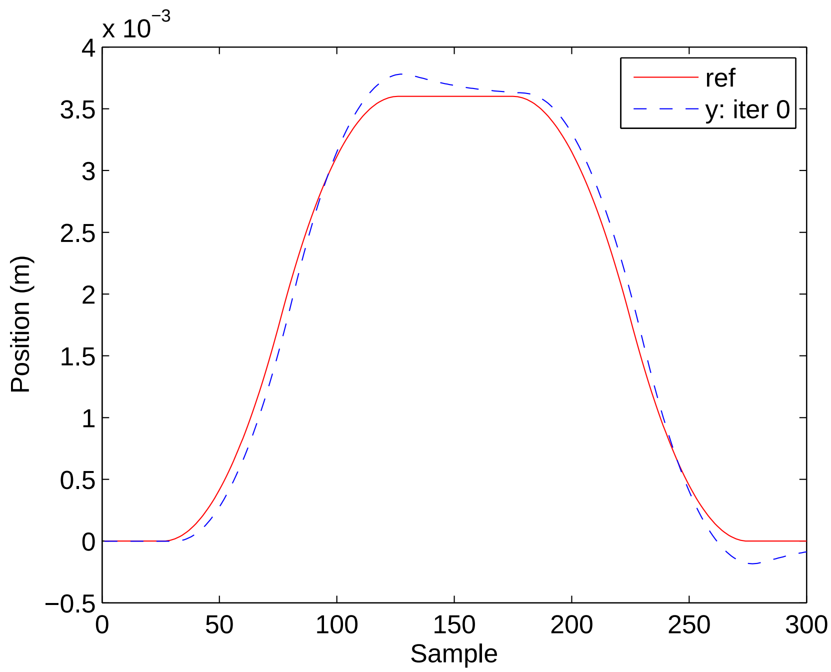


Figure 8.12: Reference trajectory and the actual wafer-stage position without ERC

By the nature of the process, the trajectory is repeatedly applied. Figure 8.13 presents the experimental results of the tracking errors for the first twenty repetitions, where the top and the bottom plots provide respectively the position errors without repetitive control and with the proposed ERC. No transient control for α is applied in Figure 8.13. We can observe that repetitive control has greatly reduced the tracking errors at the steady state. The 3σ value reduces from 3.814×10^{-4} m at the first repetition to 4.160×10^{-6} m at the 20^{th} repetition, indicating a 99.7% reduction.

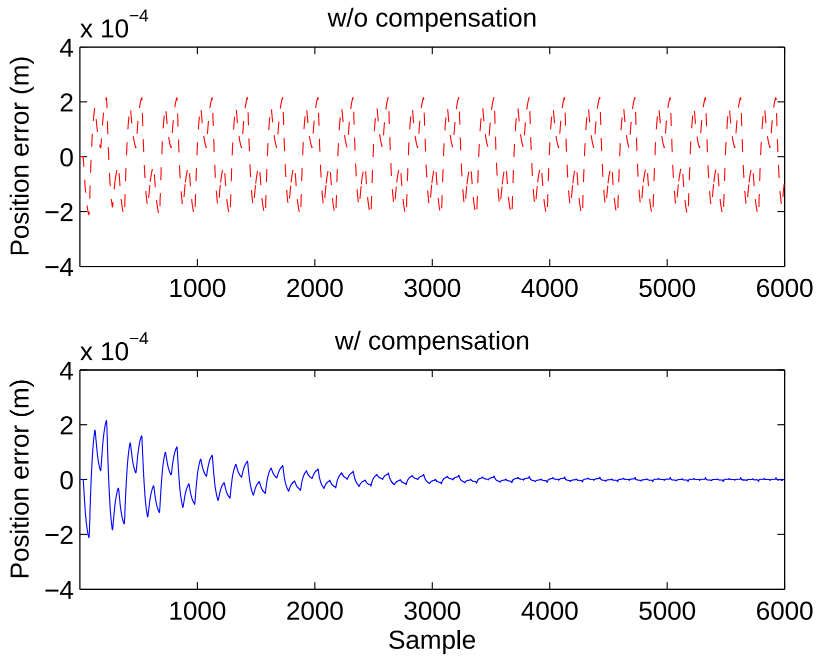


Figure 8.13: Tracking errors with ERC but without transient control

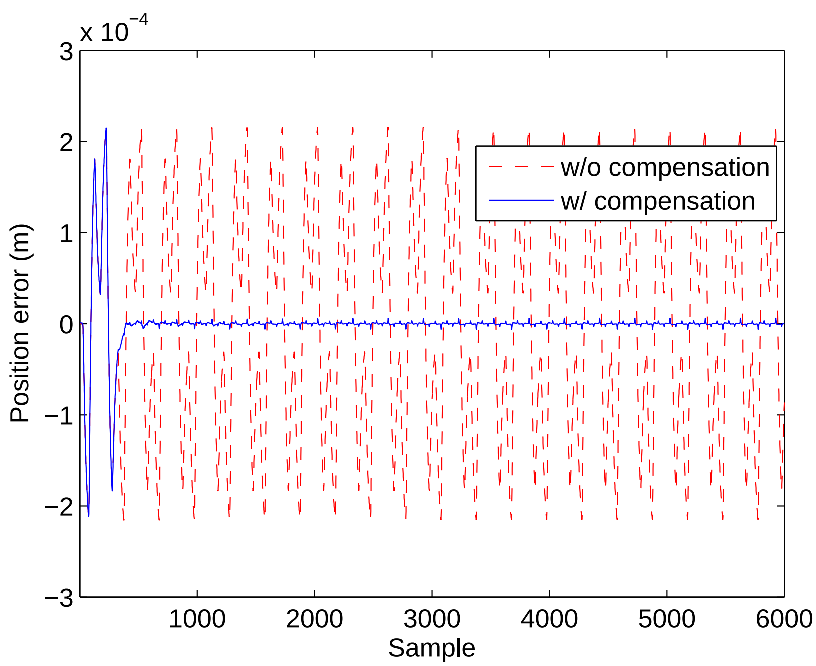


Figure 8.14: Tracking errors with ERC and transient control

The proposed algorithm in Section 8.5 is then applied to additionally accelerate the transient response. Figure 8.14 shows the results for ERC with transient control. Comparing the results with that in Figure 8.13, we can see that the transient duration has been significantly reduced while at the same time the steady-state performance has been preserved.

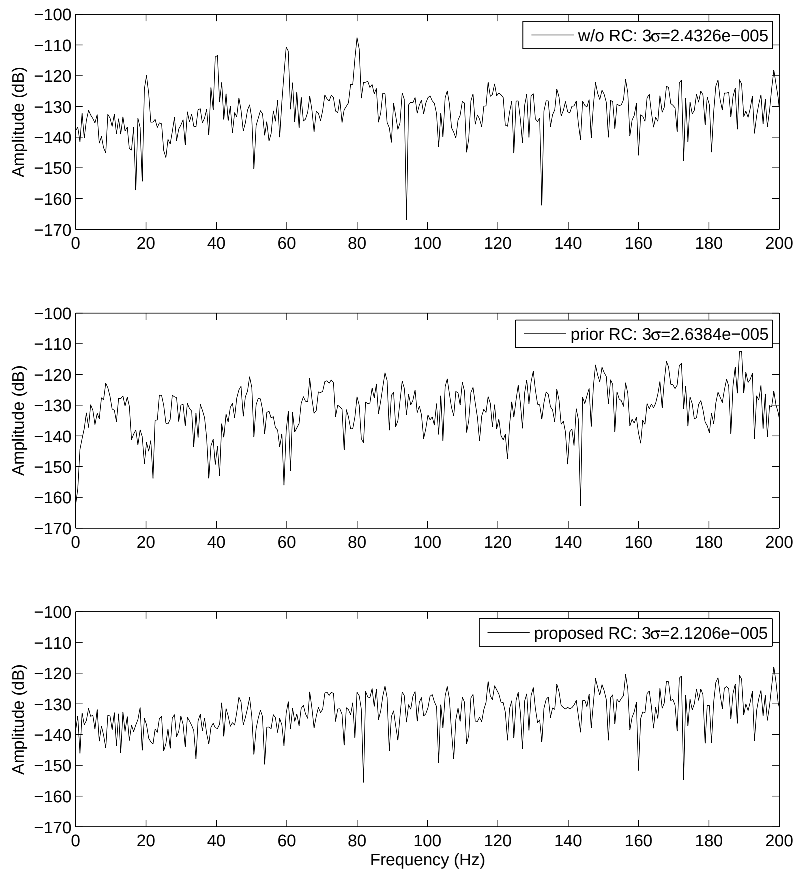


Figure 8.15: Spectra of the tracking errors under different RC schemes

To compare the performance of ERC with that of a conventional RC, a reference trajectory that consists of four sinusoidal components at 20, 40, 60, and 80 Hz is tested. Additionally a random disturbance obeying a normal distribution is applied to the system to examine the performance of the algorithms under noisy environments. Figure 8.15 shows the spectra of the resulting tracking errors. Without repetitive control, large peaks appear at 20, 40, 60, and 80 Hz in the first subplot. Using a conventional RC ($\alpha = 0$), spectral peaks at the repetitive frequencies are removed as shown in the middle plot. However, since all error components in the previous repetition are applied, the non-repetitive errors can be seen to have increased (see the additional spectral peaks at the non-repetitive frequencies). This is also well explained by the loop-shaping results in Figure 8.4. In the proposed scheme, only the repetitive components are reduced, and no visual amplification of the non-repetitive disturbances is observed.

8.8 Notes and references

Connections of ERC With Prior Repetitive-control Schemes With (8.17), the equivalent feedback controller in Figure 8.1 is

$$C_{eq}(z) = \frac{q(z,z^{-1})(1-\alpha^N)z^{-N}\hat{P}^{-1}(z) + (1-\alpha^N z^{-N})C(z)}{1-\left[1-(1-\alpha^N)(1-q(z,z^{-1}))\right]z^{-N}}.$$

For the ideal case of $q(z,z^{-1}) = 1$ (perfect disturbance rejection),

$$C_{eq}(z) = \frac{(1-\alpha^N)z^{-N}\hat{P}^{-1}(z) + (1-\alpha^N z^{-N})C(z)}{1-z^{-N}}.$$

One can remark that the internal model is absorbed in the loop in a series-parallel fashion (the two terms in the numerator of $C_{eq}(z)$ are in parallel form, and the common part $1/(1-z^{-N})$ is in series with them).

Table 8.2 summarizes the equivalent overall feedback controllers in different repetitive-control schemes. The ideal forms in the second and the third columns provide perfect disturbance rejection but are highly sensitivity to model mismatches. low-pass filters in the form of $q(z,z^{-1})$ or $q(s)$ are used in the robust versions. $C(z)$ and $C(s)$ denote the baseline feedback controllers. On the fourth line of Table 8.2, $P_{ZPET}^{-1}(z)$ in prototype RC denotes the ZPET inverse [85] that approximates $P^{-1}(z)$.

Several connections can be made from Table 8.2. First, comparing "Prototype RC" with "Proposed ERC with an FIR Q", we can observe that the former can be regarded as a special case of the latter with $C(z) = 0$ and $\hat{P}^{-1}(z) = P_{ZPET}^{-1}(z)$. Second, if we replace $z^{-N}\hat{P}^{-1}(z)$ with $P^{-1}\sum_m w_m z^{-mN}$ and let $C(z) := C_o(z)\sum_m w_m z^{-mN}$, then the high-order RC can be realized in a similar fashion as the proposed ERC with an FIR Q filter.

It can now be seen that with an *FIR* Q filter, the proposed ERC has close connections with prior RC schemes. From the second and the third rows of Table 8.2, an IIR Q provides a different integration of the internal model and introduces the additional design freedom of α.

Table 8.2: Equivalent feedback controllers in repetitive control schemes

	Ideal form	(expanded) Ideal form	Robust version
Proposed ERC w/ an IIR Q	$\frac{C(z)+z^{-m}\hat{P}^{-1}(z)Q(z)}{1-z^{-m}Q(z)}$, $Q(z)=\frac{(1-\alpha^N)z^{-N+m}}{1-\alpha^N z^{-N}}$	$\frac{(1-\alpha^N)z^{-N}\hat{P}^{-1}(z)+(1-\alpha^N z^{-N})C(z)}{1-z^{-N}}$	$\frac{q(z,z^{-1})(1-\alpha^N)z^{-N}\hat{P}^{-1}(z)+(1-\alpha^N z^{-N})C(z)}{1-[1-(1-\alpha^N)(1-q(z,z^{-1}))]z^{-N}}$
Proposed ERC w/ an FIR Q	$\frac{C(z)+z^{-m}\hat{P}^{-1}(z)Q(z)}{1-z^{-m}Q(z)}$, $Q(z)=z^{-N+m}$	$\frac{z^{-N}\hat{P}^{-1}(z)+C(z)}{1-z^{-N}}$	$\frac{q(z,z^{-1})z^{-N}\hat{P}^{-1}(z)+C(z)}{1-q(z,z^{-1})z^{-N}}$
Prototype RC [95, 108]	$\frac{k_r z^{-N}}{1-z^{-N}}P_{ZPET}^{-1}(z)$, $k_r \in (0,2)$	$\frac{k_r z^{-N}}{1-z^{-N}}P_{ZPET}^{-1}(z)$	$\frac{k_r q(z,z^{-1})z^{-N}}{1-q(z,z^{-1})z^{-N}}P_{ZPET}^{-1}(z)$
Plug-in RC [99, 104, 105, 106, 107, 112, 94]	$C(s)(1+F(s)\frac{e^{-T_p s}}{1-e^{-T_p s}})$, $F(s)$ differs in specific papers.	$C(s)\frac{(1-e^{-T_p s})+F(s)e^{-T_p s}}{1-e^{-T_p s}}$	$C(s)\frac{(1-q(s)\cdot e^{-T_p s})+q(s)F(s)\cdot e^{-T_p s}}{1-q(s)\cdot e^{-T_p s}}$
High-order RC [109, 110]	$\frac{C(z)T_n^{-1}(z)\sum_m w_m z^{-mN}}{1-\sum_m w_m z^{-mN}}$, $T_n \approx \frac{P(z)C(z)}{1+P(z)C(z)}$	$\frac{(P^{-1}(z)+C(z))\sum_m w_m z^{-mN}}{1-\sum_m w_m z^{-mN}}$	$\frac{C(z)T_n^{-1}(z)(\sum_m w_m z^{-mN})q(z,z^{-1})}{1-(\sum_m w_m z^{-mN})q(z,z^{-1})}$

Chapter 9

Decoupled Disturbance Observer for DISO Systems

9.1 Introduction

In Section 4.5, we have discussed the concept of disturbance observers and its connection to pseudo Youla-Kucera parameterization, mainly for the control of SISO systems. The central concept of a disturbance observer is that, if the plant dynamics can be properly inverted, then the equivalent input disturbance can be extracted from the control signal and the measured plant output. This design principle has clear intuitions for SISO systems. For dual-input-single-output (DISO), and more generally multiple-input-single-output (MISO) systems, the situation is more complex. The main difficulty is that, there are multiple actuators (and hence multiple control inputs) while only one combined output signal $y \triangleq \sum_{i=1}^{N} y_i$ is measured. It becomes nontrivial to generate estimated disturbances for each control channel using just the single output y. It also remains unclear how the compensation effort should be distributed to each actuator, as the disturbance information is coupled with the cross-channel control inputs in the output signal.

In this chapter, we discuss a decoupled disturbance observer (DDOB) to address the nontrivial model inversion and the distribution of compensation efforts in MISO systems. We focus on the separation between the external disturbances and the internal control actions. This enables us to have a partial-inverse based disturbance-rejection scheme, where no cross-channel coupling effects enter as internal disturbances, and designers have the flexibility to distribute the compensation effort according to the mechanical properties of each actuator as well as the disturbance characteristics. These features make DDOB especially beneficial, e.g., for vibration rejection, where the disturbance consists of energy components at different frequency ranges, and it is ideal for each control channel to be flexibly adjusted for customized servo enhancement.

9.2 DDOB Structure: Open-loop Configuration

Consider a sampled-data DISO system $P(z) = [P_1(z),\ P_2(z)]$ with the input-output (IO) relation:

$$Y\left(z\right) = P_1(z)U_1\left(z\right) + P_2(z)U_2\left(z\right) + D\left(z\right) \tag{9.1}$$

where $Y\left(z\right)$, $U_i\left(z\right)$ $(i = 1, 2)$ and $D\left(z\right)$ represent respectively the Z transforms of the plant output, the plant inputs, and the lumped external disturbance.

Figure 9.1 shows the structure of the proposed DDOB for the second channel $P_2(z)$. The idea is to apply the compensation signal $c_2\left(k\right)$ to this second actuator, such that the overall lumped disturbance $d\left(k\right)$ is compensated. Here $\hat{P}_i(z)$ is the nominal model of $P_i(z)$, and m_i is the relative degree of $\hat{P}_i(z)$.

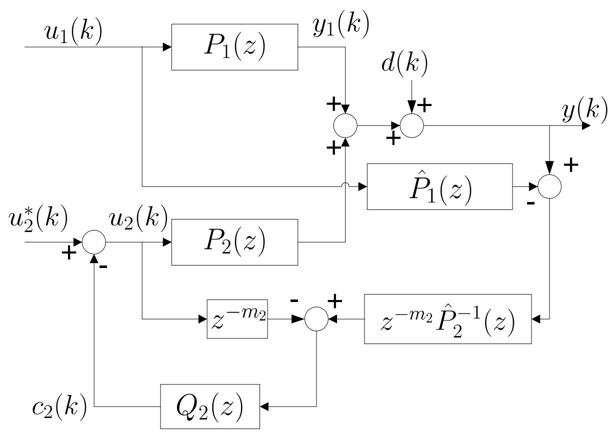


Figure 9.1: Block diagram of DDOB for $P_2(z)$

Time-domain Disturbance-rejection Criteria

From Figure 9.1, the output of $Q_2(z)$ is

$$C_2\left(z\right) = Q_2(z)\left\{z^{-m_2}\hat{P}_2(z)\left[Y\left(z\right) - \hat{P}_1(z)U_1\left(z\right)\right] - z^{-m_2}U_2\left(z\right)\right\} \tag{9.2}$$

which is equivalent to, after substituting in (9.1),

$$\begin{aligned} C_2(z) = Q_2(z)\Big[&z^{-m_2}\hat{P}_2^{-1}(z)\Big(P_1(z) - \hat{P}_1(z)\Big))U_1(z) \\ &+z^{-m_2}\Big(\hat{P}_2^{-1}(z)P_2(z) - 1\Big)U_2(z)\Big] + Q_2(z)z^{-m_2}\hat{P}_2^{-1}(z)D(z). \end{aligned} \tag{9.3}$$

If $P_i(z) = \hat{P}_i(z)$, (9.3) reduces to

$$C_2\left(z\right) = Q_2(z)z^{-m_2}\hat{P}_2^{-1}(z)D\left(z\right). \tag{9.4}$$

Recall in Figure 9.1, that

$$Y(z) = P_1(z)U_1(z) + P_2(z^{-1})U_2^*(z) + (D(z) - P_2(z)C_2(z))$$

where $U_1(z)$ and $U_2^*(z)$ are from the outer-loop control design and $D(z) - P_2(z)C_2(z)$ explains how the external disturbance is compensated in the inner DDOB loop. Applying (9.4), we have

$$D(z) - P_2(z)C_2(z) = \left(1 - z^{-m_2}Q_2(z)\right)D(z). \tag{9.5}$$

Equation (9.5) contains no information about the first controlled channel. It is now seen that the disturbance rejection is entirely decoupled to the second channel, and relies on the design of a single filter $Q_2(z)$. By the above construction, the disturbance $D(z)$ is decoupled from $Y_1(z)$, the position output of the first actuator, and "observed" in the compensation signal $C_2(z)$.

Model-following Property

One can remark that when $P_i(z)$ differs from $\hat{P}_i(z)$, the model mismatch is absorbed as an internal disturbance in (9.3) (see the first two terms in the square brackets). Notice that $C_2 = U_2 - U_2^*$ in (9.3). Solving for $U_2(z)$ and substituting the result to (9.1), we can obtain the IO relation:

$$Y(z) = G_{yd}(z)D(z) + G_{yu_1}(z)U_1(z) + G_{yu_2^*}(z)U_2^*(z),$$

where the three transfer functions are given by

$$\begin{aligned}
G_{yd}(z) &= 1 - \frac{\hat{P}_2^{-1}(z)P_2(z)z^{-m_2}Q_2(z)}{1 + \left(\hat{P}_2^{-1}(z)P_2(z) - 1\right)z^{-m_2}Q_2(z)} \\
G_{yu_1}(z) &= P_1(z) - \frac{\hat{P}_2^{-1}(z)P_2(z)\left(P_1(z) - \hat{P}_1(z)\right)z^{-m_2}Q_2(z)}{1 + \left(\hat{P}_2^{-1}(z)P_2(z) - 1\right)z^{-m_2}Q_2(z)} \\
G_{yu_2^*}(z) &= \frac{P_2(z)}{1 + \left(\hat{P}_2^{-1}(z)P_2(z) - 1\right)z^{-m_2}Q_2(z)}.
\end{aligned}$$

If $z^{-m_2}Q(z) = 1$, the above reduces to

$$G_{yd}(z) = 0,\ G_{yu_1}(z) = \hat{P}_1(z),\ G_{yu_2^*}(z) = \hat{P}_2(z). \tag{9.6}$$

Here $G_{yd}(z) = 0$ explains the disturbance-rejection result in Section 9.2. Additionally, we observe that the dynamics between the nominal inputs (u_1 and u_2^*) and the output is now forced to follow the nominal model $\hat{P}_i(z)$ $(i = 1, 2)$—thus the rejection of modeling mismatch within the DDOB loop. DDOB hence has the nominal-model-following property. Notice that although it is not practical to have $z^{-m_2}Q(z) = 1$ over the entire frequency region, (9.6) equally holds if we replace z with $e^{j\omega}$, in which case the model following is enforced at the frequencies where $e^{-m_2 j\omega}Q(e^{j\omega}) = 1$.

Operation of Two DDOBs

Swapping every applicable sub-index between 1 and 2 in the preceding discussions, we get the DDOB for $P_1(z)$. By linearity and (9.4), if two DDOBs operate simultaneously, the disturbance compensation is achieved by

$$D\left(z\right)-P_1(z)C_1\left(z\right)-P_2(z)C_2\left(z\right)=\left(1-z^{-m_1}Q_1(z)-z^{-m_2}Q_2(z)\right)D\left(z\right). \tag{9.7}$$

One can remark that if a single DDOB already achieves canceling the disturbance (i.e., $D(z)-P_1(z)C_1(z)$ approximates 0), then the second DDOB is not necessary and we should set $Q_2(z)=0$. This is the ideal situation when one actuator alone can effectively handle all the disturbances. In practice, this may not always be feasible due to the mechanical limitation of the actuator. One approach to utilize (9.7) is to make

$$z^{-m_1}Q_1(z)+z^{-m_2}Q_2(z)\approx 1$$

in the interested frequency region. Since there is only one constraint and two filters to design, the selection of $Q_1(z)$ and $Q_2(z)$ will not be unique. We propose to apply frequency-dependent DDOBs based on the actuator dynamics and disturbance properties. For example, in HDD applications, the VCM actuator ($P_1(z)$ in Figure 9.1) has a large actuation range and the microactuator ($P_2(z)$ in Figure 9.1) suits only for small-range positioning. Additionally, $\hat{P}_1^{-1}(z)$ has properties similar to a double differentiator in the high-frequency region [28, 113], yielding large high-frequency noises in the output of $\hat{P}_1^{-1}(z)$. Such actuator dynamics renders DDOB for $P_1(z)$ to have increased difficulties as the disturbance frequency gets higher and higher. The microactuator on the other hand has a model of a DC gain plus resonances above 4 kHz, and a better signal-to-noise ratio during implementation of $\hat{P}_2^{-1}(z)$. From the above considerations, in the low-frequency region, we can apply DDOB to the large-stroke VCM actuator, by assigning $Q_1(z)$ to be a low-pass/band-pass filter and $Q_2(e^{j\omega})\approx 0$. At middle and high frequencies, the precise and faster-response microactuator can be more effectively used. This is achieved by assigning $Q_1(e^{j\omega})\approx 0$ and $Q_2(z)$ to have a band-pass structure. Throughout this chapter, unless otherwise stated, we assume the above decoupled disturbance-rejection scheme.

Extension to General MISO Systems

In the construction of DDOB in Figure 9.1, we have applied the $\hat{P}_1(z)$ block to remove the coupling of $u_1(k)$ in $y(k)$. Extending this concept to a general multiple-input-single output system, we obtain Figure 9.2, which depicts the block diagram for designing DDOB for the n-th actuation stage. Two paths are constructed for each control signal $u_i(k)$, $i=1,\ldots,n-1$: the first through the physical actuator dynamics $P_i(z)$, the second through the model $\hat{P}_i(z)$. Notice the minus sign after each $\hat{P}_i(z)$ block. The effect of $u_i(k)$ is thus removed from $y(k)$, and $d(k)$ is the only remaining signal component that flows into the block $z^{-m_n}\hat{P}_n^{-1}(z)$. Using analogous analysis as that in Section 9.2, we can get

$$D\left(z\right)-P_n(z)C_n\left(z\right)=\left(1-z^{-m_n}Q_n(z)\right)D\left(z\right) \tag{9.8}$$

if $P_i(z) = \hat{P}_i(z)$. Equation (9.8) has the same structure as (9.5). The same design techniques can thus be applied to design $Q_n(z)$.

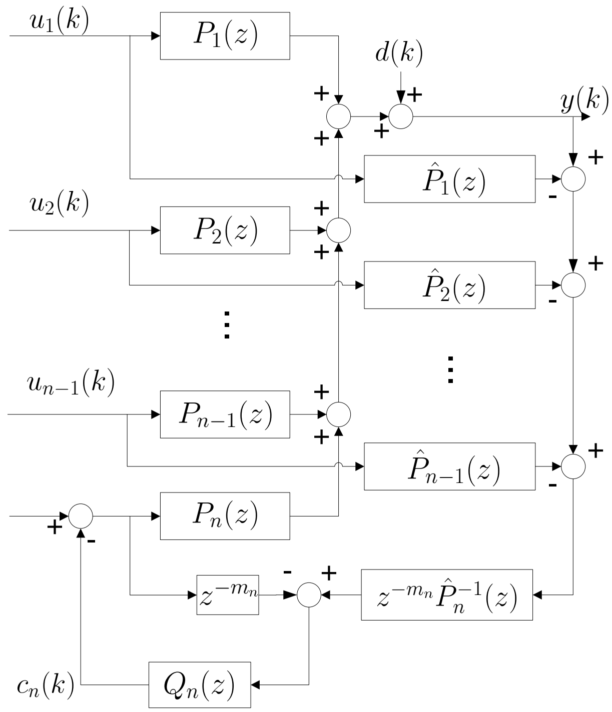


Figure 9.2: DDOB for general MISO systems

9.3 Nominal Stability and Frequency-domain Loop-shaping Criteria

This section discusses the design criteria and the nominal stability when DDOB is applied to a closed loop consisting of the DISO plant and a baseline feedback controller $C(z) = [C_1(z), C_2(z)]^T$. Figure 9.3 shows the closed-loop controller implementation. We focus on the regulation problem to reject $d(k)$, and assume r is zero for stability and loop-shaping analysis. We will present analysis of DDOB for the secondary actuator. The result for the first actuator is immediate after inter-changing the sub-indexes between 1 and 2 in the transfer functions.

For simplified analysis, the influence of DDOB can be absorbed into the second-channel controller. We have

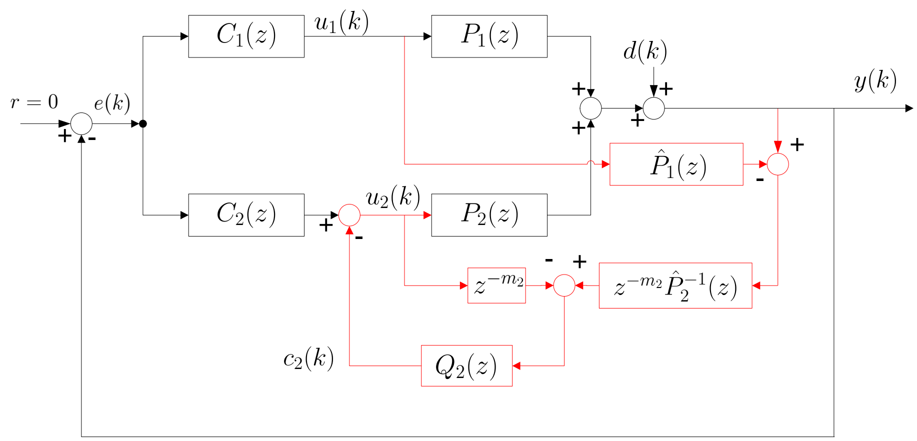


Figure 9.3: Block diagram of the closed loop with DDOB for $P_2(z)$

Proposition 9.1. *Figure 9.3 is equivalent to Figure 9.6, with*

$$C_{2,s}(z) = \frac{1}{1 - z^{-m_2}Q_2(z)} \tag{9.9}$$

$$C_{2,p}(z) = \left[1 + \hat{P}_1(z)C_1(z)\right] z^{-m_2}\hat{P}_2^{-1}(z)Q_2(z). \tag{9.10}$$

Proof. Splitting the output of $Q_2(z)$ into two parts, and relocating the summing junction after $\hat{P}_1(z)$, we get Figure 9.4a. Since the reference is zero, Figure 9.4a is equivalent to Figure 9.4b. Finally, noting that the $\hat{P}_1(z)C_1(z)$ block in Figure 9.4b does not influence the path from $C_1(z)$ to $P_1(z)$, we obtain Figure 9.5a, which is equivalent to Figures 9.5b and 9.6, with the add-on serial and parallel terms given by (9.9) and (9.10). □

Remark. Figure 9.5a and Figure 9.3 can both be implemented in practice. Figure 9.3 is more suited for regulation while Figure 9.5a works for both tracking and regulation problems.

From Figure 9.6, the loop transfer function from the feedback error to the output $y(k)$ is

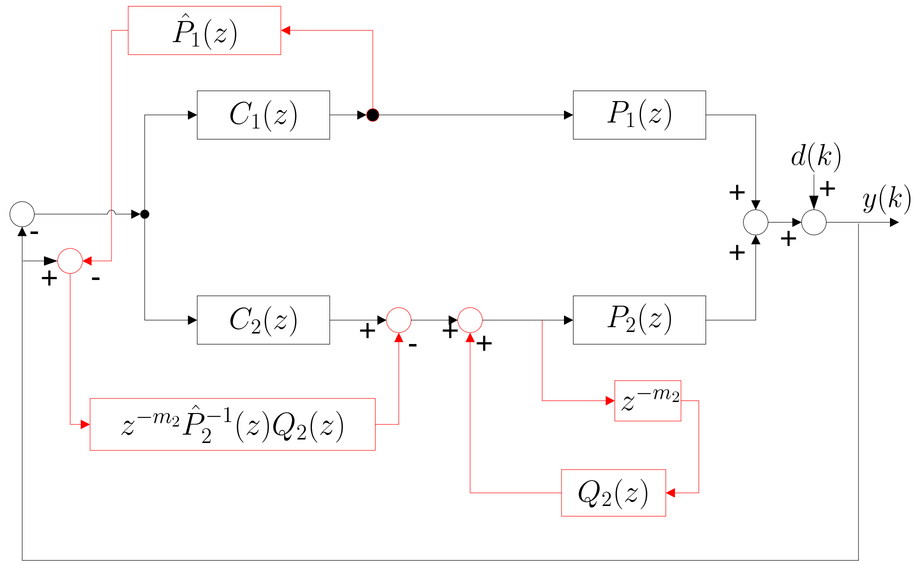
$$L(z) \tag{9.11}$$

$$= P_1(z)C_1(z) + P_2(z)C_{2,s}(z)\left(C_2(z) + C_{2,p}(z)\right) \tag{9.12}$$

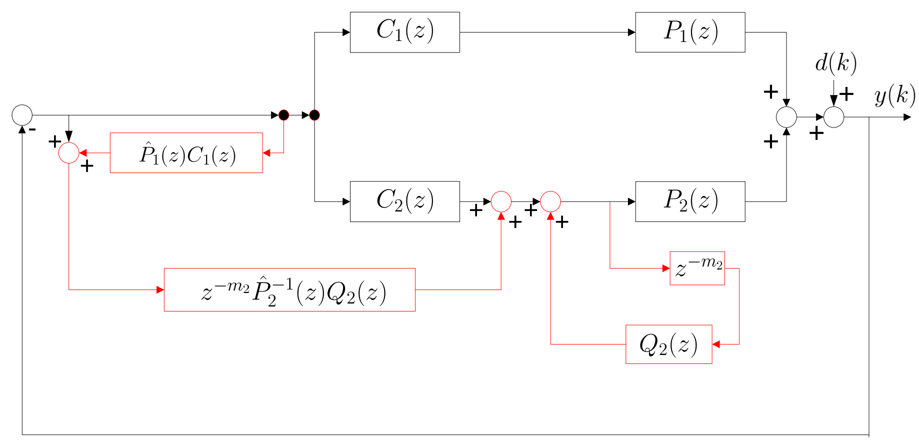
$$= P_1(z)C_1(z) + P_2(z)\frac{C_2(z) + \left(1 + \hat{P}_1(z)C_1(z)\right)z^{-m_2}\hat{P}_2^{-1}(z)Q_2(z)}{1 - z^{-m_2}Q_2(z)}. \tag{9.13}$$

If $P_i = \hat{P}_i$, (9.13) simplifies to

$$L(z) = \frac{P_1(z)C_1(z) + P_2(z)C_2(z) + z^{-m_2}Q_2(z)}{1 - z^{-m_2}Q_2(z)}.$$



(a) An equivalent form of the system in Figure 9.3: the input to $z^{-m_2}\hat{P}_2(z)$ is relocated; the summing junction before $Q_2(z)$ is separated



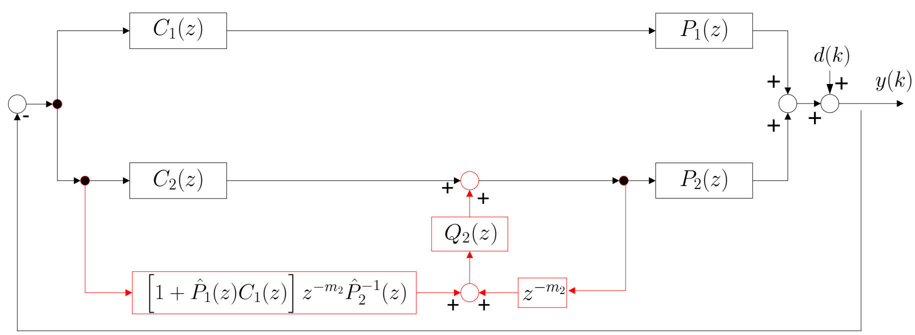
(b) An equivalent form of Figure 9.6: the signs of the signals are changed after another relocation of block diagrams

Figure 9.4: Block diagram transformation for Figure 9.3

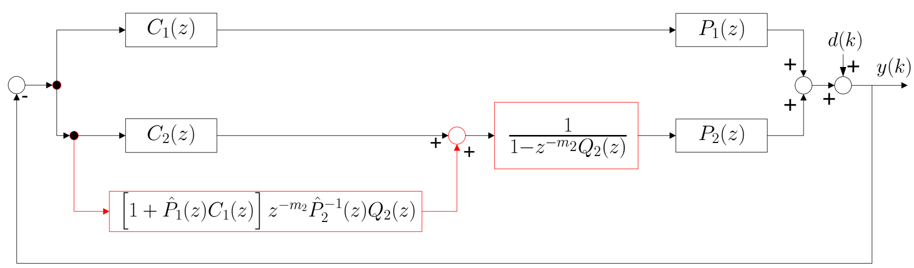
The sensitivity function of the closed-loop system is therefore given by

$$S(z) = \frac{1}{1+L(z)} = \frac{1 - z^{-m_2} Q_2(z)}{1 + P_1(z)C_1(z) + P_2(z)C_2(z)}. \tag{9.14}$$

Notice that $1/(1 + P_1(z)C_1(z) + P_2(z)C_2(z))$ is the baseline closed-loop sensitivity function. $S(z)$ therefore is stable as long as $Q_2(z)$ is stable. Additionally, similar to SISO pseudo YK parameterization, $1 - z^{-m_2}Q_2(z)$ can be applied as a frequency-domain design criteria for loop shaping. Specifically, from (9.14), the complementary sensitivity function is



(a) An equivalent form of Figure 9.4b



(b) An equivalent form of Figure 9.5a

Figure 9.5: Block diagram transformation for Figure 9.4: DDOB is decomposed to series and parallel modules

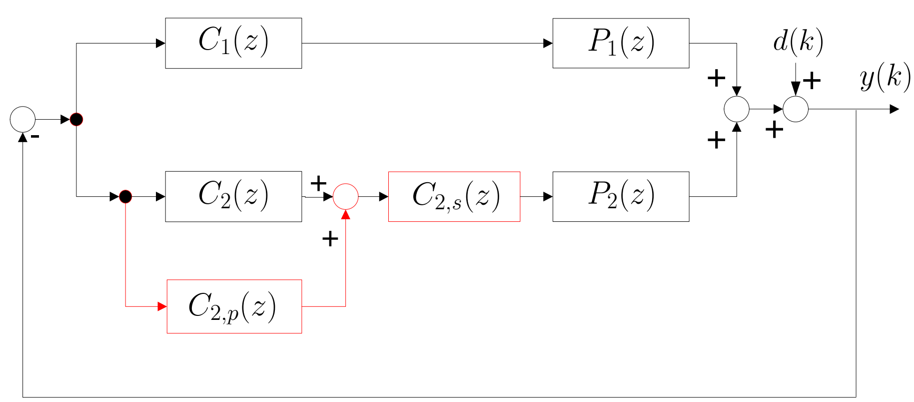


Figure 9.6: An equivalent block diagram of the system in Figure 9.5

$$T(z) = \frac{P_1(z)\,C_1(z) + P_2(z)\,C_2(z) + z^{-m_2}Q_2(z)}{1 + P_1(z)\,C_1(z) + P_2(z)\,C_2(z)}. \tag{9.15}$$

In the frequency regions where $z^{-m_2}Q_2(z)$ is approximately 1, $S(z) \approx 0$ in (9.14) and $T(z) \approx 1$ in (9.15), i.e., the closed-loop system has enhanced performance of disturbance rejection and reference following. When $z^{-m_2}Q_2(z)$ is approximately 0, $S(z)$ and $T(z)$ are close to their baseline versions (without DDOB) and the original system response is preserved. One can notice that the DDOB inherits the affine loop-shaping criteria about SISO YK parameterization. Indeed, DDOB is a special case of MIMO Youla-Kucera parameterization. Before connecting these two concepts, we provide first the full version of the nominal-stability condition:

Theorem 9.1. *(DDOB nominal stability) Given an internally stable baseline feedback system in Figure 9.3, if the exact model of the plant is available and the following conditions hold:*
(i) $\hat{P}_1(z)$ and $C(z)$ are stable;
(ii) $\hat{P}_2(z)$ is a minimum-phase system;
then the closed-loop system in Figure 9.3 is internally stable as long as $Q_2(z)$ is stable.

Proof. From (9.13), under the stated conditions, the closed-loop characteristic polynomial comes from

$$1 + P_1C_1 + P_2\frac{C_2 + \left(1 + \hat{P}_1C_1\right) z^{-m_2}\hat{P}_2^{-1}Q_2}{1 - z^{-m_2}Q_2} = 0$$

namely,

$$1 + P_1C_1 + P_2C_2 - z^{-m_2}Q_2 - z^{-m_2}Q_2P_1C_1 + z^{-m_2}Q_2P_2\hat{P}_2^{-1} + z^{-m_2}Q_2P_2\hat{P}_2^{-1}\hat{P}_1C_1 = 0$$

If $P_i = \hat{P}_i$, then the two pairs of terms $(-z^{-m_2}Q_2,\ z^{-m_2}Q_2P_2\hat{P}_2^{-1})$ and $(-z^{-m_2}Q_2P_1C_1,\ z^{-m_2}Q_2P_2\hat{P}_2^{-1}\hat{P}_1C_1)$ get canceled. However the canceled terms will still contribute to internal states. The actual closed-loop characteristic polynomial is

$$D_{Q_2}D_{P_2}N_{\hat{P}_2}D_{\hat{P}_1}D_{C_1}\left(D_{P_1}D_{P_2}D_{C_1}D_{C_2} + N_{P_1}N_{C_1}D_{P_2}D_{C_2} + N_{P_2}N_{C_2}D_{P_1}D_{C_1}\right)$$

where $N_{(\cdot)}$ and $D_{(\cdot)}$ denote respectively the numerator and the denominator of a transfer function.

Notice that

$$D_{P_1}D_{P_2}D_{C_1}D_{C_2} + N_{P_1}N_{C_1}D_{P_2}D_{C_2} + N_{P_2}N_{C_2}D_{P_1}D_{C_1}$$

is the characteristic polynomial for the baseline system. The internal stability follows readily from the assumptions. □

9.4 DDOB is A Special Youla-Kucera Parameterization for DISO systems

Standard MIMO YK Parameterization

We first extend the discussions of Chapter 4 and review YK parameterization for general MIMO systems. Dimensions of transfer functions now play important roles. We have:

Definition 9.1. (**right coprime**) Suppose $N \in \mathcal{S}^{n_y \times n_u}$ and $D \in \mathcal{S}^{n_u \times n_u}$. The pair (N, D) is called right coprime over $\mathcal{S}$ if there exists $U \in \mathcal{S}^{n_u \times n_y}$, $V \in \mathcal{S}^{n_u \times n_u}$ such that

$$UN + VD = I_{n_u}.$$

If D^{-1} is additionally realizable, then (N, D) is called a **right-coprime factorization** of a rational and proper transfer function $G = ND^{-1} \in \mathcal{R}^{n_y \times n_u}$.

(**left coprime**) Suppose $N \in \mathcal{S}^{n_y \times n_u}$ and $D \in \mathcal{S}^{n_y \times n_y}$. The pair (N, D) is called left coprime over $\mathcal{S}$ if there exists $U \in S^{n_u \times n_y}$, $V \in \mathcal{S}^{n_y \times n_y}$ such that

$$NU + DV = I_{n_y}.$$

If D^{-1} is additionally realizable, then (N, D) is called a **left-coprime factorization** of $G = D^{-1}N \in \mathcal{R}^{n_y \times n_u}$.

Theorem 9.2. *Consider a n_u-input-n_y-output plant P. Let P have the right-coprime factorization $P = ND^{-1}$, with $N \in \mathcal{S}^{n_y \times n_u}$ and $D \in \mathcal{S}^{n_u \times n_u}$. Assume a stabilizing controller $C = XY^{-1}$ (in a negative feedback loop), with $X \in \mathcal{S}^{n_u \times n_y}$ and $Y \in \mathcal{S}^{n_y \times n_y}$ being the right-coprime factorization elements of C. Then the set of all stabilizing controllers for P is given by*

$$\left\{ (X + DQ)(Y - NQ)^{-1} : \ Q \in \mathcal{S}^{n_u \times n_y} \right\} \tag{9.16}$$

or, in the left-coprime format,

$$\left\{ \left(\bar{Y} - \bar{Q}\bar{N} \right)^{-1} \left(\bar{X} + \bar{Q}\bar{D} \right) : \ \bar{Q} \in \mathcal{S}^{n_u \times n_y} \right\} \tag{9.17}$$

where $P = \bar{D}^{-1}\bar{N}$ is the left-coprime factorization of P with $\bar{N} \in \mathcal{S}^{n_y \times n_u}$, $\bar{D} \in \mathcal{S}^{n_y \times n_y}$; $C = \bar{Y}^{-1}\bar{X}$ is the left-coprime factorization of C with $\bar{X} \in \mathcal{S}^{n_u \times n_y}$, $\bar{Y} \in \mathcal{S}^{n_u \times n_u}$. For well-posedness, the mild requirements

$$\det (Y(\infty) - N(\infty) Q(\infty)) \neq 0$$

and

$$\det \left(\bar{Y}(\infty) - \bar{Q}(\infty) \bar{N}(\infty) \right) \neq 0$$

are assumed in (9.16) and (9.17).

Proof. See [63]. □

A Special DISO YK Parameterization

For DISO systems, consider the case when $P_1(z)$ is stable and the second actuator $P_2(z)$ is a minimum-phase system that can be factorized as $P_2(z) = z^{-m_2} P_{2m}(z)$, where $P_{2m}(z)$ has a relative degree of zero.[1]

[1]The notation P_{2m} reads "the transfer function of the second actuator, with the m-delay term separated".

Note that $P_{2m}^{-1}(z) = z^{-m_2} P_2^{-1}(z)$ is stable from the minimum-phase assumption, and that z^{-m_2} is stable by definition. We have $z^{-m_2} \in \mathcal{S}$, $P_{2m}^{-1}(z) \in \mathcal{S}$, and hence

$$P_2(z) = \frac{z^{-m_2}}{P_{2m}^{-1}(z)} = \frac{z^{-m_2}}{z^{-m_2} P_2^{-1}(z)}$$

is a valid coprime factorization for $P_2(z)$. The full DISO plant $P(z)$ can then be (left-coprime) factorized as

$$P(z) = [P_1(z),\ P_2(z)] = \left(P_{2m}^{-1}(z)\right)^{-1} \left[P_{2m}^{-1}(z) P_1(z),\ z^{-m_2}\right]. \tag{9.18}$$

For simplicity, we will also make the practical assumption that the baseline stabilizing controllers $C_1(z)$ and $C_2(z)$ are stable—namely, $C_1(z),\ C_2(z) \in \mathcal{S}$—so that the following is a valid (left) coprime factorization for the baseline controller:

$$C(z) = \begin{bmatrix} C_1(z) \\ C_2(z) \end{bmatrix} = I_{2\times 2}^{-1} \begin{bmatrix} C_1(z) \\ C_2(z) \end{bmatrix}. \tag{9.19}$$

From (9.17) in Theorem 9.2, all the stabilizing controllers can be factorized as

$$\begin{aligned} &C_{all}(z) = \\ &\left(I - \begin{bmatrix} Q_1(z) \\ Q_2(z) \end{bmatrix} [P_{2m}^{-1}(z) P_1(z),\ z^{-m_2}]\right)^{-1} \left(\begin{bmatrix} C_1(z) \\ C_2(z) \end{bmatrix} + \begin{bmatrix} Q_1(z) \\ Q_2(z) \end{bmatrix} P_{2m}^{-1}(z)\right). \end{aligned} \tag{9.20}$$

Introduce now the dummy variable

$$\begin{aligned} &M(z) \\ &\triangleq P(z) \times \left(I - \begin{bmatrix} Q_1(z) \\ Q_2(z) \end{bmatrix} [P_{2m}^{-1}(z) P_1(z),\ z^{-m_2}]\right)^{-1} && (9.21) \\ &= \left(P_{2m}^{-1}(z)\right)^{-1} [P_{2m}^{-1}(z) P_1(z),\ z^{-m_2}] \left(I - \begin{bmatrix} Q_1(z) \\ Q_2(z) \end{bmatrix} [P_{2m}^{-1}(z) P_1(z),\ z^{-m_2}]\right)^{-1} \\ &= \frac{[P_1(z),\ z^{-m_2} P_{2m}(z)]}{1 - P_{2m}^{-1}(z) P_1(z) Q_1(z) - z^{-m_2} Q_2(z)} && (9.22) \end{aligned}$$

where in the last equality of (9.22), we have used the matrix identity

$$B(I + AB)^{-1} = (I + BA)^{-1} B$$

and the fact that

$$I - [P_{2m}^{-1}(z) P_1(z),\ z^{-m_2}] \begin{bmatrix} Q_1(z) \\ Q_2(z) \end{bmatrix}$$

is a scalar transfer function.

The closed-loop sensitivity function thus can be expressed as

$$\begin{aligned} S(z) &= (1+P(z)C_{all}(z))^{-1} \\ &= \left\{1+M(z)\times\left(\begin{bmatrix} C_1(z) \\ C_2(z)\end{bmatrix}+\begin{bmatrix} Q_1(z) \\ Q_2(z)\end{bmatrix}P_{2m}^{-1}(z)\right)\right\}^{-1}, \end{aligned}$$

which is equivalent to [by using (9.22)]

$$S(z) = \frac{1-P_{2m}^{-1}(z)P_1(z)Q_1(z)-z^{-m_2}Q_2(z)}{1+P_1(z)C_1(z)+P_2(z)C_2(z)}. \tag{9.23}$$

Notice that

$$S_o(z) = \frac{1}{1+P_1(z)C_1(z)+P_2(z)C_2(z)} = \frac{1}{1+P(z)C(z)}$$

is the sensitivity function of the baseline feedback system. Therefore

$$S(z) = S_o(z)\left\{1-P_{2m}^{-1}(z)P_1(z)Q_1(z)-z^{-m_2}Q_2(z)\right\} \tag{9.24}$$

which reduces to (9.14) for DDOB if $Q_1(z)=0$. Indeed the feedback controller also becomes the same as that in DDOB. To see this, when $Q_1(z)=0$, we have the following simplification of (9.20):

$$\begin{aligned} &C_{all}(z) \\ &= \left(I-\begin{bmatrix} 0 \\ Q_2(z)\end{bmatrix}\left[P_{2m}^{-1}(z)P_1(z),\ z^{-m_2}\right]\right)^{-1}\left(\begin{bmatrix} C_1(z) \\ C_2(z)\end{bmatrix}+\begin{bmatrix} 0 \\ Q_2(z)\end{bmatrix}P_{2m}^{-1}(z)\right) \\ &= \begin{bmatrix} 1 & 0 \\ -Q_2(z)P_{2m}^{-1}(z)P_1(z) & 1-z^{-m_2}Q_2(z)\end{bmatrix}^{-1}\begin{bmatrix} C_1(z) \\ C_2(z)+Q_2(z)P_{2m}^{-1}(z)\end{bmatrix} \\ &= \frac{1}{1-z^{-m_2}Q_2(z)}\begin{bmatrix} 1-z^{-m_2}Q_2(z) & 0 \\ Q_2(z)P_{2m}^{-1}(z)P_1(z) & 1\end{bmatrix}\begin{bmatrix} C_1(z) \\ C_2(z)+Q_2(z)P_{2m}^{-1}(z)\end{bmatrix} \\ &= \begin{bmatrix} C_1(z) \\ \frac{C_2(z)+Q_2(z)P_{2m}^{-1}(z)[1+P_1(z)C_1(z)]}{1-z^{-m_2}Q_2(z)}\end{bmatrix}. \end{aligned} \tag{9.25}$$

On the other hand, from Figure 9.6 and Equations (9.9)–(9.10), the overall equivalent feedback controller in DDOB based control is

$$\begin{aligned} C'(z) &= \begin{bmatrix} C_1(z) \\ C_{2,s}(z)\left(C_2(z)+C_{2,p}(z)\right)\end{bmatrix} \\ &= \begin{bmatrix} C_1(z) \\ \frac{C_2(z)+\left[1+\hat{P}_1(z)C_1(z)\right]z^{-m_2}\hat{P}_2^{-1}(z)Q_2(z)}{1-z^{-m_2}Q_2(z)}\end{bmatrix}, \end{aligned}$$

which is nothing but (9.25) when $P_i=\hat{P}_i$ (notice that $P_{2m}^{-1}=z^{-m_2}P_2^{-1}$).

Removal of the assumption on $C_2(z)$ being stable: Similar to the derivation of the pseudo Youla-Kucera parameterization in Chapter 4, we see that DDOB itself does not require $C_2(z)$ to be strictly stable (Theorem 9.1). If we additionally want to enable the parameterization of all stabilizing controllers, stability of $C_2(z)$ is required for (9.19) to be a valid coprime factorization for the special DISO YK parameterization.

Changes to the sensitivity function: Note that (9.24) shares a similar characteristics as (9.7). Designing $Q_1(z) = z^{-m_1} P_{2m}(z)\hat{P}_1^{-1}(z)$ in (9.24) actually recovers (9.7). DDOB therefore provides an intuitive explanation of the DISO Youla-Kucera parameterization, from the view point of disturbance cancellation.

9.5 Robust Stability

This section analyzes the robust-stability condition when the plant is perturbed to

$$\tilde{P}_1(e^{j\omega}) = P_1(e^{j\omega})\left(1 + W_1(e^{j\omega})\Delta_1(e^{j\omega})\right) \tag{9.26}$$

$$\tilde{P}_2(e^{j\omega}) = P_2(e^{j\omega})\left(1 + W_2(e^{j\omega})\Delta_2(e^{j\omega})\right), \tag{9.27}$$

where $W_1(e^{j\omega})$ and $W_2(e^{j\omega})$ are frequency weighting functions, and the multiplicative uncertainties satisfy $|\Delta_i(e^{j\omega})| \leq 1$. Notice that since the perturbed plant is for stability analysis rather than H_2/H_∞ synthesis, we do not restrict $W_i(e^{j\omega})$ to come from the frequency response of a transfer function.

As the DISO plant is a special MIMO system, the μ-analysis (see, e.g., [114]) tool can be applied to derive the robust stability condition. Usually the structured singular value μ is only approximated by its upper bound. However, for the special case of DISO plants, we have a closed-form solution for μ:

Theorem 9.3. *The closed-loop system in Figure 9.3 is stable w.r.t. the perturbed plant (9.26) and (9.27), if and only if the nominal stability in Theorem 9.1 holds and the following structured singular value μ satisfies $\mu(\omega) < 1$ for any $\omega \in [0, \pi]$*

$$\mu = \frac{|1 - z^{-m_2}Q_2|\,|P_1C_1|\,|W_1|}{|1 + P_1C_1 + P_2C_2|} + \frac{|P_2C_2 + (1 + P_1C_1)\,z^{-m_2}Q_2|\,|W_2|}{|1 + P_1C_1 + P_2C_2|}. \tag{9.28}$$

Proof. Consider first the general closed-loop system for DISO plants under perturbation, as shown in Figure 9.7, wherein $\tilde{C}_i$'s are the equivalent feedback controller. Figure 9.7 can be transformed to the generalized representation in Figure 9.8. From μ-analysis, the closed-loop system is stable w.r.t. the plant perturbations if and only if G is stable and the structured singular value of G satisfies: $\forall\omega$, $\mu_\Delta\left(G(e^{j\omega})\right) < 1$. In Figure 9.8, consider the smallest (in the sense of H_∞ norm) perturbation Δ such that the following stability boundary is attained:

$$\det\left(I + \Delta(e^{j\omega})G(e^{j\omega})\right) = 0. \tag{9.29}$$

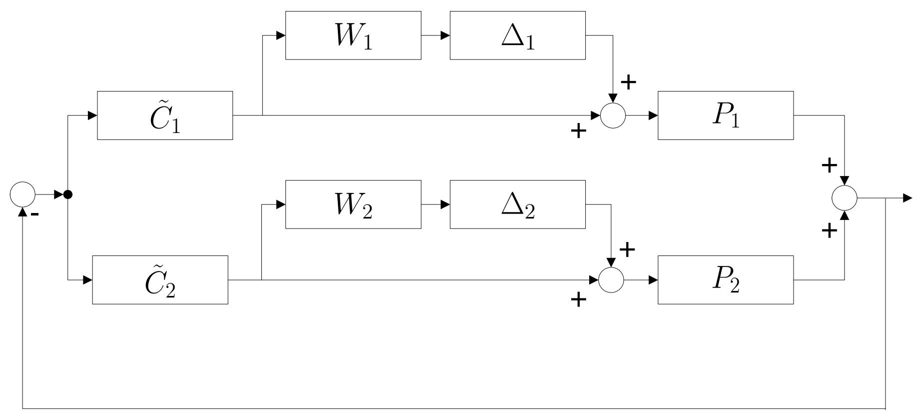


Figure 9.7: General closed-loop system for DISO plants under perturbations

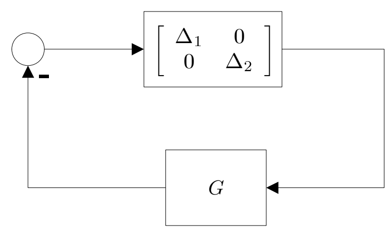


Figure 9.8: Generalized block diagram of Figure 9.7

After standard block-diagram analysis, the generalized plant G can be shown to be

$$G = \frac{W\tilde{C}P}{1 + P\tilde{C}} = \frac{1}{1 + P_1\tilde{C}_1 + P_2\tilde{C}_2} \begin{bmatrix} W_1\tilde{C}_1 \\ W_2\tilde{C}_2 \end{bmatrix} \begin{bmatrix} P_1 & P_2 \end{bmatrix}. \tag{9.30}$$

Substituting $\Delta = diag\{\Delta_1, \Delta_2\}$ and (9.30) to (9.29) yields (for simplified notation, the frequency index is omitted)

$$\begin{aligned} &\det(I + \Delta G) \\ =&\det\left(I + \frac{\begin{bmatrix} \Delta_1 & 0 \\ 0 & \Delta_2 \end{bmatrix}}{1 + P_1\tilde{C}_1 + P_2\tilde{C}_2} \begin{bmatrix} W_1\tilde{C}_1 \\ W_2\tilde{C}_2 \end{bmatrix} \begin{bmatrix} P_1 & P_2 \end{bmatrix} \right) \\ =&1 + \frac{P_1\Delta_1W_1\tilde{C}_1 + P_2\Delta_2W_2\tilde{C}_2}{1 + P_1\tilde{C}_1 + P_2\tilde{C}_2}, \end{aligned} \tag{9.31}$$

where the last equality has used the identity $\det(I + AB) = \det(I + BA)$.

Combining (9.31) and (9.29), the minimum-H_∞-norm perturbation is obtained when

$|\Delta_1| = |\Delta_2| =: |\Delta_0|$ and

$$1 - \left| \frac{P_1 W_1 \tilde{C}_1}{1 + P_1 \tilde{C}_1 + P_2 \tilde{C}_2} \right| |\Delta_0| - \left| \frac{P_2 W_2 \tilde{C}_2}{1 + P_1 \tilde{C}_1 + P_2 \tilde{C}_2} \right| |\Delta_0| = 0.$$

By definition, the structured singular value is

$$\mu = \frac{1}{|\Delta_0|} = \frac{\left| P_1 \tilde{C}_1 \right| |W_1| + \left| P_2 \tilde{C}_2 \right| |W_2|}{\left| 1 + P_1 \tilde{C}_1 + P_2 \tilde{C}_2 \right|}. \tag{9.32}$$

When DDOB is in the feedback loop as shown in Figure 9.6, we have $\tilde{C}_1 = C_1$ and $\tilde{C}_2 = C_{2,s}\left(C_2 + C_{2,p}\right)$, where $C_{2,s}$ and $C_{2,p}$ are respectively given by (9.9) and (9.10). Therefore, letting $\hat{P}_i = P_i$ and after simplifications, we get

$$\begin{aligned} \tilde{C}_1 &= C_1 \\ \tilde{C}_2 &= \frac{C_2 + (1 + P_1 C_1)\, z^{-m_2} P_2^{-1} Q_2}{1 - z^{-m_2} Q_2} \end{aligned}$$

which, combined with (9.32), gives the explicit form of (9.28).

□

Notice that μ is linear w.r.t. $|W_1|$ and $|W_2|$. Overall (9.28) indicates that in the regions where a good model is available for the plant, the structured singular value is small and we have flexible design freedom in $Q_2(z)$. If $e^{-m_2 j\omega} Q_2(e^{j\omega}) = 0$, DDOB is turned off at this frequency and (9.28) is simply the structured singular value of the baseline feedback system. This infers that the baseline system needs to be robustly stable. In the frequency region where $e^{-m_2 j\omega} Q_2(e^{j\omega})$ is close to unity, $\mu \approx |W_2|$ and the robust stability depends on the magnitude of model uncertainty of the secondary actuator in this region. Therefore, if $|W_2|$ goes beyond one at certain frequencies (particularly near Nyquist frequency), certainly one should not apply the model based DDOB here, and the magnitude of $Q(z)$ should be kept small.

9.6 Decoupled Sensitivity and DDOB for DISO systems

This section discuss a reduced-order implementation of DDOB. The main result is that, DDOB provides a natural enhancement scheme to the decoupled sensitivity (DS) feedback design idea, which is the most popular design technique in dual-stage HDDs (see, e.g., [115, 116, 117]).

In Figure 9.9, the DS controller has the transfer function

$$C(z) = \begin{bmatrix} C_v(z)\left(1 + \hat{P}_2(z) C_m(z)\right) \\ C_m(z) \end{bmatrix}$$

Here the sub-indexes v and m represent VCM and microactuator respectively. Consider first the baseline system formed only by $P_1(z)$, $P_2(z)$ and $C(z)$. The idea of decoupled sensitivity design is that, if $P_2(z) = \hat{P}_2(z)$, direct computation gives that

$$1 + P(z)C(z) = (1 + P_1(z)C_v(z))\,(1 + P_2(z)C_m(z))\,.$$

Consequently, the total sensitivity function $S(z) = 1/(1 + P(z)C(z))$ is decoupled to the cascaded connection of $1/(1 + P_1(z)C_v(z))$ and $1/(1 + P_2(z)C_m(z))$.

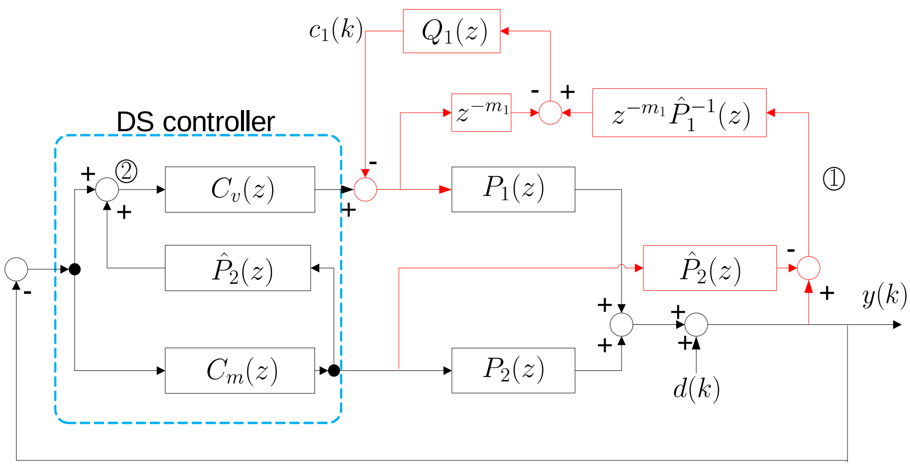


Figure 9.9: Control of dual-stage systems with DDOB and decoupled sensitivity

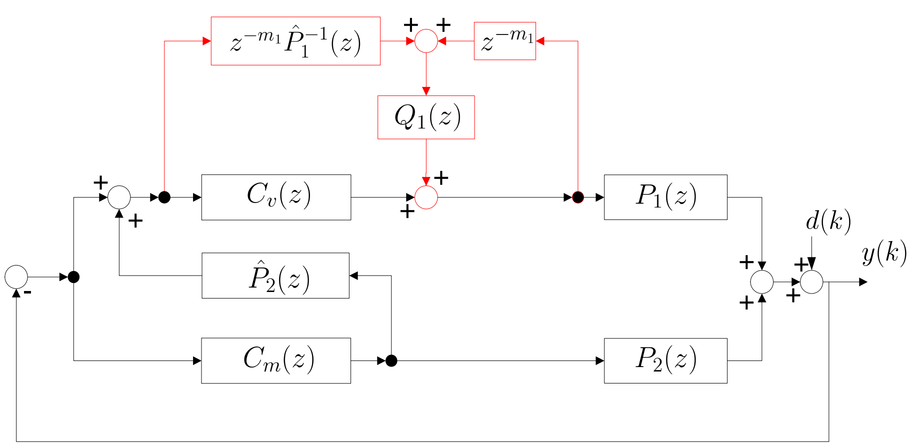


Figure 9.10: A reduced-order implementation of Figure 9.9

Consider the combined implementation of DS and DDOB for the first actuator in Figure 9.9. Notice that $\hat{P}_2(z)$ appears in both DDOB and the DS controller. In addition, since the reference is zero in regulation control, the signal at ① is the negative of the signal at

②. We can thus replace the input of $z^{-m_1}\hat{P}_1^{-1}(z)$ by the negative of the component at ②. With some block-diagram relocation and sign rearrangement, we can obtain in Figure 9.10, the equivalent realization of the block diagram in Figure 9.9. We have now saved the computation of one $\hat{P}_2(z)$ block. Figure 9.10 is simpler to implement and analyze as well.

9.7 Case Study: Control of Dual-stage HDDs

The proposed DDOBs are applied in this section to the dual-stage benchmark introduced in Section 3.2 on page 30. A set of disturbance data is obtained from audio-vibration experiments on an actual HDD. We will design DDOBs for both the Voice-Coil-Motor (VCM) and microactuator (MA) actuators. The former is denoted as VCM DDOB, and the latter as MA DDOB.

Two major resonances exist in the VCM plant, and are compensated via two notch filters at 3.0 kHz and 6.5 kHz. The 11-order resonance-compensated VCM model is treated as a generalized plant $P_1(z)$. $\hat{P}_1(z)$ is chosen to be a second-order transfer function that captures the friction mode at around 60 Hz in Figure 3.11. As the notch filters introduced some additional phase loss to $P_1(z)$, overall $\hat{P}_1(z)$ contains a two-step delay, i.e., $m_1 = 2$. Through the above design, both the magnitude and the phase of $P_1(z)$ are well captured by $\hat{P}_1(z)$ at frequencies up to around 6 kHz.

The microactuator also contains two resonances that are compensated by notch filters (at 6.5 kHz and 9.6 kHz). This actuator is a minimum-phase system (DC gain plus resonances) by nature, which simplifies the $\hat{P}_2(z)$ design. We directly model $\hat{P}_2(z)$ to include the resonances and have $m_2 = 1$.

The baseline feedback loop uses the decoupled sensitivity design in Section 9.6, with the magnitude responses of the decoupled sensitivities $1/(1+P_1(z)C_v(z))$ and $1/(1+P_2(z)C_m(z))$ plotted in Figure 9.11.

The flexibility of DDOB is explained in Figure 9.12, where we use the direct Q-design method in Section 5.4 [see (5.33) and (5.34) on page 73] to obtain

$$Q_1(z) = \frac{\left(\alpha^2 - 1 - a^2(\alpha - 1)\right) - (\alpha - 1)\,az^{-1}}{1 + \alpha a z^{-1} + \alpha^2 z^{-2}}$$
$$Q_2(z) = \frac{(\alpha - 1)\left(a + (1 + \alpha)\,z^{-1}\right)}{1 + \alpha a z^{-1} + \alpha^2 z^{-2}}$$

and vary the coefficient $a = -2\cos(2\pi\Omega_{Hz}T_s)$ to test the system performance at different frequencies. Six frequency values are evaluated, with the resulting sensitivity functions plotted in Figure 9.12. The first three results come from VCM DDOB, with $Q_2(z) = 0$; and $Q_1(z)$ centered at 500 Hz, 900 Hz, and 1500 Hz respectively. The remaining three are generated by MA DDOB, with $Q_1(z) = 0$; and $Q_2(z)$ centered at 2300 Hz, 3100 Hz, and 3900 Hz. It is observed that by simple alternation of one coefficient in the Q filters, the servo loop can be customized to a great extent.

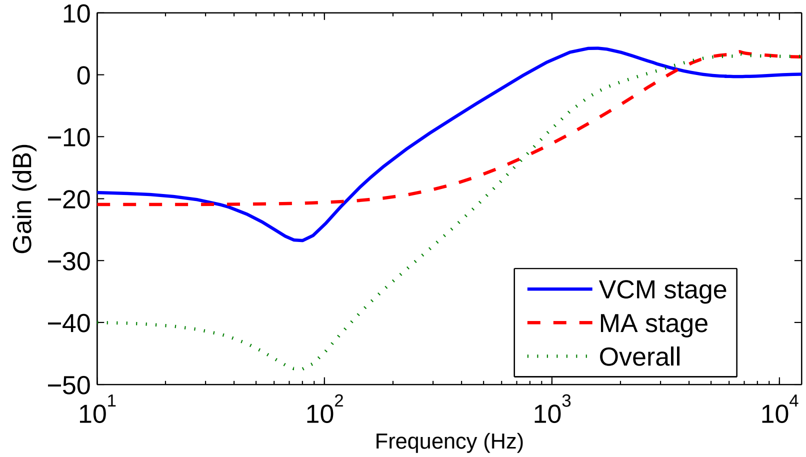


Figure 9.11: Magnitude responses in the decoupled-sensitivity design

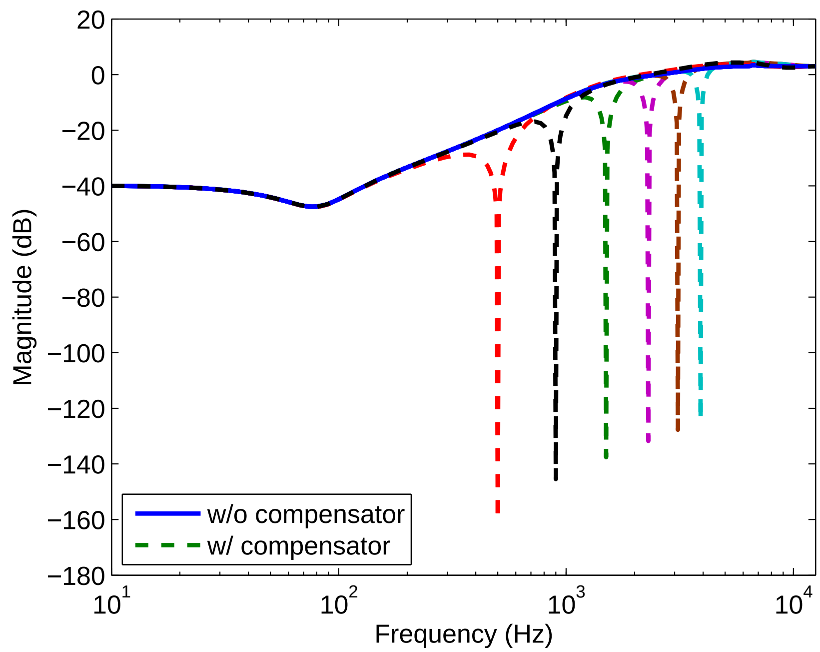


Figure 9.12: Magnitude responses of the sensitivity functions with different Q-filter configurations

For the application to audio-vibration rejection, Figures 9.13 and 9.14 present respectively the frequency- and time-domain PES signals. Two Q filters are used, one at 1200 Hz for the VCM DDOB, another at 2900 Hz for the MA DDOB. A scaling factor of 0.9 was used to slightly reduce the Q-filter gains to avoid excessive waterbed effect (see Chapter 6). It can be observed that the spectral peaks at the corresponding frequencies are significantly reduced by DDOBs, and that the magnitudes of position errors are decreased to half of the original values.

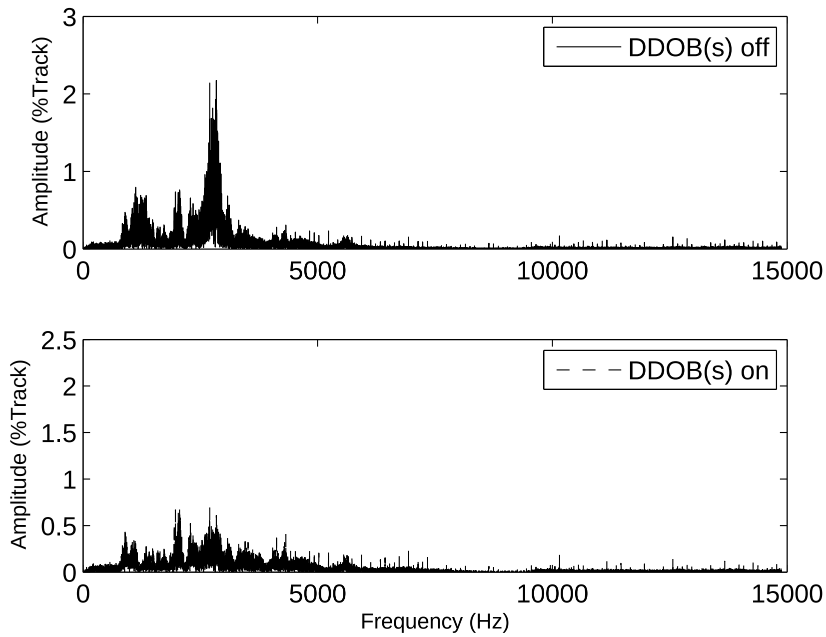


Figure 9.13: Spectra of the position error signals using a projected disturbance profile from actual experiments

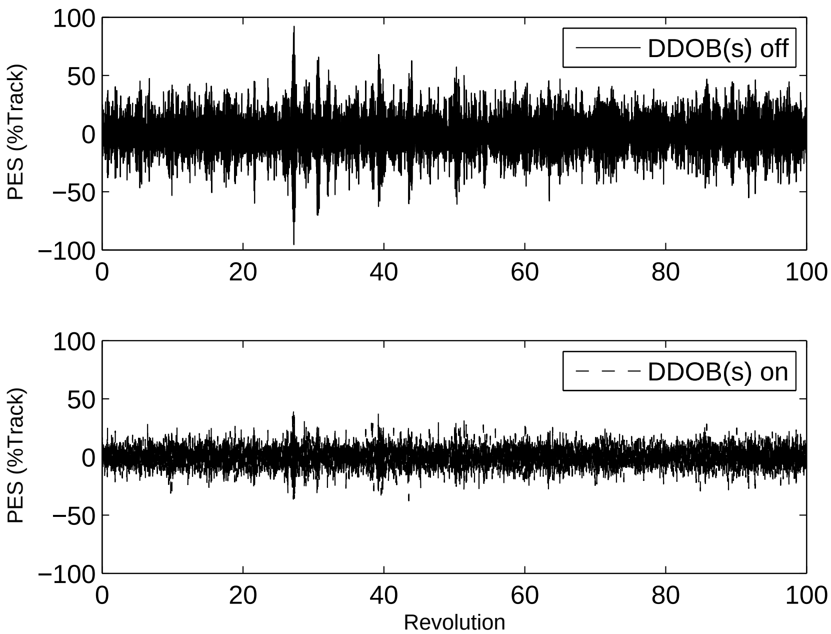


Figure 9.14: Time traces of the position error signals in Figure 9.13

Part III

Adaptive Band-limited Local Loop Shaping

Chapter 10

Parameter Adaptation Algorithms

10.1 Overview

To place enhanced servo at the desired frequency regions, we need knowledge of the spectral peaks in the error/disturbance signals. This is particularly true for loop-shaping applications such as Figures 1.3 to 1.5 on page 8. For tracking control, the trajectory may be pre-defined offline, and thus available for customized servo design. For regulation problems, however, spectral peaks of the disturbances can be unknown and even time-varying. In this chapter, we study approaches to automate the identification process of the spectral peaks.

The overall adaptive SISO pseudo Youla-Kucera parameterization scheme is shown in Figure 10.1. A similar construction can be made for the MISO problem in Chapter 9. The central adaptation algorithm will be the same for both cases.

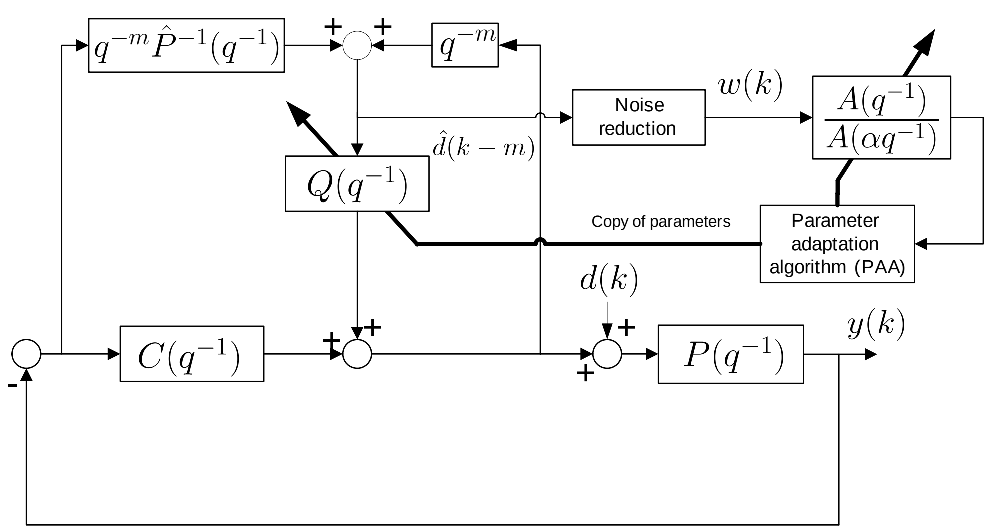


Figure 10.1: Structure of adaptive pseudo YK parameterization for regulation control

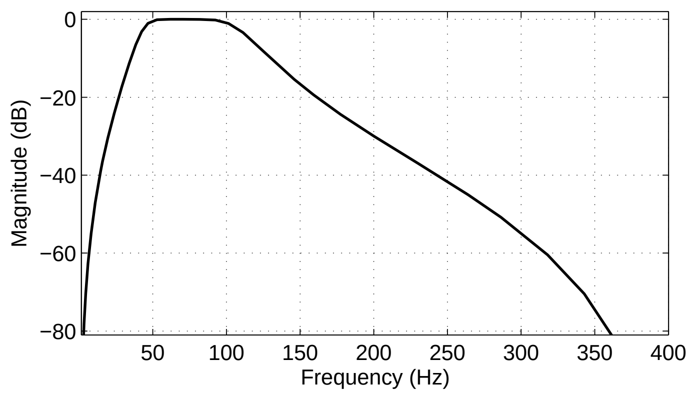


Figure 10.2: An example bandpass filter

Recall that the input to the Q filter is a delayed and noisy estimate of the actual disturbance. We can use this $\hat{d}(k-m)$ signal for parameter adaptation on $Q(q^{-1})$. To see the reason of the noise-reduction block in Figure 10.1, note that the residual output is

$$\begin{aligned} y(k) &= S\left(q^{-1}\right)P\left(q^{-1}\right)d(k) \\ &\approx \left(1-q^{-m}Q\left(q^{-1}\right)\right)\frac{P\left(q^{-1}\right)}{1+P\left(q^{-1}\right)C\left(q^{-1}\right)}d(k) \\ &\triangleq \left(1-q^{-m}Q\left(q^{-1}\right)\right)y_o(k) \end{aligned} \tag{10.1}$$

where $y_o(k)$ is the baseline output without the add-on pseudo YK compensation scheme. Using $P(q^{-1})/[1+P(q^{-1})C(q^{-1})]$ (or a nominal version of it) as the noise-reduction block makes $w(k)\approx y_o(k)$ and the adaptation hence directly focus on $y_o(k)$, which is invariant with respect to $Q(q^{-1})$.

Remark. Of course, if prior knowledge about a coarse region of the disturbances is available, additional filtering (using e.g., a bandpass filter similar to that in Figure 10.2) can be applied in the noise-reduction block. This will further improve the signal-to-noise ratio in the adaptation.

For the moment, we focus on the adaptation for narrow-band loop shaping, and assume the disturbance consists of n independent sinusoidal components with additive noises. Generalizations will be made in the next chapter.

For the considered class of problem, we have designed, in (10.1),

$$1-q^{-m}Q\left(q^{-1}\right)=\frac{A\left(q^{-1}\right)}{A_Q\left(q^{-1}\right)}K\left(q^{-1}\right)$$

where

$$
\begin{aligned}
A(q^{-1}) &= \prod_{i=1}^{n}\left(1-2\cos(\omega_i)q^{-1}+q^{-2}\right) &(10.2)\\
&= 1+a_1q^{-1}+\cdots+a_nq^{-n}+\cdots+a_1q^{-2n+1}+q^{-2n} &(10.3)\\
A_Q(q^{-1}) &= A(\alpha q^{-1})\\
&= \prod_{i=1}^{n}\left(1-2\alpha\cos(\omega_i)q^{-1}+\alpha^2q^{-2}\right) &(10.4)\\
&= 1+a_1\alpha q^{-1}+\cdots+a_n\alpha^nq^{-n}+\cdots+a_1\alpha^{2n-1}q^{-2n+1}+\alpha^{2n}q^{-2n} &(10.5)
\end{aligned}
$$

and $K(q^{-1})$ is [Section 5.4]

$$K\left(q^{-1}\right)=k_0+k_1q^{-1}+\ldots$$

or a structured IIR filter [Section 5.3]:

$$K\left(q^{-1}\right)=\sum_{i=1}^{m}\begin{pmatrix} m\\ i\end{pmatrix}\left[-\frac{A(q^{-1})}{A(\alpha q^{-1})}\right]^{i-1}. \tag{10.6}$$

In the special case where $m=1$, we have $K(q^{-1})=1$, and can simply construct adaptation based on $v(k)=A(q^{-1})/A(\alpha q^{-1})\,w(k)$. When $m>1$, the direct input-output dynamics in (10.1) is nonlinear w.r.t. the coefficients $[a_1,a_2,\ldots,a_n]^T$ or $[\omega_1,\omega_2,\ldots,\omega_n]^T$ in $Q(q^{-1})$. However, noticing that in

$$\overline{v}(k)=\left(1-q^{-m}Q\left(q^{-1}\right)\right)w(k)=K\left(q^{-1}\right)\frac{A(q^{-1})}{A(\alpha q^{-1})}w(k),$$

$K(q^{-1})$ is an auxiliary filter and the main shape of $1-q^{-m}Q(q^{-1})$ is from the notch filter $A(q^{-1})/A(\alpha q^{-1})$, we can thus still use

$$\boxed{v(k)=\frac{A(q^{-1})}{A(\alpha q^{-1})}w(k)} \tag{10.7}$$

as the model for adaptation. Indeed, in (10.6), $K(q^{-1})$ is nothing but a combination of normalized notch filter, and does not change the major loop shape outside the desired servo-enhancing frequencies.

Remark. Notice that if $w(k)$ is composed of pure sinusoidal signals and the notch filter $A(q^{-1})/A(\alpha q^{-1})$ is configured to have the correct center frequencies, then (10.7) will generate a null output $v(k)$.

The general adaptation algorithm has the iterative structure of

$$\hat{\theta}(k)=\hat{\theta}(k-1)+[\text{Adaptation gain}\times\text{Error}]. \tag{10.8}$$

Depending on whether we choose directly the filter coefficient

$$\theta = [a_1, a_2, \ldots, a_n]^T \tag{10.9}$$

or other generalized versions of the coefficients for the cascaded filter

$$\theta = [\omega_1, \omega_2, \ldots, \omega_n]^T, \tag{10.10}$$

different parameter adaptation algorithms (PAA) can be formed. We discuss the two cases for (10.9) and (10.10), respectively, in Section 10.2 and Section 10.3.

As $A(q^{-1})/A(\alpha q^{-1})$ and $Q(q^{-1})$ share the same parameter vector θ, after adaptation on $A(q^{-1})/A(\alpha q)^{-1}$, the identified $\hat{\theta}$ is directly implemented in $Q(q^{-1})$ as shown in Figure 10.1.

10.2 Adaptation of the Direct-filter Structure

The Adaptation Model and the Predictor

Equation (10.7) gives

$$A\left(\alpha q^{-1}\right) v(k) = A\left(q^{-1}\right) w(k). \tag{10.11}$$

Using (10.5) and (10.3), we get the direct difference equation of (10.11)

$$\begin{aligned} v(k) = {} & w(k) + w(k-2n) - \alpha^{2n} v(k-2n) + a_n \left[w(k-n) - \alpha^n v(k-n)\right] \\ & + \sum_{i=1}^{n-1} a_i \left[w(k-i) + w(k-2n+i) - \alpha^i v(k-i) - \alpha^{2n-i} v(k-2n+i)\right], \end{aligned} \tag{10.12}$$

which has the vector-form representation:

$$v(k) = \psi(k-1)^T \theta + \left(w(k) + w(k-2n) - \alpha^{2n} v(k-2n)\right). \tag{10.13}$$

with the parameter vector θ and the regressor vector $\psi(k-1)$ defined as

$$\theta = [a_1, a_2, \ldots, a_n]^T \tag{10.14}$$

$$\psi(k-1) = [\psi_1(k-1), \psi_2(k-1), \ldots, \psi_n(k-1)]^T \tag{10.15}$$

$$\begin{aligned} \psi_i(k-1) = {} & w(k-i) + w(k-2n+i) \\ & - \alpha^i v(k-i) - \alpha^{2n-i} v(k-2n+i); \quad i = 1, ..., n-1 \end{aligned} \tag{10.16}$$

$$\psi_n(k-1) = w(k-n) - \alpha^n v(k-n). \tag{10.17}$$

Consider now the adaptive version of (10.11):

$$\hat{A}\left(\alpha q^{-1}, \hat{\theta}(k)\right) \hat{v}(k) = \hat{A}\left(q^{-1}, \hat{\theta}(k)\right) w(k). \tag{10.18}$$

Replacing $\theta(k)$ and $v(k)$ with $\hat{\theta}(k)$ and $\hat{v}(k)$ respectively in (10.13), we get

$$\hat{v}(k) = \overline{\psi}(k-1)^T \hat{\theta}(k) + w(k) + w(k-2n) - \alpha^{2n}\hat{v}(k-2n). \tag{10.19}$$

Here $\hat{\theta}(k)$ is the estimate of θ; $\overline{\psi}(k-1)$ consists of

$$\begin{aligned} \overline{\psi}_i(k-1) &= w(k-i) + w(k-2n+i) \qquad (10.20)\\ &\quad - \alpha^i \hat{v}(k-i) - \alpha^{2n-i}\hat{v}(k-2n+i); \quad i = 1, ..., n-1, \\ \overline{\psi}_n(k-1) &= w(k-n) - \alpha^n \hat{v}(k-n). \qquad (10.21) \end{aligned}$$

Note that for an ideal $Q(q^{-1})$, the output $v(k)$ equals zero. The signal $\hat{v}(k)$ is thus an estimation error signal:

$$e(k) = \hat{v}(k) - 0 \tag{10.22}$$

yielding

$$e(k) = \phi(k-1)^T \hat{\theta}(k) + w(k) + w(k-2n) - \alpha^{2n} e(k-2n) \tag{10.23}$$

with

$$\begin{aligned} \phi(k-1) &= [\phi_1(k-1), \ldots, \phi_n(k-1)]^T \\ \phi_i(k-1) &= w(k-i) + w(k-2n+i) \qquad (10.24)\\ &\quad - \alpha^i e(k-i) - \alpha^{2n-i} e(k-2n+i); \quad i = 1, ..., n-1 \\ \phi_n(k-1) &= w(k-n) - \alpha^n e(k-n). \qquad (10.25) \end{aligned}$$

$e(k)$ and $\hat{\theta}(k)$ are respectively the *a posteriori* estimation error and the *a posteriori* parameter estimate. Replacing $\hat{\theta}(k)$ by $\hat{\theta}(k-1)$ in (10.23), we obtain the *a priori* estimation error

$$e^o(k) = -(-\phi(k-1))^T \hat{\theta}(k-1) + \left(w(k) + w(k-2n) - \alpha^{2n} e(k-2n)\right). \tag{10.26}$$

At time k, *before* computation of the up-to-date parameter estimate $\hat{\theta}(k)$, $e^o(k)$ (rather than $e(k)$) is the available error from the latest information $\hat{\theta}(k-1)$.

The application of *a posteriori* information in (10.26) is essential for adaptation in noisy environments. Notice that $e(k)$ is more accurate than $e^o(k)$, since the former is updated by a more recent coefficient vector $\hat{\theta}(k)$.

Directly applying $e^o(k)$ to update $\hat{\theta}(k)$ will yield a PAA that requires

$$\frac{1}{A(\alpha q^{-1})} - \frac{1}{2}$$

to be strictly positive real (SPR) for adaptation stability. This condition is usually challenging to satisfy in practice. We apply instead the PAA using the output error predictor with a fixed compensator, which greatly relaxes the stability condition and provide good performance in noisy environments [118].

The idea is to apply a filtered version of $e(k)$ as the adaptation error, to update $\hat{\theta}(k)$ (in the *a posteriori* sense). The fixed compensator $C(q^{-1})$ is proposed to be given by

$$C\left(q^{-1}\right) = 1 + c_1\alpha q^{-1} + \cdots + c_n\alpha^n q^{-n} + \cdots + c_1\alpha^{2n-1}q^{-2n+1} + \alpha^{2n}q^{-2n}, \tag{10.27}$$

i.e., by replacing every a_i in $A(\alpha q^{-1})$ by c_i. We will discuss shortly the reasons for this choice.

The *a posteriori* adaptation error $\epsilon(k)$ is therefore

$$\epsilon(k) = C\left(q^{-1}\right)e(k) = e(k) + \alpha^{2n}e(k-2n) + \varphi(k-1)^T\theta_c. \tag{10.28}$$

where

$$\begin{aligned}
\theta_c &= [c_1, c_2, \ldots, c_n]^T, \\
\varphi(k-1) &= [\varphi_1(k-1), \varphi_2(k-1), \ldots, \varphi_n(k-1)]^T, \\
\varphi_i(k-1) &= \alpha^i e(k-i) + \alpha^{2n-i}e(k-2n+i);\ i = 1, ..., n-1, \\
\varphi_n(k-1) &= \alpha^n e(k-n).
\end{aligned}$$

At time instant k, the previous $e(k-i)$ is available $\forall i = 1, \ldots 2n$. Yet from (10.19), $e(k)$ can only be updated after $\hat{\theta}(k)$ has been obtained. The implementable *a priori* adaptation error is given by

$$\epsilon^0(k) = e^0(k) + \alpha^{2n}e(k-2n) + \varphi(k-1)^T\theta_c. \tag{10.29}$$

Estimation of $\hat{\theta}(k)$ can now be performed through the following PAA:

$$\hat{\theta}(k) = \hat{\theta}(k-1) + \frac{F(k-1)(-\phi(k-1))\epsilon^0(k)}{1 + \phi(k-1)^T F(k-1)\phi(k-1)}, \tag{10.30}$$

$$F(k) = \frac{1}{\lambda(k)}\left[F(k-1) - \frac{F(k-1)\phi(k-1)\phi^T(k-1)F(k-1)}{\lambda(k) + \phi^T(k-1)F(k-1)\phi(k-1)}\right], \tag{10.31}$$

where $F(k)$ and $\lambda(k)$ are respectively the adaptation gain and the forgetting factor. $\lambda(k)$ is a positive real number no larger than one, and can be designed depending on the nature of the process [55].

Remark. It may appear strange at first, to have a negative sign in front of $\phi(k-1)$ *in* (10.30)*. We explain next why this had happened and how it may be transformed to a standard form. Instead of* (10.22)*, we can define the estimation error from the other way around:*

$$\bar{e}(k) = v(k) - \hat{v}(k) = -\hat{v}(k) \tag{10.32}$$

which is more conventionally adopted. Equation (10.32) *gives, after substitution to* (10.19)*–*(10.25)*,*

$$\bar{e}(k) = -\phi(k-1)^T\hat{\theta}(k) - w(k) - w(k-2n) - \alpha^{2n}\bar{e}(k-2n) \tag{10.33}$$

with

$$\begin{aligned}\phi_i(k-1) &= w(k-i) + w(k-2n+i) \\ &\quad + \alpha^i \bar{e}(k-i) + \alpha^{2n-i}\bar{e}(k-2n+i); \quad i = 1, ..., n-1 \\ \phi_n(k-1) &= w(k-n) + \alpha^n \bar{e}(k-n).\end{aligned}$$

After noting $e(k) = -\bar{e}(k)$, *we can see that* (10.23) *and* (10.33) *are equivalent. If we had followed the notation of* (10.32), *we will then have* $\bar{\epsilon}(k) = C(q^{-1})\bar{e}(k) = -\epsilon(k)$ *and* $\bar{\epsilon}^o(k) = -\epsilon^o(k)$, *yielding*

$$-\phi(k-1)\epsilon^o(k) = \phi(k-1)\bar{\epsilon}^o(k).$$

Thus (10.30) *can be equivalently represented as*

$$\hat{\theta}(k) = \hat{\theta}(k-1) + \frac{F(k-1)\phi(k-1)\bar{\epsilon}^0(k)}{1 + \phi(k-1)^T F(k-1)\phi(k-1)},$$

which is in the standard PAA form. The reason for the difference in notation is that in (10.23), $\partial e(k)/\partial\hat{\theta}(k) = \phi(k-1)$ *while in* (10.33), $\partial\bar{e}(k)/\partial\hat{\theta}(k) = -\phi(k-1)$. *Before filtering by the compensator* $C(q^{-1})$, *the two PAAs aim at minimizing in the following negative directions of the gradient vector:*

$$-\nabla_{\hat{\theta}(k)}\left\{\frac{1}{2}e^2(k)\right\} = -\frac{\partial\frac{1}{2}e^2(k)}{\partial\hat{\theta}(k)} = -e(k)\frac{\partial e(k)}{\partial\hat{\theta}(k)} \tag{10.34}$$

$$-\nabla_{\hat{\theta}(k)}\left\{\frac{1}{2}\bar{e}^2(k)\right\} = -\bar{e}(k)\frac{\partial\bar{e}(k)}{\partial\hat{\theta}(k)}. \tag{10.35}$$

For (10.23), (*10.34*) *equals* $-\phi(k-1)e(k)$, *while for* (10.33), (*10.35*) *equals* $\phi(k-1)\bar{e}(k)$.

In practice, to apply Equations (10.30)-(10.31), one computes $\phi(k-1)$ from (10.24) to (10.25); gets $e(k)$ and $e^o(k)$ from (10.23) and (10.26); and then obtains $\epsilon^o(k)$ from (10.29).

Due to the fact that the more accurate *a posteriori* $e(k)$ is used in the adaptation, the above PAA maintains good performance when the adaptation input $w(k)$ contains additional noise terms. The parameters converge to their true values if the initial conditions are not far away from the global minima. The algorithm is subjected to the following sufficient but not necessary stability condition [118]: the transfer function

$$\frac{C(q^{-1})}{A(\alpha q^{-1})} - \frac{1}{2} \tag{10.36}$$

should be SPR.

For the design of $C(q^{-1})$, it is common practice in the field of system identification, to apply first some stable adaptation algorithm and assign the resulting rough estimate of

$A(\alpha q^{-1})$ to $C(q^{-1})$. Under such situations, $C(q^{-1})$ is close to $A(\alpha q^{-1})$, (10.36) is approximately $1/2$, and thus SPR. The design of the compensator in (10.27) was meant for this reason.

More specific, to obtain θ_c, the coefficients of $C(q^{-1})$, we propose the series-parallel predictor[1] in the next subsection. This type of predictor is always stable and actually provides unbiased estimation when the adaptation input is noise free [118].

Initialization with a Series-Parallel Predictor

Recall the adaptation model in (10.13), that

$$v(k) = \psi(k-1)^T \theta + \big(w(k) + w(k-2n) - \alpha^{2n} v(k-2n)\big). \tag{10.37}$$

If we do not use the *a posteriori* term $e(k)$ but only replace θ by $\hat{\theta}(k-1)$, we get the *a priori* prediction of $v(k)$:

$$\hat{v}^o(k) = \psi(k-1)^T \hat{\theta}(k-1) + \big(w(k) + w(k-2n) - \alpha^{2n} v(k-2n)\big). \tag{10.38}$$

The *a priori* prediction error is given by

$$e^o(k) = v(k) - \hat{v}^o(k). \tag{10.39}$$

$e^o(k)$ will be applied as the adaptation source. It can be seen that the series-parallel predictor is much more simplified compared to the parallel predictors. The computation can be further reduced by the observation that for a *tuned* $Q(q^{-1})$, the ideal output $v(k)$ is zero. This simplifies the regression vector to:

$$\psi_i(k-1) = w(k-i) + w(k-2n+i); \quad i = 1, ..., n-1 \tag{10.40}$$

$$\psi_n(k-1) = w(k-n), \tag{10.41}$$

and the estimation error in (10.39) now becomes

$$e^o(k) = -\psi(k-1)^T \hat{\theta}(k-1) - (w(k) + w(k-2n)). \tag{10.42}$$

With the above information, the following RLS type parameter adaptation algorithm (PAA) can be constructed:

$$\hat{\theta}(k) = \hat{\theta}(k-1) + \frac{F(k-1)\psi(k-1)e^o(k)}{1 + \psi(k-1)^T F(k-1)\psi(k-1)} \tag{10.43}$$

$$F(k) = F(k-1) - \frac{F(k-1)\psi(k-1)\psi(k-1)^T F(k-1)}{1 + \psi(k-1)^T F(k-1)\psi(k-1)}. \tag{10.44}$$

[1]Correspondingly, the predictors in Section 10.2 are known as the parallel predictors, for their use of the a *posteriori* information.

The adaptation algorithm is always stable. The proof is provided in Appendix A.5 and [6]. [6] also provides a different (and slightly less complex) derivation of the PAA.

When parameters have converged for the above PAA, the value of $\hat{\theta}$ is applied to initialize $C(q^{-1})$ and $\hat{Q}(q^{-1})$. It will be shown in later sections, that this transient period of running the series-parallel prediction is actually quite small, and we can move very quickly to the high-performance parallel predictor.

Dynamic Switching Between the Two-stage Adaptation

This section discusses how to connect the two developed PAAs. Conventionally, switching between two sets of predictors is usually done by choosing a fixed transition time instant. As for the confirmation of parameter convergence, common choices are to monitor either the adaptation gain or the adaptation error. For the former case, since the adaptation gain $F(k)$ is in general a matrix, some operation such as taking the trace is needed to judge the size of $F(k)$. For the latter, monitoring the adaptation error is well-suited for processes with high signal-to-noise ratio. However, when the parameters are biased or the adaptation algorithm has local convergence, it is difficult to decide the threshold for convergence. We propose an algorithm to automate the switching by directly focusing on the values of the parameter estimate $\hat{\theta}(k)$. It is considered that $\hat{\theta}(k)$ has converged if

1. $\max\left(|\delta\hat{\theta}(k)|\right) \triangleq \max\left(|\hat{\theta}(k) - \hat{\theta}(k-1)|\right)$ is less than a pre-defined tolerance.

2. condition 1) holds continuously for a number of samples.

One can remark that the proposed algorithm is an approximation of the Cauchy criterion [119] for convergence. Note also, that the tolerance in condition 1) is directly related to the parameters and is much easier to choose than the threshold of the adaptation gain or that of the adaptation error.

Obtaining the Frequencies from the Identified Parameters

The frequencies and the identified parameters are mapped by

$$\prod_{i=1}^{n}\left(1 - 2\cos(\omega_i)q^{-1} + q^{-2}\right) = 1 + a_1 q^{-1} + \cdots + a_n q^{-n} + \cdots + a_1 q^{-2n+1} + q^{-2n}. \quad (10.45)$$

For the simplest case where $n = 1$, we have $a_1 = -2\cos\omega_1$, from which we can compute $\omega_1 = 2\pi\Omega_1 T_s$, where the unit of Ω_1 is Hz. The parameter a_1 is online updated and Ω_1 can be computed offline for algorithm tuning.

For $n > 1$, as $(1 - 2\cos(\omega_i)q^{-1} + q^{-2}) = (1 - e^{j\omega_i}q^{-1})(1 - e^{-j\omega_i}q^{-1})$, the values of ω_i can be computed offline via calculating the angle of the complex roots of

$$1 + a_1 q^{-1} + \cdots + a_n q^{-n} + \cdots + a_1 q^{-2n+1} + q^{-2n} = 0.$$

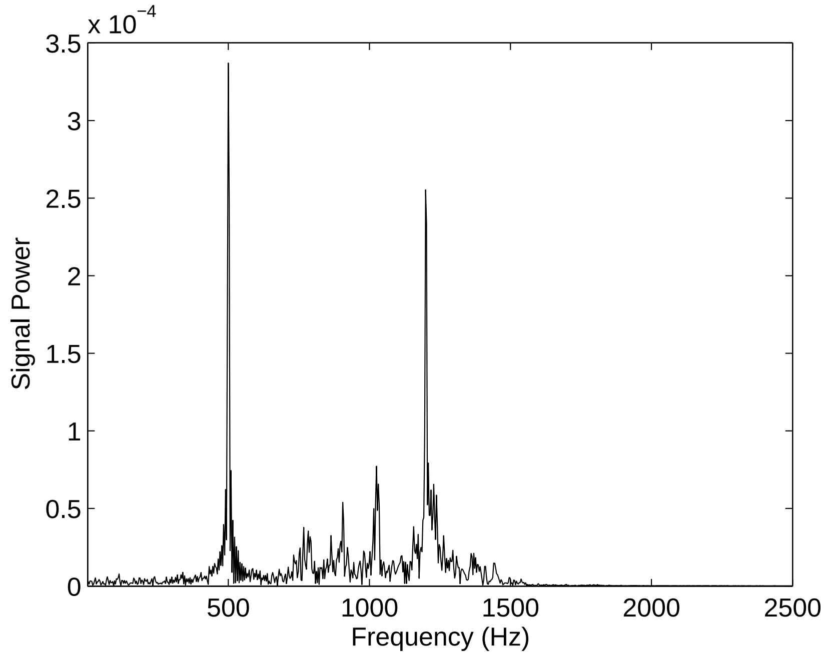


Figure 10.3: Spectra of the adaptation input

Parameter Initialization

When initializing $\hat{\theta}(0)$, we can first estimate the disturbance frequencies $\{\hat{\omega}_i(0)\}_{i=1}^{n}$, and then expand the product on the left hand side of (10.45), to obtain $\{\hat{a}_i(0)\}_{i=1}^{n}$ for implementation in the PAA.

An Example

Figures 10.3 to 10.4 show details of an application of the PAA to the HDD benchmark in Section 3.2. The input to the adaptation, namely, the signal $w(k)$, has the spectrum shown in Figure 10.3, where two major peaks and several side peaks are present.

Figure 10.4 shows the identified frequencies using the two-stage adaptation scheme. The first stage, the series-parallel predictor, was turned on at around 1.25-th revolution (1 revolution = 8.33 ms). The dynamic switching algorithms then operated to find a continuous 30-sample window where all online identified coefficients did not have abrupt changes in values. We can see that at around 1.35-th revolution, the first stage converged close to the actual frequency values. At around 1.48-th revolution, the auto-switching algorithm made the transition, from the series-parallel predictor to the more accurate parallel predictor, and gave unbiased estimation.

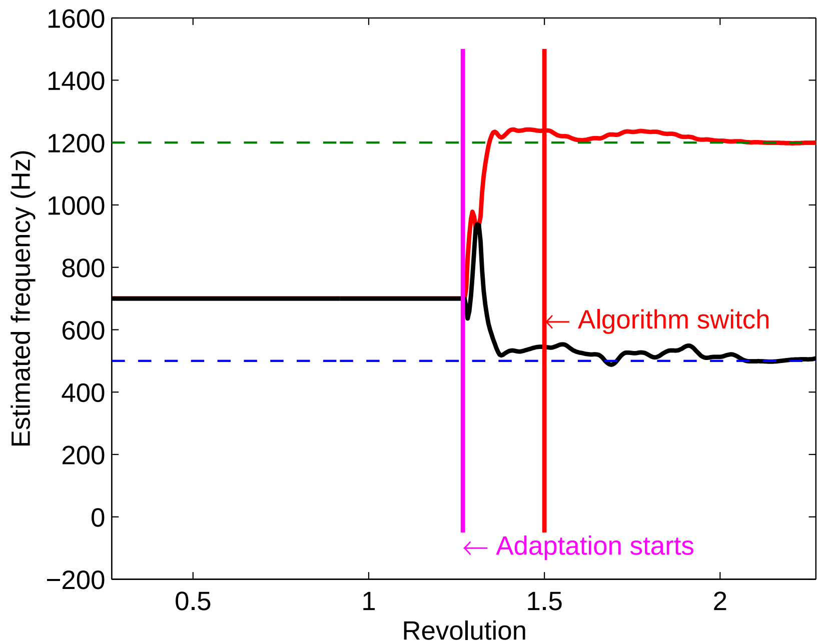


Figure 10.4: Adaptive identification of two vibrations at 500 and 1200 Hz

10.3 Adaptation of the Cascaded-filter Structure

Instead of using the direct filter structure $A\left(q^{-1}\right) = 1 + a_1 q^{-1} + \cdots + a_n q^{-n} + \cdots + a_1 q^{-2n+1} + q^{-2n}$ in $A\left(q^{-1}\right)/A\left(\alpha q^{-1}\right)$, in this section, we discuss adaptation on the cascaded filter structure:

$$y\left(k\right) = H\left(q^{-1}\right) w\left(k\right), \ H\left(q^{-1}\right) = \frac{A\left(\beta q^{-1}\right)}{A\left(\alpha q^{-1}\right)} = \prod_{i=1}^{n} H_o\left(\omega_i, q^{-1}\right), \tag{10.46}$$

where

$$H_o\left(\omega_i, q^{-1}\right) = \frac{1 - 2\beta q^{-1}\cos\omega_i + \beta^2 q^{-2}}{1 - 2\alpha q^{-1}\cos\omega_i + \alpha^2 q^{-2}}. \tag{10.47}$$

Here, we have made the structure of the notch filter more general, with β not necessarily being one, but $\alpha < \beta \leq 1$.

Remark. Cascaded filters are very popular in digital signal processing. One main advantage of using this filter structure is that it is numerically more efficient and stable for high-order filters [120].

Updating directly ω_i gave rise to [121], which was later refined in [122]. We notice, however, from the control aspect of view, that the value of ω_i is not directly needed ($\cos\omega_i$ is the term really implemented in the controller). To directly estimate ω_i, [121] and [122] needed to calculate trigonometric functions within each iteration, which can be expensive or inconvenient in applications where the computation power is limited. We introduce $\theta_i = \cos(\omega_i)$, and construct adaptation as follows:

In (10.47), letting

$$A_o\left(\theta_i, \gamma q^{-1}\right) \triangleq 1 - 2\gamma q^{-1}\theta_i + \gamma^2 q^{-2},\ \gamma = \alpha, \beta \tag{10.48}$$

and introducing the unknown parameter vector $\theta = [\theta_1, \theta_2 \cdots \theta_n]^T$, we can simplify (10.46) to

$$H\left(\theta, q^{-1}\right) = \prod_{i=1}^{n} H_o\left(\theta_i, q^{-1}\right) = \prod_{i=1}^{n} \frac{A_o\left(\theta_i, \beta q^{-1}\right)}{A_o\left(\theta_i, \alpha q^{-1}\right)}, \tag{10.49}$$

where we used the notation $H\left(\theta, q^{-1}\right)$ to emphasize that this is a transfer function with unknown parameter θ. The objective of the PAA is to find the best parameter estimate, such that the following cost function is minimized

$$V_k = \sum_{j=1}^{k} \frac{1}{2}\left[e^o\left(j\right)\right]^2, \tag{10.50}$$

where $e^o\left(k\right) = H\left(\theta, q^{-1}\right) w\left(k\right)$ is the output error.

The transfer function $H\left(\theta, q^{-1}\right)$ is nonlinear in θ. To find the best estimation, the celebrated Gauss-Newton Recursive Prediction Error Method (RPEM) (Chapter 11 in [55]) suggests to apply the following iterative formulas

$$\hat{\theta}\left(k\right) = \hat{\theta}\left(k-1\right) + \frac{F(k-1)\psi(k-1)e^o(k)}{\lambda\left(k\right) + \psi^T(k-1)F(k-1)\psi(k-1)} \tag{10.51}$$

$$F(k) = \frac{1}{\lambda\left(k\right)}\left[F(k-1) - \frac{F(k-1)\psi(k-1)\psi^T(k-1)F(k-1)}{\lambda\left(k\right) + \psi^T(k-1)F(k-1)\psi(k-1)}\right], \tag{10.52}$$

where $\psi(k-1) = [\psi_1(k-1), \ldots, \psi_n(k-1)]^T$, $\psi_i(k-1) = -\partial e^o\left(k\right)/\partial\hat{\theta}_i\left(k-1\right)$ is the i-th element of the regressor vector (in the negative direction of the gradient of $e^o\left(k\right)$), and $\lambda\left(k\right)$ is the forgetting factor.

The above modified algorithm has several nice properties:

1. stability of the Gauss-Newton RPEM is guaranteed if $H(\hat{\theta}\left(k\right), q^{-1})$ is stable during the adaptation [55], which can be easily checked by monitoring if $|\hat{\theta}_i\left(k\right)| < 1$, due to our cascaded construction of (10.49).

2. $\hat{\theta}\left(k\right)$ unbiasedly converges to a local minimum [55].

3. it inherits most of the advantages of [47, 122], such as fast convergence, computational efficiency, and numerical robustness. Moreover, it does not require computing sine and cosine functions.

Algorithm

Similar to [122], for obtaining first $e^o(k) = H(\theta, q^{-1})\, w(k)$, we introduce

$$x_j(k) = \prod_{i=1}^{j} H_o\left(\theta_i, q^{-1}\right) w(k), \tag{10.53}$$

from which we have

$$x_j(k) = \frac{1 - 2\beta\theta_i q^{-1} + \beta^2 q^{-2}}{1 - 2\alpha\theta_i q^{-1} + \alpha^2 q^{-2}} x_{j-1}(k), \tag{10.54}$$

i.e., in state-space representation

$$Z_i(k+1) = \left[\begin{array}{c|c} 2\alpha\theta_i & -\alpha^2 \\ \hline 1 & 0 \end{array}\right] Z_i(k) + \left[\begin{array}{c} 1 \\ \hline 0 \end{array}\right] x_{i-1}(k) \tag{10.55}$$

$$x_i(k) = [2(\alpha - \beta)\theta_i, \beta^2 - \alpha^2] Z_i(k) + x_{i-1}(k). \tag{10.56}$$

We can then iteratively get $e^o(k)$, with $e^o(k) = x_n(k)$ and $x_0(k) = w(k)$.

The regressor vector: To get $\psi_i(k-1) = -\partial e^o(k) / \partial\hat{\theta}_i(k-1)$, we notice that

$$\begin{aligned}
\frac{\partial e^o(k)}{\partial\hat{\theta}_i(k-1)} &= \frac{\partial H\left(\hat{\theta}_i(k-1), q^{-1}\right) w(k)}{\partial\hat{\theta}_i(k-1)} \\
&= \frac{\partial H_o\left(\hat{\theta}_i(k-1), q^{-1}\right)}{\partial\hat{\theta}_i(k-1)} \prod_{j\neq i} H_o\left(\hat{\theta}_j(k-1), q^{-1}\right) w(k) \\
&= \frac{\partial H_o\left(\hat{\theta}_i(k-1), q^{-1}\right)}{\partial\hat{\theta}_i(k-1)} H_o^{-1}\left(\hat{\theta}_i(k-1), q^{-1}\right) e^o(k).
\end{aligned} \tag{10.57}$$

Using (10.48) and (10.49), we get

$$\frac{\partial H_o(\theta_i, q^{-1})}{\partial\theta_i} = \frac{\frac{\partial A_o\left(\theta_i, \beta q^{-1}\right)}{\partial\theta_i} A_o(\theta_i, \alpha q^{-1})}{A_o^2(\theta_i, \alpha q^{-1})} - \frac{A_o(\theta_i, \beta q^{-1}) \frac{\partial A_o\left(\theta_i, \alpha q^{-1}\right)}{\partial\theta_i}}{A_o^2(\theta_i, \alpha q^{-1})},$$

where $\partial A_o(\theta_i, \gamma q^{-1}) / \partial\theta_i = -2\gamma q^{-1},\ \gamma = \alpha, \beta$.

Substituting the above back to (10.57), and changing θ_i to its estimated value $\hat{\theta}_i(k-1)$, we arrive at the following simple formula:

$$\psi_i(k-1) = -2\left[e_{F_i}(\beta, k) - e_{F_i}(\alpha, k)\right], \tag{10.58}$$

where $e_{F_i}(\gamma, k) = \gamma q^{-1} / A_o\left(\hat{\theta}_i(k-1), \gamma q^{-1}\right) e^o(k),\ \gamma = \alpha, \beta$, which can again be calculated using a state-space realization

$$W_i(\gamma, k+1) = \left[\begin{array}{c|c} 2\alpha\hat{\theta}_i(k) & -\gamma^2 \\ \hline 1 & 0 \end{array}\right] W_i(\gamma, k) + \left[\begin{array}{c} 1 \\ \hline 0 \end{array}\right] e^o(k) \tag{10.59}$$

$$e_{F_i}(\gamma, k) = \left[\begin{array}{cc} \gamma, & 0 \end{array}\right] W_i(\gamma, k). \tag{10.60}$$

Implementation

The recursive parameter estimation is finally summarized as follows:

Initialization: $\alpha_o = 0.8$, $\alpha_{end} = 0.995$, $\alpha_{rate} = 0.99$,[2] $\beta = 0.9999$, $Z_i(0) = W_i(\gamma, 0) = 0$, $F(0) \approx 100/E[e^o]^2 \cdot I$, $\hat{\theta}(0) =$ initial guess of the parameters, $\lambda(0) = \lambda_0$, $\lambda(\infty) = \lambda_{end}$, $\lambda_{rate} = 0.99$.

Main loop: for $k = 1, 2, \ldots$

step 1, prediction error computation: for $i = 1 : n$

$$x_i(k) = [2(\alpha - \beta)\hat{\theta}_i(k-1), \beta^2 - \alpha^2] Z_i(k) + x_{i-1}(k), \tag{10.61}$$

with $x_0(k) = w(k)$ and $e^o(k) = x_n(k)$.

step 2, regressor vector computation: for $i = 1 : n$

$$e_{F_i}(\gamma, k) = \begin{bmatrix} \gamma, & 0 \end{bmatrix} W_i(\gamma, k), \ \gamma = \alpha, \beta \tag{10.62}$$

$$\psi_i(k-1) = -2\left(e_{F_i}(\beta, k) - e_{F_i}(\alpha, k)\right). \tag{10.63}$$

step 3, parameter update using (10.51) and (10.52).

step 4, projection of unstable parameters: for $i = 1 : n$, if $|\hat{\theta}_i(k)| > 1$, $\hat{\theta}_i(k) = \hat{\theta}_i(k-1)$.

step 5, *a posteriori* prediction error $\bar{e}(k)$ computation and state vector update: for $i = 1 : n$

$$\bar{x}_i(k) = \left[2(\alpha - \beta)\hat{\theta}_i(k), \beta^2 - \alpha^2\right] Z_i(k) + \bar{x}_{i-1}(k) \tag{10.64}$$

$$Z_i(k+1) = \left[\begin{array}{c|c} 2\alpha\hat{\theta}_i(k) & -\alpha^2 \\ \hline 1 & 0 \end{array}\right] Z_i(k) + \left[\begin{array}{c} 1 \\ \hline 0 \end{array}\right] \bar{x}_{i-1}(k), \tag{10.65}$$

with $\bar{x}_o(k) = w(k)$ and $\bar{e}(k) = \bar{x}_n(k)$.

$$W_i(\gamma, k+1) = \left[\begin{array}{c|c} 2\alpha\hat{\theta}_i(k) & -\gamma^2 \\ \hline 1 & 0 \end{array}\right] W_i(\gamma, k) + \left[\begin{array}{c} 1 \\ \hline 0 \end{array}\right] \bar{e}(k), \tag{10.66}$$

for $\gamma = \alpha, \beta$.

step 6, forgetting factor and notch filter shape coefficient update: replace α by $\alpha_{end} - [\alpha_{end} - \alpha]\,\alpha_{rate}$, and

$$\lambda(k+1) = \lambda_{end} - [\lambda_{end} - \lambda(k)]\,\lambda_{rate}. \tag{10.67}$$

Remark. Similar to Section 10.2, the *a posteriori* information is applied to improve the estimation precision. As long as the initial parameter guesses are not too far away from the true values, the estimation is unbiased in noisy environments [55].

[2] α is designed to increase exponentially from α_o to α_∞, at the rate of α_{rate}, such that the notches get sharper and sharper to better capture the narrow-band frequencies.

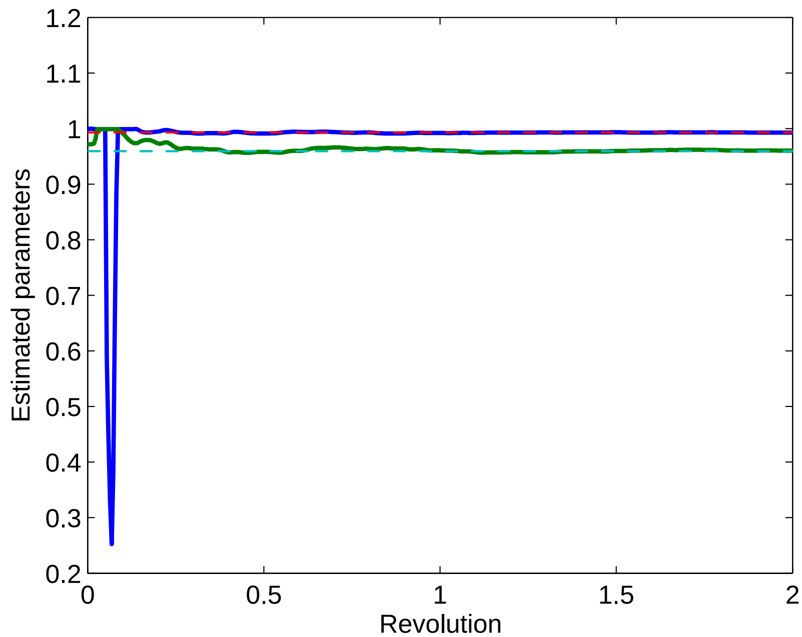


Figure 10.5: Online identified parameters for the two main spectral peaks in Figure 10.3

An Example

Consider the same adaptation input as shown in Figure 10.3 on page 162. Figure 10.5 shows the online identification of the parameters $\hat{\theta}_1$ and $\hat{\theta}_2$. Figure 10.6 shows the equivalent online frequency estimation, via the transformation $\hat{\Omega}_1 = \cos^{-1}(\hat{\theta}_1)/(2\pi T_s)$ and $\hat{\Omega}_2 = \cos^{-1}(\hat{\theta}_2)/(2\pi T_s)$. The convergence speed is again very fast, and the transient period is much less than one revolution of the disk rotation (1 revolution = 83.3 ms).

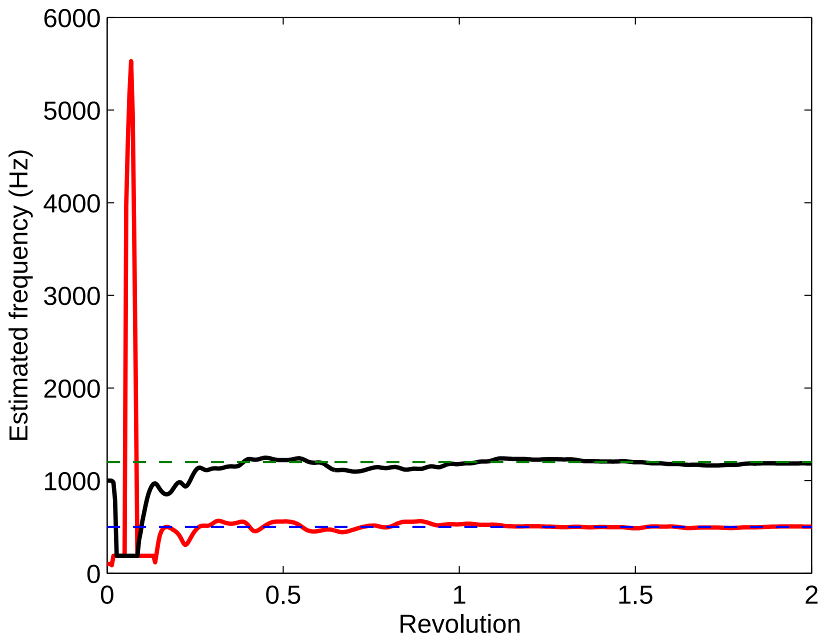


Figure 10.6: Frequency values computed from the data in Figure 10.5

10.4 Notes and Additional Discussions

The main results in this chapter are based on theories of system identification and adaptive control. The notation and equation style follow those of [55, 118, 56].

For more details of algorithm implementation, readers can refer to, respectively, [3, 7], [6], and [9], for applications of the PAAs in Sections 10.2, 10.2, and 10.3.

Order of Adaptation and Parameter Convergence

In the derivations of this chapter, we have assumed that the number of the adaptation parameters in $\{\theta_i\}_{i=1}^n$ is the same as the number of frequency components in the actual disturbance signal. The choice of the size of $\{\theta_i\}_{i=1}^n$ (which we denote as the order of adaptation) requires designers to have some engineering judgment during implementation. If there are more than n, say $r\,(> n)$, narrow-band signals in $w\,(k)$, the parameters will converge to a local optimal point[3] which corresponds to minimizing n components of the signal $w\,(k)$. This result is demonstrated via the example shown in Figure 10.7. If, on the contrary, $r < n$, then we will still be able to identify r frequency components in $w\,(k)$. Notice however, that in this case the adaptation model is over-determined and the optimum parameter estimate will not be unique. For the cascaded adaptation where $\theta_i = \cos\,(\omega_i)$,

[3]The location of the local optima depends on the initialization of the parameter estimate.

$n-r$ values of $\{\theta_i\}_{i=1}^n$ will have no determined convergence points. For the direct adaptation where $\theta_i = a_i$, we have

$$1+a_1\alpha q^{-1}+\cdots+a_n\alpha^n q^{-n}+\cdots+a_1\alpha^{2n-1}q^{-2n+1}+\alpha^{2n}q^{-2n} = \prod_{i=1}^{r}\left(1-2\alpha\cos(\omega_i)q^{-1}+\alpha^2 q^{-2}\right)$$
$$\times \prod_{i=r+1}^{n}\left(1-2\alpha\cos(\omega_i)q^{-1}+\alpha^2 q^{-2}\right). \quad (10.68)$$

For the last $n-r$ terms on the right hand side of (10.68), the frequency coefficients ω_i will also be undetermined.

Overall Closed-loop Stability with the PAAs

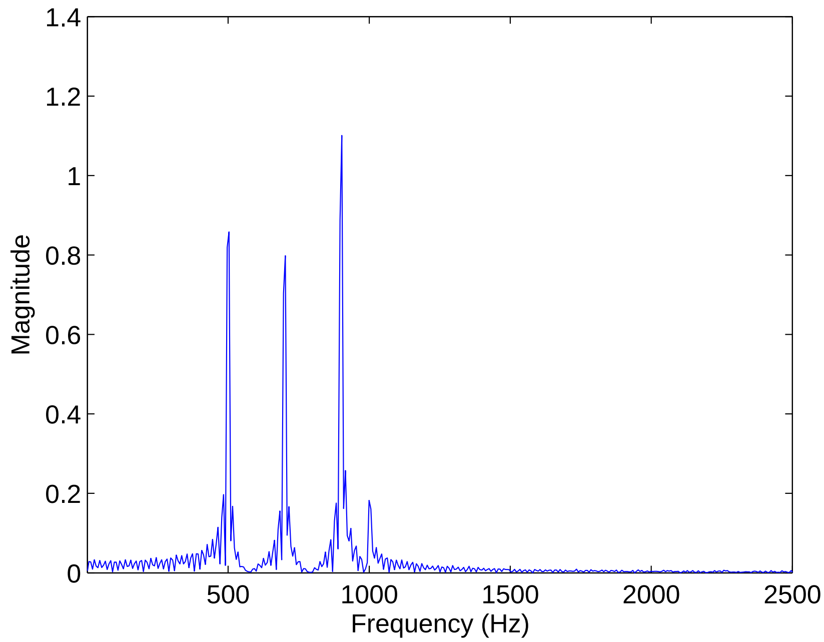
The stability and SPR conditions in this chapter apply only to the parameter adaptation algorithms. For the overall closed-loop system in Figure 10.1, we have discussed the stability condition in Chapter 4: given that $Q(z)$ is stable, if we have a good model match between $\hat{P}$ and P, then stability is guaranteed in the framework of Youla-Kucera parameterization.

Notice that the poles of the Q filter are the roots of $A(\alpha q^{-1})$, which satisfies

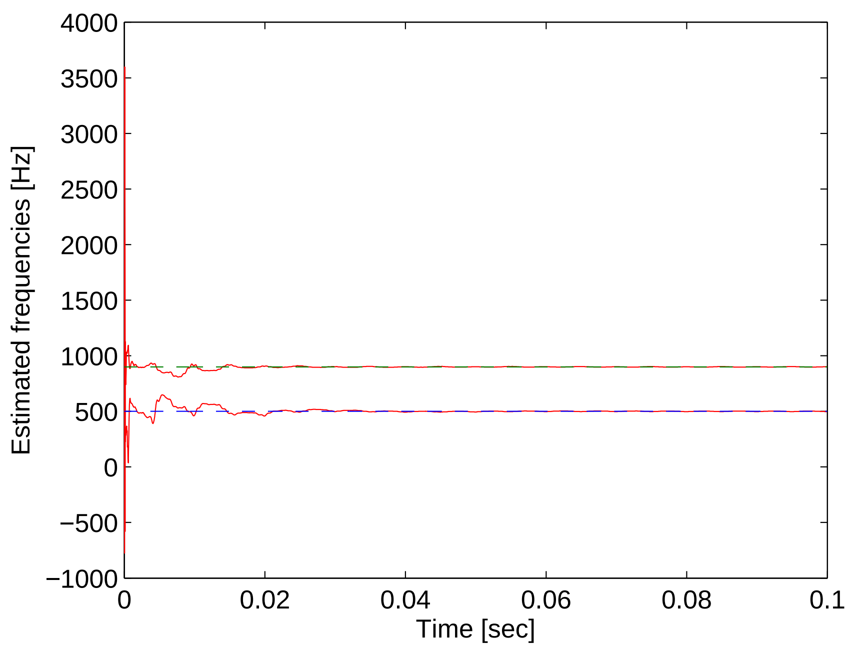
$$A(\alpha q^{-1}) = 1+a_1\alpha q^{-1}+\cdots+a_n\alpha^n q^{-n}+\cdots+a_1\alpha^{2n-1}q^{-2n+1}+\alpha^{2n}q^{-2n} \quad (10.69)$$
$$= \prod_{i=1}^{n}\left(1-2\alpha\cos(\omega_i)q^{-1}+\alpha^2 q^{-2}\right) \quad (10.70)$$
$$= \prod_{i=1}^{n}\left(1-\alpha e^{j\omega_i}q^{-1}\right)\left(1-\alpha e^{-j\omega_i}q^{-1}\right)$$

For the cascaded-filter adaptation in Section 10.3, we directly adapt $\theta_i = \cos(\omega_i)$, and have the stability-monitoring step to make sure $\left|\hat{\theta}_i\right| < 1$. Hence at each time instance, stability of $Q(z)$ is guaranteed for the YK parameterization.

For the direct-filter adaptation, $\{a_i\}_{i=1}^n$ are estimated. Notice that we have designed $\alpha \in (0,1)$. When the order of adaptation matches the number of frequency components in the actual disturbance, the parameters will converge towards the true values corresponding to (10.70), yielding the stability of $Q(z)$. Simulation and experimental verifications show that the roots of $Q(z)$ are also stable during the adaptation transient. Actually, as $\alpha \longrightarrow 0$, all roots of $A(\alpha q^{-1})$ converge to the origin, and the Q filter becomes an FIR. In practice, it may however be preferred not to turn on the compensation when the filter parameters are rapidly changing. In this case, designers can delay the Q-filter implementation for a few steps during the initial adaptation transient.



(a) Spectrum of actual disturbance



(b) Identified frequencies for $n = 2$ in the adaptation model $A\left(q^{-1}\right)/A\left(\alpha q^{-1}\right)$

Figure 10.7: Narrow-band disturbance identification example: adaptation input has more frequency components than the adaptation model in the cascaded-filter adaptation

Chapter 11

Implementations of Adaptive Local Loop Shaping

11.1 Remarks About Practical Implementation

We have discussed the general design concepts of LLS, pseudo YK parameterization, and adaptive Q filters. In this section, we provide several notes that are relevant to practical implementation.

Algorithm Tuning

For filters in the structure of $A\left(q^{-1}\right)/A\left(\alpha q^{-1}\right)$, the width of the notch shape is determined by the parameter α. When α is smaller than but very close to one, $A\left(q^{-1}\right)/A\left(\alpha q^{-1}\right)$ is almost one except at the center frequencies of the notch shape. The closed-loop robust stability will then be quite easy to satisfy. In practice, it is suggested to always start with an α that is close to one, and gradually reduce it to reach the desired attenuation bandwidth, without violating the stability conditions.

Order of the Internal Model

By construction, the proposed PAAs in Chapter 10 have good robustness against adaptation noise. It is yet always beneficial to apply as most engineering experiences as possible, and obtain a close estimation of n, the number of narrow-band disturbances. If indeed no information on n is available, several *series-parallel* predictors (which are low-cost in computation compared to the parallel predictors) in Section 10.2 can be run as shown in Figure 11.1, where a group of predictors with different orders are applied for the same adaptation input. We can then compare the resulted estimation errors (see an example in Figure 11.2), and choose the smallest PAA order that brings the error down to a satisfactory level.

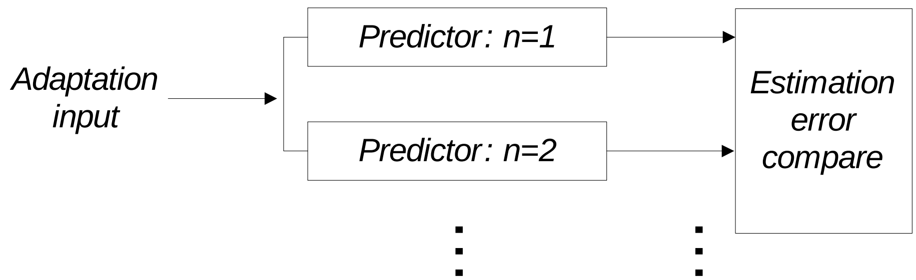


Figure 11.1: Flow chart to determine the number of spectral peaks online

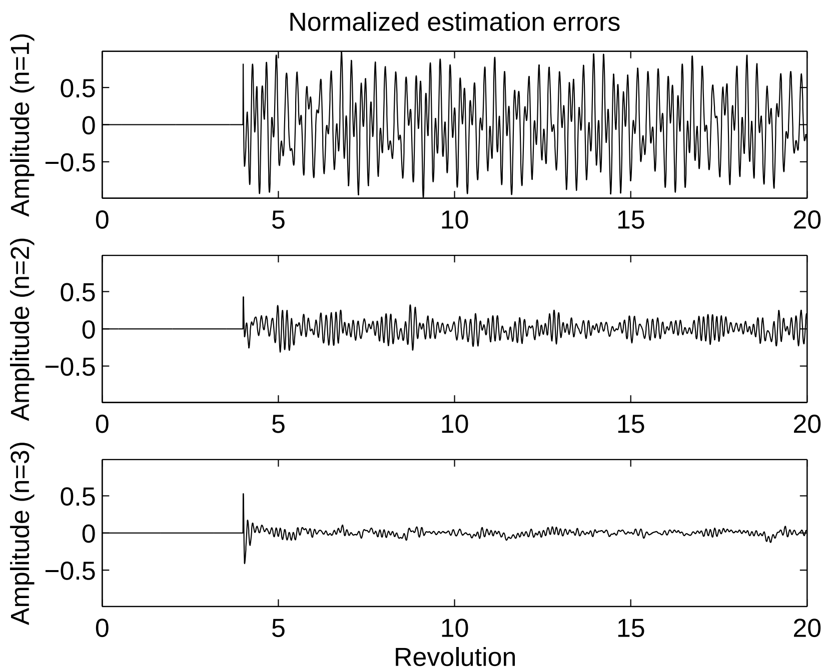


Figure 11.2: An example for identifying the number of multiple narrow-band disturbances

Disturbance Detection

The adaptive compensator may not need to be run for the entire time. We can include a disturbance detector to provide instructions on the timing of the adaptation. As an example, in HDD servo control, every effort is made towards reducing the Track Mis-Registration ($TMR = 3\sigma$ value of the position error, where σ denotes the standard deviation). Figure 11.3 presents one example disturbance detector, where the TMR is calculated online in a moving window of 20 samples. When online TMR exceeds the normal operation bound (in this example, 15% Track Pitch), the detector reports alert to the central adaptive compensator and turns on the adaptation. Figure 11.3 is the online result of an application example to be discussed in Section 11.2. We see that large error occurs after 1 revolution and is quickly reduced to be below 15% Track Pitch. This is the result of fast disturbance rejection after the adaptive compensation scheme is turned on.

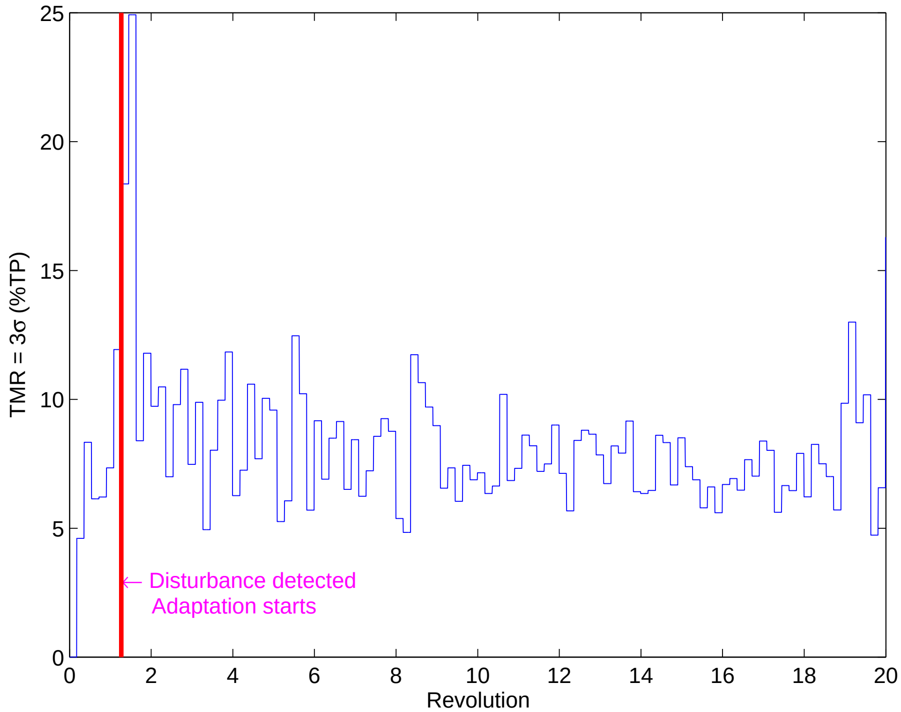


Figure 11.3: Online TMR monitoring in an HDD example

The Plant and the Resonances

One may notice that, in several of the application examples discussed so far, we have chosen to include notch filters as part of the plant. In this way, the "notched system" has a much smoother frequency response, and can be approximated by a low-order nominal model $\hat{P}(z)$. It should however be noted that notch filters will inevitably introduce more phase delays to the plant, thus increasing the difficulty of loop-shaping algorithms.

In the proposed YK parameterization schemes, it is also an option to model the real plant (without the notch filters) accurately when constructing $\hat{P}(z)$. We would then have a higher-order model, but with less plant delays and easier Q designs.

If we are interested in servo enhancement at frequencies much below the resonant frequencies, the effect of phase change due to notch filters will be insignificant. The performance difference of the two modeling schemes will then be quite small. Designers can make a selection based on the particular servo requirement and the specific implementation cost.

Time-varying Disturbances

In the general parameter adaptation algorithm

$$\begin{aligned}\hat{\theta}(k+1) =& \hat{\theta}(k) + [\text{Adaptation gain} \times \text{Error}]\\ F(k+1) =& \frac{1}{\lambda(k)}\left[F(k) - \frac{F(k)\,\phi(k)\,\phi(k)^T F(k)}{\lambda(k) + \phi(k)^T F(k)\,\phi(k)}\right],\end{aligned}$$

it is well-known that altering the forgetting factor is the key for time-varying parameter adaptation [55, 118]. For a non-unity $\lambda(k) = \lambda$, the overall objective at time k changes from

$$\min \sum_{i=1}^{k} \text{Error}^2$$

to

$$\min \sum_{i=1}^{k} \lambda^{k-i}\text{Error}^2.$$

A smaller $\lambda(k)\,(\in [0,1])$ thus gives faster forgetting of the history errors. We suggest the following tuning rules for disturbances whose frequencies change w.r.t. time:

(i) rapid initial convergence: initialize $\lambda(k)$ to be exponentially increasing from λ_0 to 1 in the first several samples, where λ_0 can be taken to be between $[0.92, 1)$

(ii) disturbances with sudden parameter changes: for such applications, when the prediction error encounters a sudden increase, indicating a change of disturbance characteristics, reduce $\lambda(k)$ to a small value $\underline{\lambda}$ (e.g. 0.92), and then increase it back to its steady-state value $\overline{\lambda}$ (usually close to 1), using the formula $\lambda(k) = \overline{\lambda} - \lambda_{\text{rate}}(\overline{\lambda} - \lambda(k-1))$, with λ_{rate} preferably in the range $(0.9, 0.995)$.

(iii) disturbances with continuously changing parameters: in this case, keep $\lambda(k)$ strictly smaller than one, using e.g., a constant $\lambda(k)(\in [0,1])$.

11.2 Application: Adaptive Rejection of Narrow-band and Audio Vibrations in HDDs

We discuss next application of the SISO pseudo Youla-Kucera parameterization, with direct Q design [Section 5.4] and direct-filter adaptation [Section 10.2], to the single-stage HDD simulation benchmark in Section 3.2.

The frequency response of the plant is as shown in Figure 3.9 on page 32. The solid line of Figure 11.4 presents the sensitivity function of the baseline feedback system.

Two narrow-band disturbances at 500 Hz and 1200 Hz were injected at the end of the first revolution to test the system performance. The upper plot in Figure 11.5a shows the time

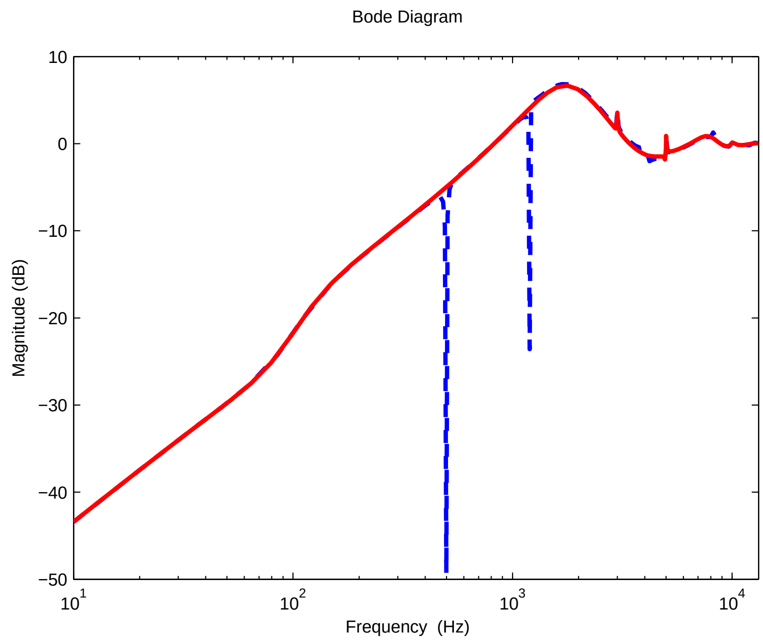


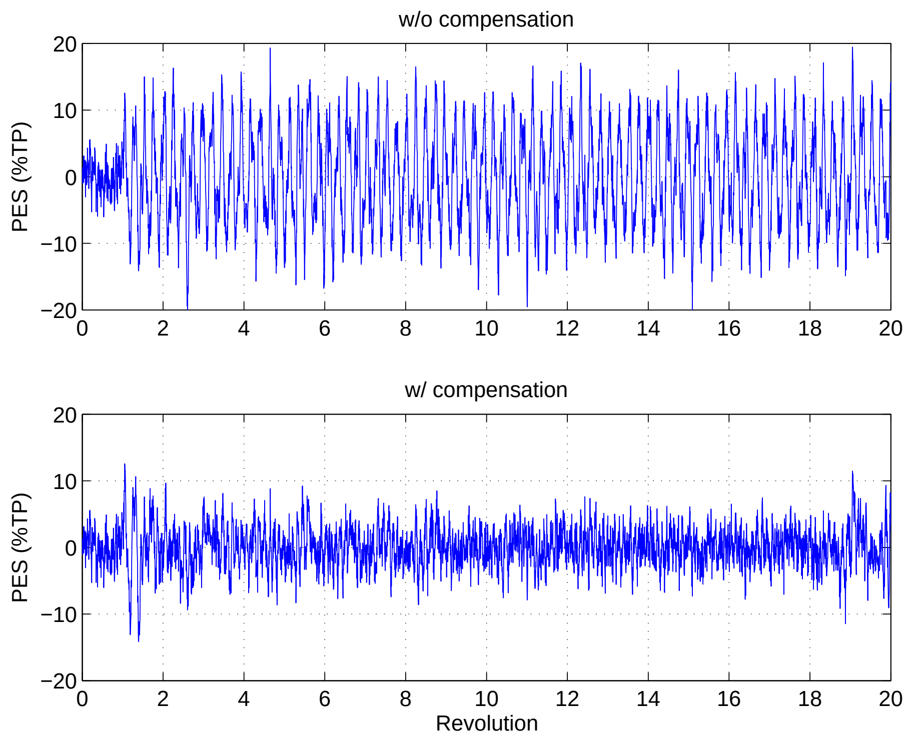
Figure 11.4: Frequency responses of the sensitivity functions with and without compensation

trace of the position error signal (PES) without adaptive pseudo YK parameterization. The corresponding steady-state frequency spectrum is plotted in the dotted line in Figure 11.5b. It is observed that with only the baseline feedback controller, the PES had strong energy components at 500 Hz and 1200 Hz; the TMR was 21.56% Track Pitch (TP). Notice that 1200 Hz is higher than the bandwidth of the baseline servo system. Without compensation, disturbance at this frequency was actually amplified.

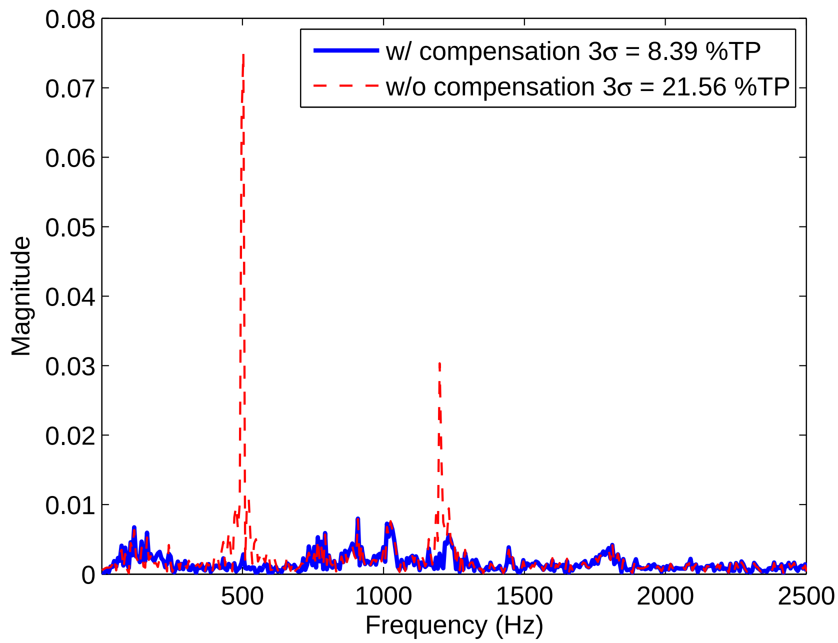
We have seen how the disturbance detector works in Figure 11.3. After successful disturbance detection, adaptation was turned on at 1.25-th revolution. Figure 11.6 shows the identified disturbance frequencies.[1] At around 1.48-th revolution, the series-parallel predictor automatically switched to the more accurate parallel predictor, and the disturbance compensation was turned on.

The frequency- and time-domain compensation results are shown in Figures 11.4, 11.5b, and 11.5a. After convergence, the closed-loop sensitivity function had the frequency response as shown in the dotted line of Figure 11.4. Correspondingly in Figure 11.5b, the major narrow-band components were fully attenuated, while the frequency spectra in other regions were hardly influenced. In the time domain, the lower plot in Figure 11.5a demonstrates the PES time trace with compensation, where we note the PES dropped to be significantly below 10% TP. The corresponding TMR (after convergence) was reduced to 8.39% TP–less than one half of the value without compensation.

[1] Recalled from Figure 10.4 on page 163.



(a) Time traces of PES with and without compensation



(b) Frequency spectra of PES with and without compensation

Figure 11.5: Time- and frequency-domain PES reduction

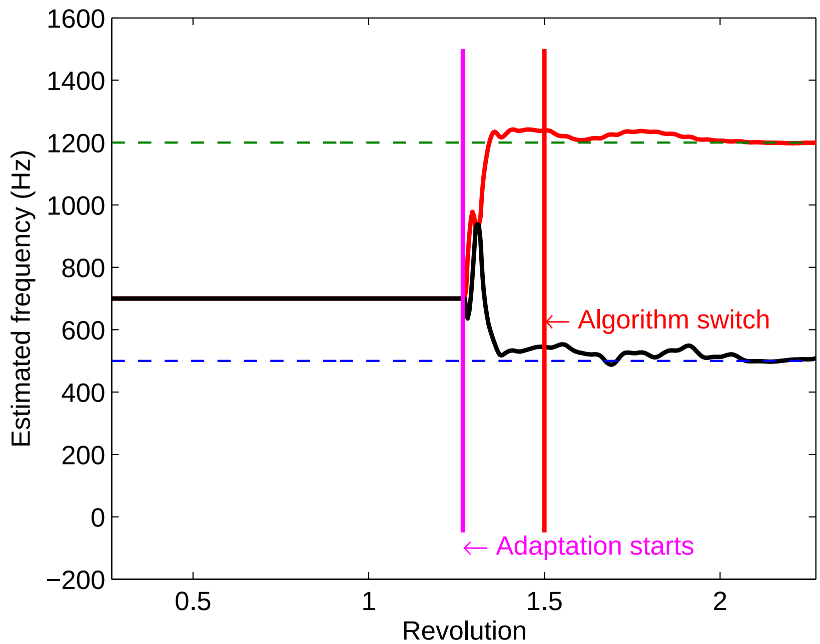
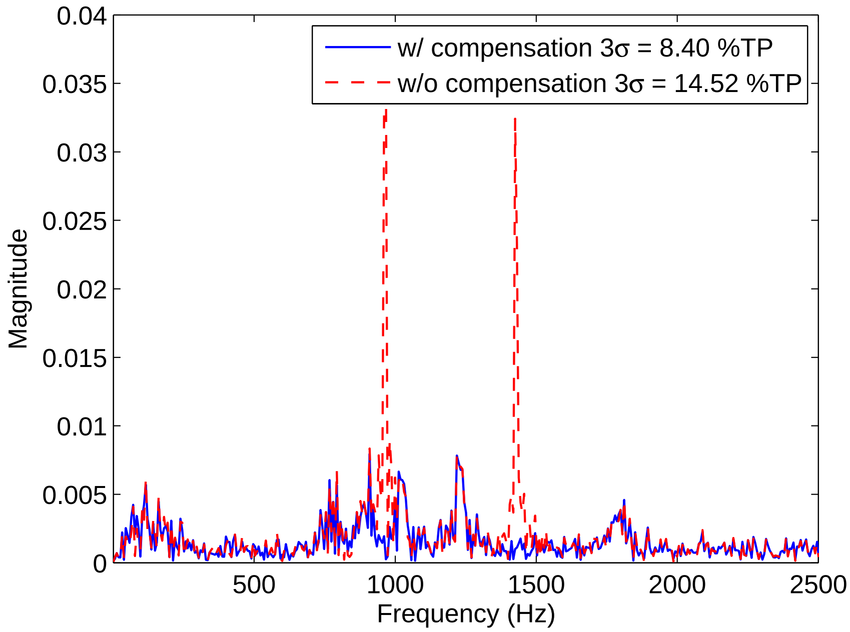


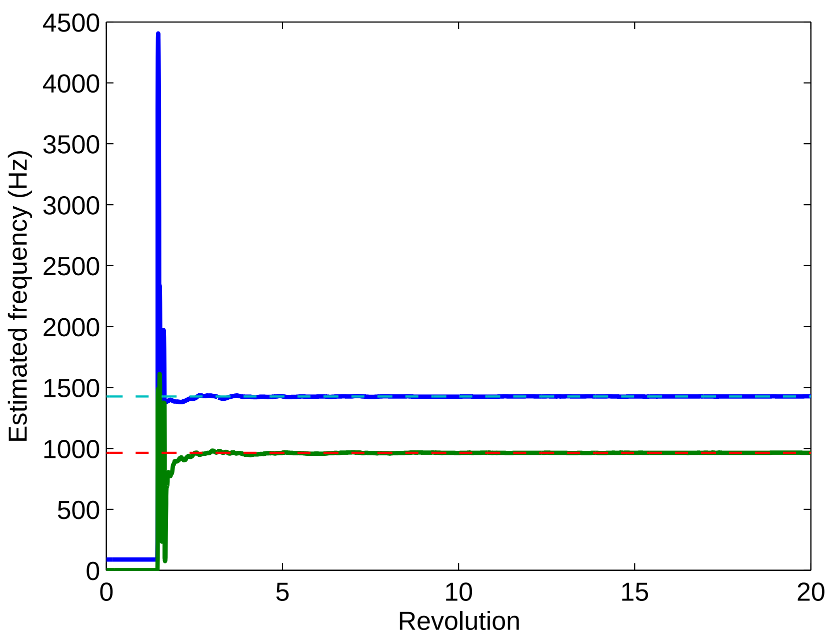
Figure 11.6: Identification of the disturbance frequencies in Figure 11.5a

Figure 11.7a and Figure 11.7b show the performance of the algorithm at higher frequencies and with the parameters initialized far away from their true values. Notice that in this case study, more small spectral peaks appear as input noise to the PAA (see Figure 11.7a), yet the algorithm found and fully rejected the two major disturbance components.

In Figure 11.4, we used an α very close to one for narrow-band disturbance rejection. The adaptation also works for wider-band disturbances, by choosing a smaller value of α (see an example loop shape in Figure 6.2 on page 93). Figures 11.8a and 11.8b show the performance of the algorithm for rejecting the audio vibrations discussed in Section 3.2. It is clear that the proposed algorithm correctly located and attenuated the spectral peaks of the vibrations. The influence of the large error peaks are indeed significant to the system: after removal of them, the 3σ value of the errors reduced from 24.18 to 13.14, namely, we have a 45.66 percent performance gain.

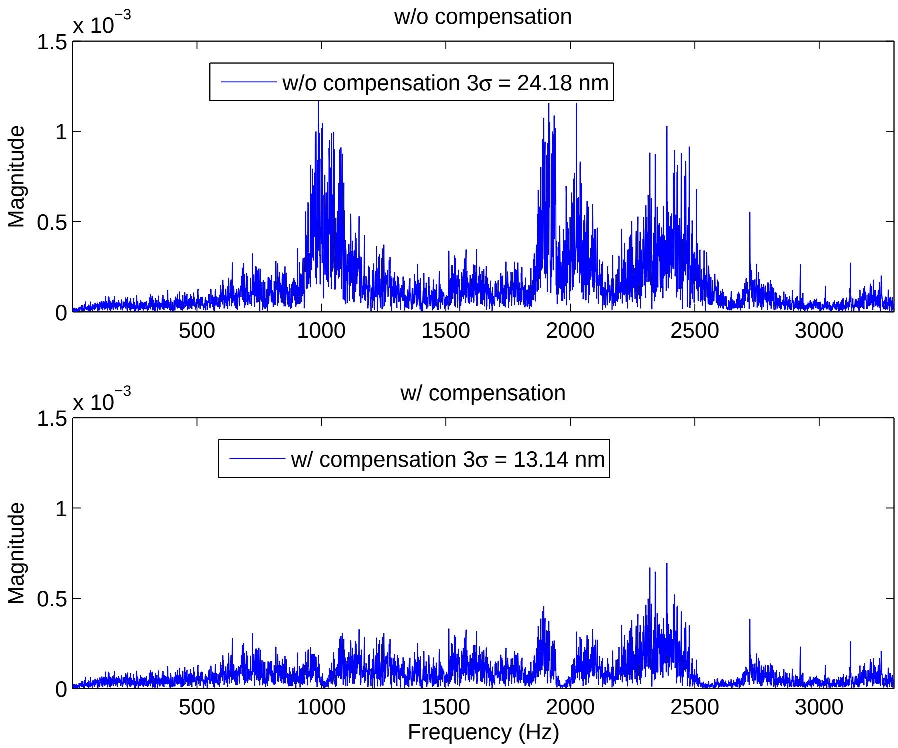


(a) Frequency spectra of PES with and without compensation

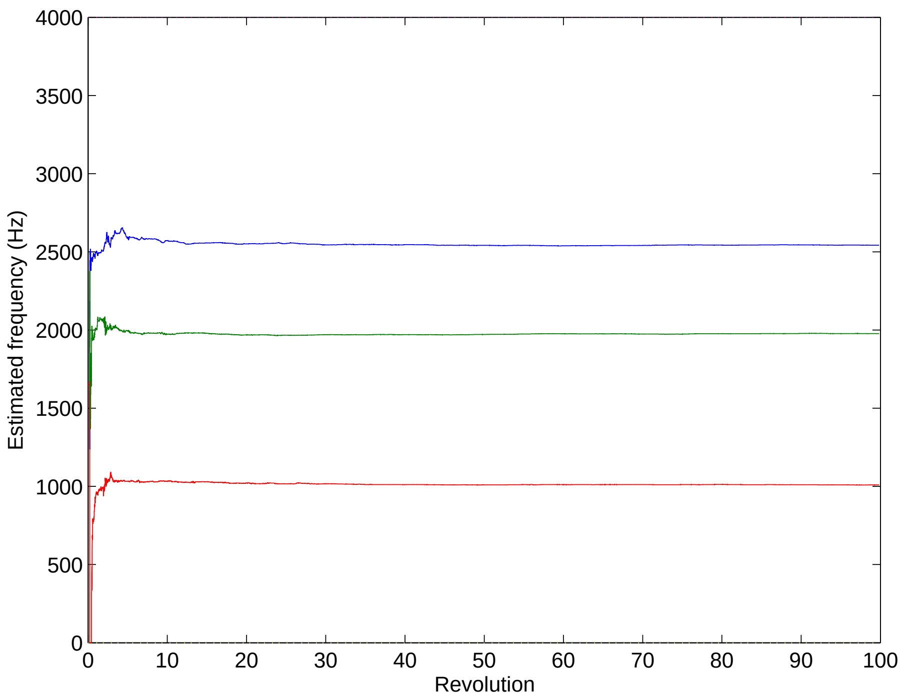


(b) Online identification results

Figure 11.7: Rejecting two narrow-band disturbances at 964 Hz and 1426 Hz



(a) Frequency-domain results



(b) Parameter adaptation for achieving Figure 11.8a

Figure 11.8: Rejecting multi-band audio-vibrations at high frequencies

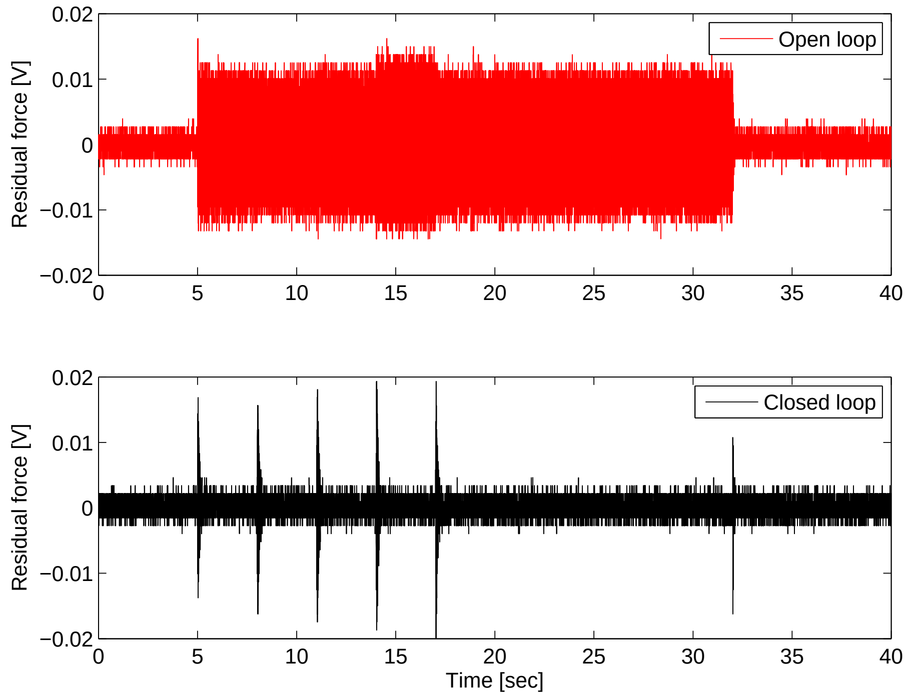
11.3 Application: Time-varing Vibration Rejection on an Active Suspension

Recall the results of narrow-band vibration rejection on the active suspension in Section 5.6. In all the discussed tests there, the vibration frequencies were actually unknown [25], and the controller parameters were updated online via the PAA in Section 10.2. The steady-state performance has been extensively shown in Figures 5.7 to 5.12, we now provide the part of the results about adaptation.

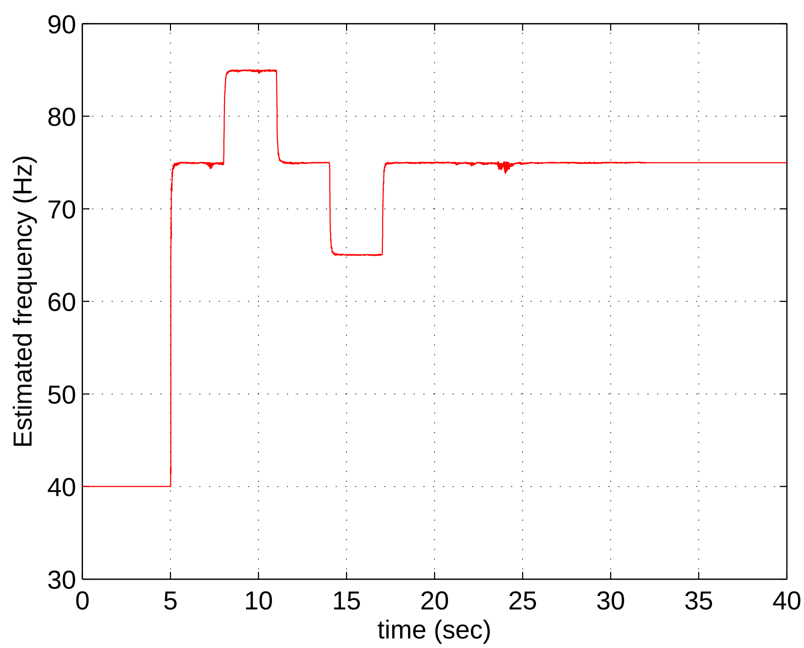
Figure 11.9a shows the time trace of the residual errors (experimental result) for a time-varying vibration with the following characteristics: the disturbance frequency changes in the pattern of null→75Hz→85Hz→75Hz→65Hz→75Hz→null, occurring respectively at 5 sec, 8 sec, 11 sec, 14 sec, 17 sec, and 32 sec. In the presence of various frequency jumps, the algorithm is seen to provide rapid and strong vibration compensation. This is due to the correct online identification of the vibration frequencies in Figure 11.9b. Here, a disturbance detector similar to that in Section 11.1 is used to automatically turn on the identification at 5 sec (the time when disturbance was injected). Comparing the data at 2 sec and 7 sec in Figure 11.9a, we see that the steady-state residual errors with the compensation scheme (data at 7 sec) has been reduced to be at the same magnitude as the baseline case (at 2 sec) where no disturbance presents.

Figure 11.10a shows the experimental results when the disturbance is a chirp vibration whose frequency sweeps between 50 Hz and 95 Hz. Figure 11.10b is the identified frequency values. Under such time-varying vibrations, we see the proposed algorithm maintains its effectiveness of compensating the disturbance. The slight degraded performance at around 14 sec, 20 sec, and 24 sec is due to the transient in the changing frequency, and the intentional limit of the performance above 86 Hz [7, 16].

Similar performances have been achieved for the case where we have multiple time-varying vibrations at different frequencies. Interested readers can refer to the details in [7, 16].

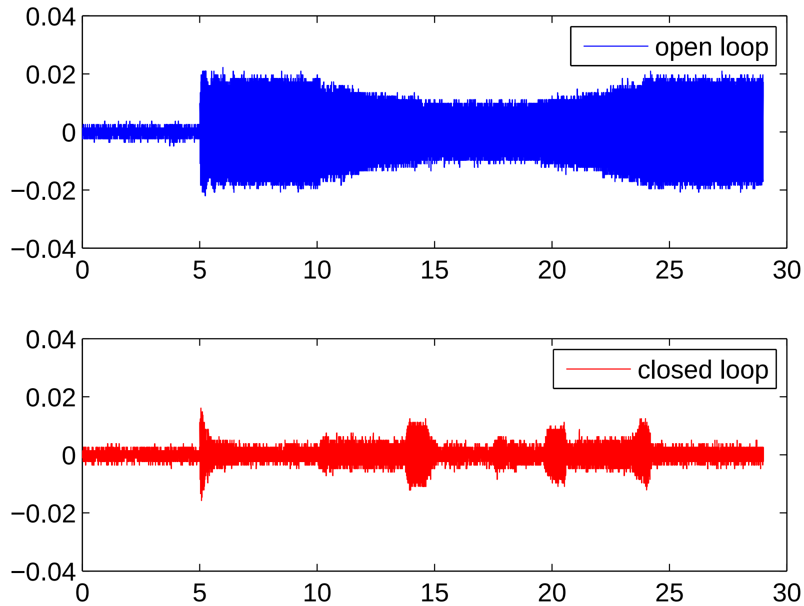


(a) Time-domain results

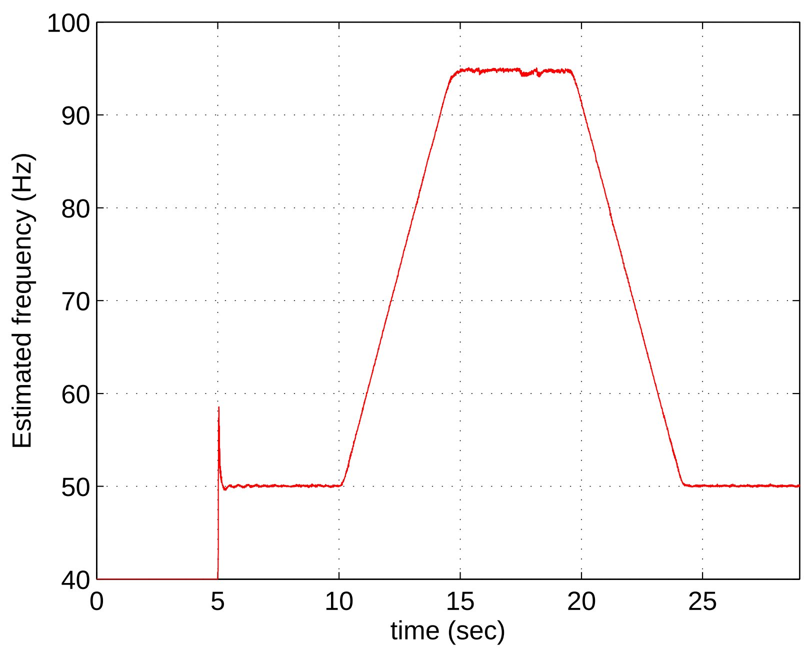


(b) Frequency identification for achieving Figure 11.9a

Figure 11.9: Experimental results of rejecting vibrations with frequency jumps



(a) Time trace



(b) Identified frequency

Figure 11.10: Experimental result of rejecting a chirp disturbance

Chapter 12

Solving the Robust Strictly Positive Real Problem Via Convex Optimization

We have seen the requirement of strictly positive realness in the parameter adaptation algorithms. We briefly described how the condition can be satisfied in Chapter 10. In this chapter, a systematic treatment of the problem is provided.

12.1 Introduction

The strictly positive real (SPR) condition of a transfer function has substantial importance in adaptive control and system identification [56, 55]. An essential problem (see, e.g., [123, 124, 125, 118] and the references therein) in a group of recursive parameter adaptation algorithms is: given an uncertain polynomial $A(\sigma)$ (σ denotes s in continuous-time problems and z in discrete-time problems), design a polynomial compensator $C(\sigma)$ such that the transfer function

$$\frac{C(\sigma)}{A(\sigma)} - \alpha \tag{12.1}$$

is SPR for all possible values of $A(\sigma)$. Here α ($\in [0,1]$) is a fixed scalar that depends on the adaptation algorithm. For instance, in the output error method with a fixed compensator, we have seen that $C(z)/A(z) - 1/2$ being SPR is crucial to assure the stability of the PAA [Section 10.2]. In this case, $A(z)$ comes from the transfer function of the plant to be identified. Conventionally, one has to guess or apply another parameter adaptation algorithm to obtain a $C(z)$ that is hopefully close to $A(z)$ [118]. The same problem occurs in the more general pseudo linear regression algorithm, where the importance of the SPR condition has been remarked in Section 8.6 of [56].

The SPR condition is a strong requirement and is not easy to guarantee for an uncertain $A(\sigma)$. Investigation of the above problem has therefore been popular in the control community. Approaches that use: (a) complex polynomial analysis [126, 127, 128, 129, 130, 123, 125, 131, 132, 133, 134]; (b) geometrical design [132, 133]; and (c) linear matrix inequalities

[135, 136] have been proposed to address the problem. More specific, [123, 131] and [134] characterized the SPR condition and discussed the case when $A(\sigma)$ belongs to a set of stable and known polynomials. [132] and [133] analyzed the situation when the uncertainty in $A(\sigma)$ comes from its root locations or bounded uncertain frequency responses. An important general classification was discussed in [126, 127, 128, 135, 129, 130, 136], where $A(\sigma)$ is assumed to lie in a known polytope, with bounded coefficients in the polynomial. Among the existing results, the majority discussed the case where $\alpha = 0$; [128] and [135] investigated the more difficult situation where $\alpha > 0$. [123, 127, 128, 129, 136, 125, 131, 135, 134, 130] mainly analyzed the continuous-time version of the problem. The discrete-time robust SPR problem has different characteristics compared to the continuous-time version [128]. Within this category, [126] provided conditions for the existence of a solution but did not discuss how to construct it; [133] showed a geometrical design approach for systems with disk uncertainties; later in [135], LMIs are formed to analyze the general SPR condition for an uncertain transfer function $G(\sigma) - \alpha$. The formation and realization of a compensator $C(\sigma)$ was however not discussed, and the equations are slightly more complex than the present approach.

The most natural (and recommended in the related text books [56, 118, 55]) way of designing $C(\sigma)$ is to make it "close" to $A(\sigma)$, such that $C(\sigma)/A(\sigma) - \alpha$ is approximately $1-\alpha$ (usually $\alpha \leq 1$). In fact, this is also substantial for parameter convergence in adaptation algorithms (see Section 4.5.4 of [137]). This aspect has however been largely discredited in previous robust SPR design algorithms.

In this chapter, we present a convex-optimization approach to address the robust SPR problem (with a general non-negative α). The common polytopic uncertainty [126, 127, 128, 135, 129, 130, 136] is adopted here. We will be focusing on the discrete-time version of the problem, partially due to its popularity in system identification and adaptive control, as we have encountered in Chapter 10, and partially because of the fact that results in the more explored continuous-time robust SPR problem do not necessarily generalize to discrete-time systems [128, 135].[1] Additional contributions are as follows. First, we provide a design approach that not only assures the robust SPR condition but also finds the optimal $C(\sigma)$ that is "closest" to $A(\sigma)$. Second, we discuss the achievement of additional optimal properties to the compensator design. This provides us the possibility to investigate several new issues. For instance, in output error based adaptation algorithms, it is favorable for the compensator to have minimum order and/or small gain in the high-frequency region.

12.2 SPR Analysis

Definition 12.1. A proper and rational discrete-time transfer function $G(z)$ is strictly positive real (SPR) if[2]

[1] An extension of the algorithm is discussed in Section 12.7, so that the continuous-time problem can be similarly addressed.

[2] Note that definitions of SPR functions are not uniform in the literature, see [138] for historical remarks.

1. $G(z)$ does not possess any pole outside of or on the unit circle in the complex plane;

2. $\forall|\omega| < \pi$, $G(e^{-j\omega}) + G(e^{j\omega}) = 2Re\{G(e^{j\omega})\} > 0$.

From the above definition, the following properties can be obtained:
If $G(z)$ is SPR, then

1. it is asymptotically stable;

2. the phase response of $G(z)$, after normalization to $[-\pi, \pi]$, lies inside the region $(-\frac{\pi}{2}, \frac{\pi}{2})$ (see, e.g., [126, 134]);

3. the Nyquist plot of $G(z)$ lies in the closed right-half complex plane [139].

Property 1 and 2 are direct results of the first and the second points in Definition 1. The third property is an equivalent statement of Property 2.

The robust SPR problem we will be solving is stated as follows:

Problem 12.1. Given $\alpha \in [0, 1]$ and a monic[3] stable polynomial

$$A(z) = 1 + a_1z^{-1} + a_2z^{-2} + \cdots + a_nz^{-n}, \tag{12.2}$$

with n unknown but bounded coefficients

$$\underline{a}_i \leq a_i \leq \bar{a}_i, \ i = 1, 2, \ldots, n, \tag{12.3}$$

find a polynomial $C(z)$ such that $\frac{C(z)}{A(z)} - \alpha$ is SPR.

Remark 12.1. Notice that the region of coefficients as specified in (12.3) need to be a subset of the stability region of $A(z)$. For example, for $A(z) = 1 + a_1z^{-1} + a_2z^{-2}$, the stability region (obtained by bilinear transformation $z = (1 + s) / (1 - s)$ and Routh test) is a reverse triangle (see Figure 12.1) defined by

$$\begin{aligned} 1 - a_1 + a_2 &> 0 \\ 1 + a_1 + a_2 &> 0 \\ 2 - 2a_2 &> 0. \end{aligned} \tag{12.4}$$

The rectangle

$$\begin{aligned} \underline{a}_1 \leq a_1 \leq \overline{a}_1 \\ \underline{a}_2 \leq a_2 \leq \overline{a}_2 \end{aligned}$$

must stay inside the shape of (12.4).

[3] A monic polynomial is a polynomial whose leading coefficient is equal to 1.

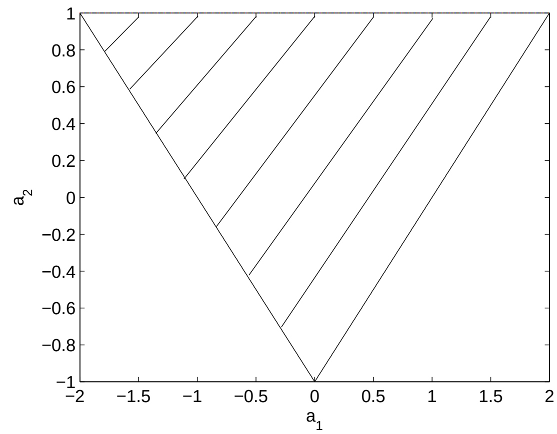


Figure 12.1: Stability region of a second-order polynomial of z^{-1}

In practical applications, α is usually strictly positive [56, 55]. In this case the problem can be normalized as shown in the following proposition.

Proposition 12.1. *For $\alpha > 0$, there exists a polynomial $C(z)$ such that $\frac{C(z)}{A(z)} - \alpha$ is SPR, if and only if $\frac{C'(z)}{A(z)} - \frac{1}{2}$ is SPR for some polynomial $C'(z)$.*

Proof. Under the assumption that $\alpha > 0$,

$$\frac{C(z)}{A(z)} - \alpha = 2\alpha\left(\frac{C'(z)}{A(z)} - \frac{1}{2}\right)$$

where $C'(z) = C(z)/(2\alpha)$. The proof follows by noting the fact that scaling a transfer function by a positive number does not change the SPR property. □

For the above normalized problem, [124] has shown that the SPR condition of

$$Re\left\{C'(e^{j\omega})/A(e^{j\omega}) - 1/2\right\} > 0$$

is equivalent to

$$|A(e^{j\omega})/C'(e^{j\omega}) - 1| < 1,$$

from which it is clear that letting $C'(z)/A(z) \approx 1$ is a feasible solution. This is the suggested way of designing the compensator $C'(z)$ in text books of system identification and adaptive control [137, 118, 56, 55], and is also important for the parameter convergence, as discussed in Section 12.1.

12.3 Polytopic Uncertainty

In this section, we briefly review the characterization of the polytopic constraint (12.3) and provide a general result of the robust SPR problem.

Notice that (12.2) can be equivalently represented as

$$A(z) = 1 + \left[z^{-1}, z^{-2}, \ldots, z^{-n}\right] \left[a_1, a_2, \ldots, a_n\right]^T. \tag{12.5}$$

Consider an n dimensional vector space that contains the coefficient vector $\left[a_1, a_2, \ldots, a_n\right]^T$. An alternative representation of (12.3) is to use the concept of convex hull (see, e.g., [60]), which states that $\left[a_1, a_2, \ldots, a_n\right]^T$ can be characterized by the extreme edge vectors that are defined by lower and upper bounds of a_i's:

$$\left[a_1, a_2, \ldots, a_n\right]^T = \sum_{j=1}^{2^n} \theta_j \left[b_{j,1}, b_{j,2}, \ldots, b_{j,n}\right]^T, \ \theta_j \geq 0, \ \sum_{j=1}^{2^n} \theta_j = 1,$$

where $b_{j,i} = \underline{a}_i$ or $\bar{a}_i$. There are 2^n edge vectors. This number can be reduced if some parameters are known a prior. Applying the above result to (12.5) yields

$$A(z) = \sum_{j=1}^{2^n} \theta_j A_j(z), \ \theta_j \geq 0, \ \sum_{j=1}^{2^n} \theta_j = 1, \tag{12.6}$$

where $A_j(z)$'s are the edge polynomials defined by $b_{j,i}$'s. Note that since $A(z)$ is stable by assumption,[4] all the $A_j(z)$'s therefore are also stable.

As a second-order example, when $A(z) = 1 + a_1 z^{-1} + a_2 z^{-2}$, with $\underline{a}_1 \leq a_1 \leq \bar{a}_1$ and $\underline{a}_2 \leq a_2 \leq \bar{a}_2$ as shown in Figure 12.2, the convex-hull formulation then says that any point inside the rectangle can be expressed as a convex combination of the four edge points: $[\underline{a}_1, \underline{a}_2]^T$, $[\bar{a}_1, \underline{a}_2]^T$, $[\underline{a}_1, \bar{a}_2]^T$, and $[\bar{a}_1, \bar{a}_2]^T$.

Equation (12.6) provides a convenient interpretation of the uncertainty in $A(z)$. Instead of a polynomial of unknown coefficients (in an entire vector space), we now have a convex combination of *a finite number of* fixed polynomials. Moreover, we have the following result:

Lemma 12.1. *If $A(z)$ is given by (12.6), then $\frac{C(z)}{A(z)} - \alpha$ is SPR if and only if $\frac{C(z)}{A_i(z)} - \alpha$ is SPR $\forall i = 1, 2, \ldots, 2^n$.*

Proof. This is an enhanced version of Lemma 3.1 in [128]. The "only if" part of the proof is trivial since $C(z)/A_i(z)$ corresponds to $\theta_j = 1$ for $j = i$ and $\theta_j = 0$ for $j \neq i$ in (12.6). The "if" part of the proof is similar to that in [128] and omitted here. □

[4] To obtain the robust stability condition of such 'interval polynomials' with bounded coefficients, the Kharitonov's Theorem [140, 141] provides a complete solution to the continuous-time domain problem and has motivated various research towards its discrete-time equivalences (see, e.g., [142, 143] and the references contained therein.)

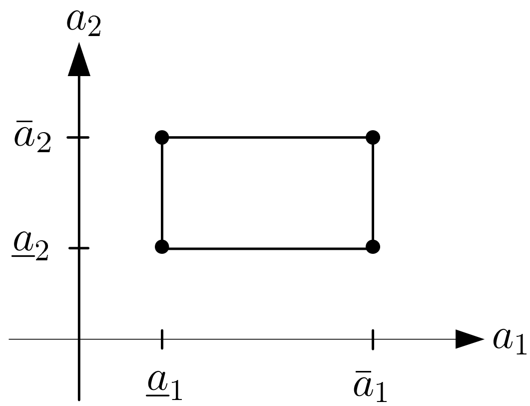


Figure 12.2: The 2D convex hull containing $[a_1, a_2]^T$

12.4 Achieving the Robust SPR Condition

Given the SPR problem, we have first shown that the uncertain $A(z)$ has the equivalent representation in (12.6). Lemma 12.1 then leads to investigation of the SPR condition for each edge transfer function $C(z)/A_j(z) - \alpha$. From Definition 12.1, the SPR condition itself is specified at an infinite amount of frequencies. We now apply the positive-real lemma to translate the infinite dimensional problem to a single LMI:

Lemma 12.2. *(positive-real lemma) A square discrete-time system*

$$G_p(z) = C_p\,(zI - A_p)^{-1}\,B_p + D_p$$

is SPR if and only if there exists a positive definite matrix $P = P^T \succ 0$ such that the following matrix inequality holds

$$\begin{bmatrix} P - A_p^T P A_p & C_p^T - A_p^T P B_p \\ C_p - B_p^T P A_p & D_p^T + D_p - B_p^T P B_p \end{bmatrix} \succ 0. \tag{12.7}$$

Proof. see [144, 145, 146]. □

Let $G(z) = C(z)/K(z) - \alpha$, where $K(z)$ represents an edge polynomial $A_j(z)$. Define

$$\begin{aligned} C(z) &= c_0 + c_1 z^{-1} + \cdots + c_l z^{-l} \\ K(z) &= 1 + k_1 z^{-1} + \cdots + k_n z^{-n}, \end{aligned} \tag{12.8}$$

The order of $C(z)$ is a design parameter here. Depending on the values of l and n, different situations exist for the design of (12.7):

Case 1: if $l \geq n$, it is straightforward to show that

$$\frac{C(z)}{K(z)} - \alpha = (c_0 - \alpha) + \frac{(c_1 - c_0 k_1)\, z^{l-1} + \cdots + (c_n - c_0 k_n)\, z^{l-n} + c_{n+1} z^{l-n-1} + \cdots + c_l}{z^l + k_1 z^{l-1} + \cdots + k_n z^{l-n}}$$

which has the following state-space realization:

$$\begin{aligned}
A_p &= \begin{bmatrix} 0_{l-1,1} & I_{l-1,l-1} \\ 0 & * \end{bmatrix}_{l\times l}, \ B_p = \begin{bmatrix} 0_{l-1,1} \\ 1 \end{bmatrix} \\
* &= [0_{1,l-n-1}, -k_n, \ldots, -k_1] \\
C_p &= \begin{bmatrix} c_l, & \ldots & , c_{n+1}, & c_n, & \ldots & , & c_1 \end{bmatrix} - c_0 \begin{bmatrix} 0_{1,l-n}, & k_n, & \ldots, & k_1 \end{bmatrix} \\
D_p &= c_0 - \alpha.
\end{aligned} \tag{12.9}$$

Here the controllable canonical form is proposed, so that when we form (12.7), the matrix on the left-hand side is linear in the decision variables $[c_0, c_1, \ldots, c_l]$.

Case 2: if $n > l$, similar analysis gives

$$\begin{aligned}
A_p &= \begin{bmatrix} 0 & 1 & 0 & \ddots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & 1 \\ -k_n & \ldots & -k_2 & -k_1 \end{bmatrix}_{n\times n}, \ B_p = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}_{n\times 1} \\
C_p &= -c_0 \begin{bmatrix} k_n & \ldots & k_{l+1} & , k_l & \ldots & , & k_1 \end{bmatrix} + \begin{bmatrix} 0_{1,n-l}, & c_l, & \ldots, & c_1 \end{bmatrix} \\
D_p &= c_0 - \alpha.
\end{aligned} \tag{12.10}$$

Equation (12.9) or (12.10) can now be applied to construct (12.7). Such constructions are repeated for each edge transfer function. We can now formulate the following feasibility problem:

$$\begin{aligned}
&\text{find } c_0, \ldots c_l \in \mathbb{R} \text{ and } P_j = P_j^T \succ 0 \\
&\text{s.t. } \begin{bmatrix} P_j - A_{p,j}^T P_j A_{p,j} & C_{p,j}^T - A_{p,j}^T P_j B_{p,j} \\ C_{p,j} - B_{p,j}^T P_j A_{p,j} & D_{p,j}^T + D_{p,j} - B_{p,j}^T P_j B_{p,j} \end{bmatrix} \succ 0, \ j = 1, 2, \ldots 2^n,
\end{aligned} \tag{12.11}$$

where for each j, $(A_{p,j}, B_{p,j}, C_{p,j}, D_{p,j})$ is defined by (12.9) or (12.10), with $K(z) = A_j(z)$. By construction, $C_{p,j}$ and $D_{p,j}$ depend affinely on c_i's. Problem (12.11) is thus a convex, actually semidefinite programming (SDP) problem, and can be solved by efficient interior-point methods (using, e.g., [84]) in convex optimization.[5]

12.5 Optimal Properties

Section 12.4 provides a tool to obtain a feasible solution. Another main design aspect is to obtain coefficients $c_0, \ldots, c_l$ in (12.11) with designer-assigned optimal properties, one of which is to keep $C(z)/A(z)$ close to 1. In this section, together with the "close-to-1" condition,

[5] When the problem is formulated for computer solvers, the positive definite constraint $P = P^T \succ 0$ is transformed to $P - \epsilon I \succeq 0$, where ϵ is a small positive number chosen as the lower bound of all the eigenvalues of P.

we provide a few examples to obtain the optimal compensator $C(z)$. The discussions are separated into subsections, but they can be combined by a weighted sum or the minimum of the weighted costs, to satisfy multiple design objectives. All the results in this section are subject to the baseline SPR requirement in Section 12.4.

$C(z)/A(z)$ Being Close to 1 The intuition and the importance of this objective has been discussed in Section 12.2. For the z-domain transfer function $C(z)/A(z)$ to be close to 1, we aim at minimizing the maximum value of $|C(e^{j\omega})/A(e^{j\omega}) - 1|$ over the entire frequency region, i.e.,

$$\min_{c_0,\dots c_l \in \mathbb{R}} \left\| \frac{C(z)}{A(z)} - 1 \right\|_\infty . \tag{12.12}$$

In a more general and flexible form, we consider

$$\min_{c_0,\dots c_l \in \mathbb{R}} \|R(z)\|_\infty \triangleq \left\| \frac{W(z)}{V(z)} C(z) - \eta \right\|_\infty \tag{12.13}$$

with $W(z) = w_0 + w_1 z^{-1} + \cdots + w_l z^{-l}$, $V(z) = 1 + v_1 z^{-1} + \cdots + v_l z^{-l}$, and $C(z) = c_0 + c_1 z^{-1} + \cdots + c_l z^{-l}$. For simplicity, it is assumed that $W(z)$, $V(z)$ and $C(z)$ have the same order. If not, one can classify different situations in a way similar to that in Section 12.4, or simply constrain the coefficients of the excessive high-order terms to be zero. Using (12.13), we can essentially constrain $C(z)$ to have an arbitrary desired (if feasible) frequency response.

Equation (12.13) can be transformed to a tractable optimization problem in a form similar to (12.11), by utilizing the bounded-real lemma.[6] To do that, we need the following system construction: notice that $H(z) \triangleq W(z)C(z) = h_0 + h_1 z^{-1} + \cdots + h_{2l} z^{-2l}$ is given by the convolution

$$\begin{bmatrix} h_0 \\ h_1 \\ \vdots \\ \vdots \\ h_{2l} \end{bmatrix} = \underbrace{\begin{bmatrix} w_0 & 0 & \cdots & \cdots & 0 \\ w_1 & w_0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ w_l & w_{l-1} & \cdots & w_1 & w_0 \\ 0 & w_l & \ddots & \ddots & w_1 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & w_{l-1} \\ 0 & \cdots & \cdots & 0 & w_l \end{bmatrix}}_{(2l+1)\times(l+1)} \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ \vdots \\ c_l \end{bmatrix}$$

[6] Recall also its application in Section 6.4.

from which $R(z) = H(z)/V(z) - \eta$ has the following state-space realization:

$$A_r = \begin{bmatrix} 0_{2l-1,1} & I_{2l-1,2l-1} \\ 0 & * \end{bmatrix}_{2l\times 2l},\ B_r = \begin{bmatrix} 0_{2l-1,1} \\ 1 \end{bmatrix} \tag{12.14}$$

$$* = [0_{1,l-1}, -v_l, -v_{l-1}, \ldots, -v_1],\ D_r = h_0 - \eta$$

$$C_r = [h_{2l},\ h_{2l-1}, \ldots, h_1] - h_0\,[0_{1,l}, v_l, \ldots, v_1].$$

The bounded-real lemma then says that (12.13) can be achieved if and only the following problem can be solved:

$$\min_{\gamma,P_r,c_i} : \gamma \geq 0 \tag{12.15}$$

$$s.t.:\ -\begin{bmatrix} A_r^T P_r A_r - P & A_r^T P_r B_r & C_r^T \\ B_r^T P_r A_r & B_r^T P_r B_r - \gamma I & D_r^T \\ C_r & D_r & -\gamma I \end{bmatrix} \succeq 0 \tag{12.16}$$

$$P_r = P_r^T \succ 0$$

Again, we used the controllable canonical form in (12.14), so that the left hand side of (12.16) is affine in γ, P_r and c_i. After adding the SPR constraint (12.11), the minimization in (12.15) remains a convex optimization problem.

A candidate polynomial $A(z)$ is needed in (12.12). For the specific polytopic uncertainty (12.3), the center of the polytope can be used.

Minimum-Order Compensator The order of the compensator is directly related to the required computation complexity in the related system identification or adaptive control problems. The common practice in system identification is to apply $l = n$ in (12.8). With the proposed algorithms, it is however possible to find the compensator $C(z)$ with the minimum number of coefficients. This can be rapidly achieved through the optimization formulation, by starting the feasibility problem (12.11) with $l = n$, and iteratively reducing l until (12.11) becomes infeasible.

A related design is to obtain sparse (having large amounts of zeros) coefficients in $C(z)$. In that case, if a feasible order l is firstly assigned, we can apply the 1-norm approximation for cardinality minimization (see, e.g., [60]), and add the cost function $\min_{c_i,P_i} \| [c_0, c_1, \ldots, c_l]^T \|_1$ to (12.11), which shall provide a sparse $[c_0, c_1, \ldots, c_l]$.

Minimum High-Frequency Gains In the output error method with a fixed compensator [Section 10.2], the output error is filtered through $C(z)$ to obtain the adaptation error, denoted as $v(k)$, for parameter identification or adaptive control. Excessive high-frequency components in $v(k)$ reduces the signal-to-noise ratio and increases the quantization error. It is therefore favorable to limit the high-frequency magnitude of $C(z)$ (and hence the high-frequency energy in $v(k)$).

Recall that the compensator is given by $C(z) = c_0 + c_1 z^{-1} + \cdots + c_l z^{-l}$, whose frequency response $C(e^{j\omega})$ at $\omega = \pi$ is $C(e^{j\pi}) = C(z)|_{z=-1}$. To minimize the high-frequency gain (at the Nyquist frequency) of $C(z)$, we can add the following objective:

$$\min \left| C(e^{j\pi}) \right| = \left| [1, -1, 1, -1, \dots] \, [c_0, c_1, \dots, c_l]^T \right| \tag{12.17}$$

which is linear in the decision variables c_i's.

12.6 Design Examples

Consider identification of the following second-order system:

$$\frac{B(z)}{A(z)} = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2}}{1 - 2\beta a_1 z^{-1} + \beta^2 z^{-2}},$$

where the damping ratio $\beta = 0.98$ and the sampling period $T_s = 1/26400$ sec. Such a transfer function generalizes various rigid-body models of mechanical systems, and can also represent the model of vibrations with a single-frequency component. Assume that we know the resonance of the system is between 700 Hz and 1000 Hz, i.e., the unknown numerator coefficient $a_1 \in [\cos(2\pi T_s \times 700), \cos(2\pi T_s \times 1000)]$. When using the output error identification method with a fixed compensator and a forgetting factor of 1, one will have $\alpha = 1/2$ in $C(z)/A(z) - \alpha$. Figure 12.3 demonstrates the frequency response of 7 possible $1/A(z)$'s uniformly sampled from the uncertainty region. One can observe that in a large frequency region, the phase responses of $1/A(z)$ is below -90 degree, i.e., $Re\left(1/A(e^{j\omega})\right) \le 0$. Therefore, $1/A(z)$ is not SPR, not to say $1/A(z) - 1/2$.

It is easy to check that $\forall a_1 \in [\cos(2\pi T_s \times 700), \cos(2\pi T_s \times 1000)]$, $A(z)$ is stable. The first condition for SPR transfer functions is thus satisfied. The two edge polynomials in this case are $A_1(z) = 1 - 2\beta \underline{a}_1 z^{-1} + \beta^2 z^{-2}$ and $A_2(z) = 1 - 2\beta \bar{a}_1 z^{-1} + \beta^2 z^{-2}$, with $\underline{a}_1 = \cos(2\pi T_s \times 700)$ and $\bar{a}_1 = \cos(2\pi T_s \times 1000)$. Formulating and solving (via [84]) the feasibility design in Section 12.4, with $l = n = 2$, we obtain

$$C(z) = 12.36 - 10.71 z^{-1} - 1.162 z^{-2}. \tag{12.18}$$

Plotting the frequency responses of the sampled $1/A(z)$ and $C(z)/A(z) - 1/2$ in Figure 12.4, one observes that $\forall \omega$, $C(z)$ is capable of providing robust compensation such that $C(e^{j\omega})/A(e^{j\omega}) - 1/2$ stays strictly in the open right-half complex plane (phase$\in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$), which, combined with the condition that $A(z)$ is always stable, indicates the success of the robust SPR design.

Notice however in Figure 12.4, that large gain variations exist in $C(z)/A(z) - 1/2$ and that $C(z)/A(z)$ is far away from the unity function. Indeed, the roots of $C(z)$ are $\{0.9637, -0.0976\}$ while the roots of $A(z)$ always appear in complex pairs and range from $\beta e^{\pm j 2\pi T_s \times 700}$ $(0.9664 \pm 0.1625i)$ to $\beta e^{\pm j 2\pi T_s \times 1000}$ $(0.9524 \pm 0.2310i)$. It is thus seen that the

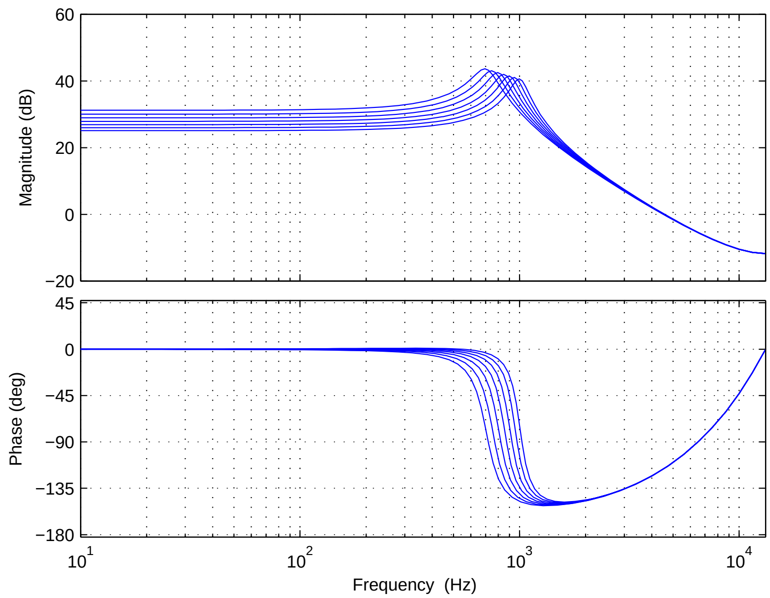


Figure 12.3: Frequency responses: samples of the uncertain $1/A(z)$

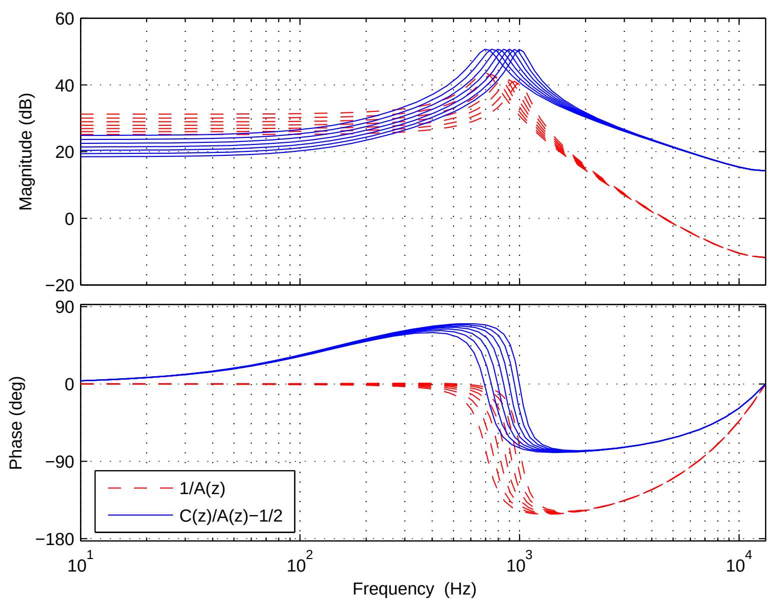


Figure 12.4: Frequency responses of the sampled $1/A(z)$ and $C(z)/A(z) - 1/2$

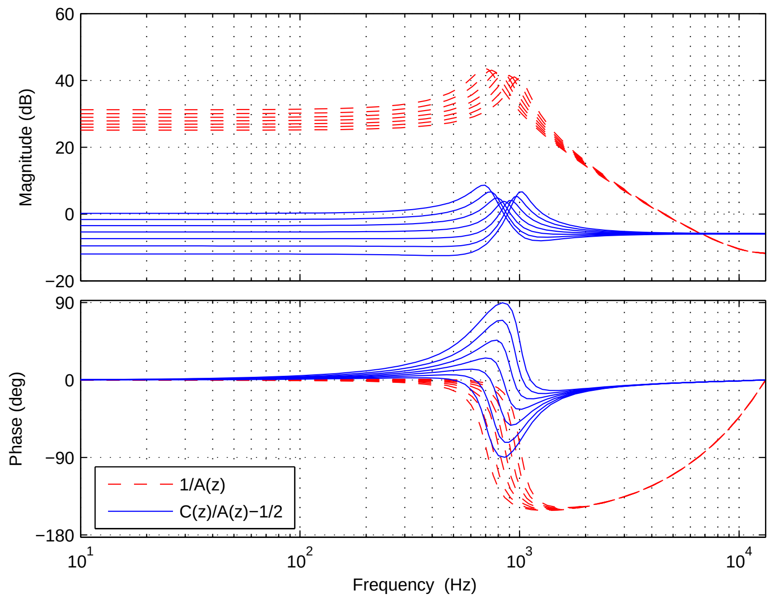


Figure 12.5: Sampled frequency responses of $1/A(z)$ and $C(z)/A(z)-1/2$: with the objective of $\min ||C(z)/A(z)-1||_\infty$

obtained $C(z)$ makes $C(z)/A(z) - 1/2$ robustly SPR, but does not reflect intuitive information about the system $A(z)$ and additionally may negatively influence convergence of the adaptation process. The algorithm in Section 12.5 is then applied to improve the result. The objective of $\min ||C(z)/A^*(z)-1||_\infty$ is enforced, with $A^*(z)$ being the center of the polytope. The resulted solution is $C(z) = 1.027 - 1.932z^{-1} + 0.9464z^{-2}$. Figure 12.5 shows the sampled frequency responses of the newly obtained $C(z)/A(z) - 1/2$. Besides the achievement of the robust SPR requirement, $C(e^{j\omega})/A(e^{j\omega}) - 1/2$ is significantly confined to be in a smaller region: compared to Figure 12.4, the phase of $C(e^{j\omega})/A(e^{j\omega}) - 1/2$ is close to 0 degree at the majority of frequencies; the magnitude response is condensed to be within 0.2503 (-12.03 dB) and 2.6931 (8.605 dB); specifically at high frequencies, $C(e^{j\omega})/A(e^{j\omega}) - 1/2$ is approximately 0.5107 (-5.836 dB). Therefore, $C(z)/A(z)$ is much closer to 1 from the optimal design. The roots of $C(z)$ in this case are $\{0.9402 \pm 0.1928i\}$, which reflect much more information about $A(z)$ compared to (12.18).

Finally we explore the optimal $C(z)$ that has the minimum high-frequency magnitude. Applying the algorithms in Section 12.5, we obtain the new optimal solution

$$C^*(z) = 6.7331 - 0.7938z^{-1} - 5.5802z^{-2}. \tag{12.19}$$

Comparing the frequency responses of (12.18) and (12.19) in Figure 12.6, and noticing the deep notch in the solid line of the magnitude responses, we can see that (12.19) indeed provides strong high-frequency gain attenuation due to the cost function design in (12.17).

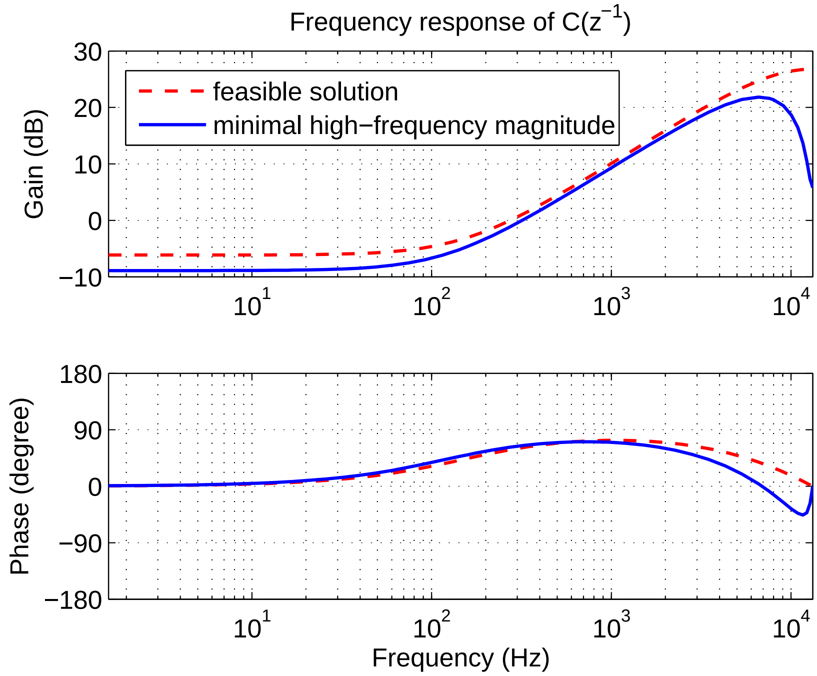


Figure 12.6: Frequency responses of the optimal $C^{*}(z)$ and the $C(z)$ from the feasibility design

12.7 Further Discussions and Extensions

In this chapter, a convex optimization approach is proposed to address the design of robust strictly positive real transfer functions. It is shown that a feasibility SDP formulation can be used to provide the compensator that achieves the desired robust SPR condition. Moreover, the important issue of maintaining the designed transfer function to be close to 1 is addressed, by adding an infinity-norm minimization in the optimization. Additional concepts of cost function design are introduced, which lead to solutions of several new problems. All the formulated optimization problems can be efficiently solved by interior-point methods in convex optimization.

Notice that although the focus has been placed on the discrete-time SPR analysis. By applying the continuous-time positive-real lemma and bounded-real lemma, the present work can be easily extended to solve the continuous-time version of the problem. There is yet one additional condition that makes the latter problem easier to solve: in the continuous-time case, for the SPR condition to hold, the relative degree of $G_p(s)$ equals zero or one [126, 139], hence greatly simplifying the formulation of matrix inequalities.

Chapter 13

Conclusions and Future Works

13.1 Concluding Remarks

Motivated by the needs from various practical problems, we have discussed the idea of local loop shaping (LLS) for linear feedback control. At a time when general (low-frequency) loop shaping is mature and easily achievable, we focused on add-on enhancements to strengthen the local (in the frequency domain) performance of the system, with the following design goals:

1. **performance**: to achieve strong error attenuation at single or multiple bands of frequencies, with minimum amplification of the sensitivity function;
2. **flexibility**: to make the algorithm easily tunable in practice, and expandable for different control tasks (e.g., repetitive control, vibration rejection, bandwidth extension, and so on);
3. **adaptability**: to build robustness to the closed loop for online *self* identification of the disturbance characteristics.

We also studied generalizations of LLS, and learned that LLS is equally achievable for MISO systems.

Besides the LLS idea, we mention the following experiences we obtained along the way:

Precision Control and Vibration Rejection

In linear systems, these problems are essentially all about loop shaping. The internal model principle provides a systematic way to customize the closed loop for enhanced loop-shaping design.

Inverse-based Control and Youla-Kucera Parameterization

By using inverse models, we gain quite a few benefits in the design and tuning of YK parameterization. After establishing the connection between disturbance observer and YK parameterization, we can see that YK schemes have actually been more broadly used (under a different name) then it is known. In motion control, model inversion is quite easy to achieve. For general systems, if a full inversion is not available, we can use optimization tools to obtain a selective model inversion.

Reduced-complexity Adaptive YK Parameterization

FIR filters are great tools for adaptive control, as they maintain their stability during change of their coefficients. This is a main reason for their popularity in literatures of adaptive YK parameterization. IIR filters, as long as its stability is carefully preserved, has the potential of more flexible and reduced-complexity implementation in adaptive control. This is the direction we pursued in the design and adaptation of pseudo YK parameterization.

Control of Waterbed Effect

In LTI systems, Bode's Integral Formula always holds (practically). However, the waterbed effect can definitely be controlled, as we can see from the YK parameterization schemes, where we translate the feedback design problem, to an affine (and hence much easier) one. In particular, in pseudo YK parameterization, the control of the waterbed effect further simplifies to minimizing the infinity norm of $1 - z^{-m}Q(z)$.

Optimization in Controls

The theory of optimization is more and more integrated in control engineering. In our study of digital-filter design, robust SPR design, and infinity-norm minimization, optimization techniques provided great benefits, e.g., to, translate infinite dimensional problems to finite ones, and to directly tell whether a problem is achievable or not. In particular, LMI and semidefinite programming are closely related to many control problems. Mature theories and tools are nowadays available for building such a linkage.

13.2 Topics of Future Research

Transient Control

Except for the enhanced repetitive control [Chapter 8] and adaptive YK parameterizations [Part II], we have been mostly concerned with the steady-state performance of linear systems. This is the case for general loop-shaping designs as well. Our experience in enhanced repetitive control clearly indicates that transient, at least in linear systems, is controllable (via time-varying and nonlinear control). There are definitely more to explore. For instance, if there are large mismatches between the initial states of the plant and its coprime-factorization model, a transient error will be generated. Although eventually it will die out (given a stable closed-loop design), the performance degradation may be too large to ignore in practical systems. Besides the plant model states, the initial states of the Q filter do not need to be zero as well. Some solution concepts can be investigated from the initial value compensation problem [147].

Youla-Kucera Parameterization for Nonlinear Systems

For the special YK parameterization scheme in Figure 4.2 on page 48, if we replace the transfer function P with a general nonlinear function $y = f(u)$, a result similar to the affine Q parameterization can be made (although the design criteria for Q need to be reconstructed). However, it gets fundamentally more challenging to construct general and pseudo YK parameterizations in nonlinear systems. Actually, general nonlinear systems do not have transfer functions, and new tools need to be developed for the analysis of YK parameterization here. Also, for the pseudo YK scheme, the inverse of nonlinear systems becomes nontrivial, and may only be achievable for a sub class of systems.

Nonlinear YK Parameterization for Linear Systems

Related to the first topic of future research, we remark that even for linear systems, the controllers can be nonlinear. In certain cases, this nonlinearity is introduced for performance purpose (e.g., the mentioned transient control problem). In other situations, nonlinear control might be necessary for safety reasons. For instance, any practical system is subjected to the problem of actuator saturation, which is an intrinsic nonlinearity. High-gain feedback control is essentially about creating stronger control efforts to the plant. Thus there is a principle conflict between high performance and actuator safety. In the add-on YK parameterization schemes, the design again narrows down to the Q-filter implementation. A meaningful extension would be to construct certain anti-windup (see. e.g., [148, 149, 150]) mechanism for recovering some performance when saturation indeed occurs.

Adaptive Repetitive Control

When the period of the reference or the disturbance is unknown or uncertain, robustness and/or adaptiveness need to be built in repetitive control. The first option has been investigated in literatures such as [151, 109, 152, 153]. Less results, however, are available for a true adaptive repetitive control algorithm that online identifies the order of the internal model $1 - z^{-N}$. For an uncertain N, robust repetitive control is a viable option, especially if the fundamental frequency of the harmonic components is not too large. If on the other hand very few knowledge about N is known, or if the period is changing, then it would be more beneficial to construct rigorous parameter adaptation algorithms for estimating N. More general, this belongs to the problem of *online* identification about the system orders, which has been much less explored in system identification theory.

Bandwidth Adaptation and Constrained Adaptation

We adapted only the center frequencies when using LLS to attenuate spectral peaks. It would be greatly beneficial also to online adaptively determine the bandwidth of the attenuation regions. However, if we are to focus only on error reduction, then standard adaptation algorithms will tend to increase the bandwidth as much as possible, making the closed-loop vulnerable to guarantee robustness and even stability. This is the trade off between performance and robustness in adaptive control. Either some constrained adaptation can be introduced to keep the system in a region of stability, or some dead zone need to be created in the PAA, to stop the adaptation when the desired performance has been achieved. The determination of the stability region and dead-zone threshold may however be not easy for different systems and disturbances.

Bibliography

[1] A. Al Mamun, G. Guo, and C. Bi, *Hard disk drive: mechatronics and control.* CRC Press, 2007.

[2] I. S. T. Group, "Technology roadmap comparisons for tape, hdd, and nand flash," http://www.digitalpreservation.gov/meetings/documents/storage12/5-Fontana-StorageMediaDenstiyfoRNANDTAPE.pdf, 2012.

[3] X. Chen and M. Tomizuka, "A minimum parameter adaptive approach for rejecting multiple narrow-band disturbances with application to hard disk drives," *IEEE Transactions on Control System Technology*, vol. 20, no. 2, pp. 408 –415, march 2012.

[4] ——, "New repetitive control with improved steady-state performance and accelerated transient," *IEEE Transactions on Control System Technology*, vol. PP, no. 99, pp. 1–1, 2013.

[5] ——, "An enhanced repetitive control algorithm using the structure of disturbance observer," in *Proceedings of 2012 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)*, july 2012, pp. 490 –495.

[6] ——, "An indirect adaptive approach to reject multiple narrow-band disturbances in hard disk drives," in *Proc. 2010 IFAC Symposium on Mechatronic Systems, Sept. 13-15, 2010 Cambridge, Massachusetts, USA*, 2010, pp. 44–49.

[7] ——, "Selective model inversion and adaptive disturbance observer for time-varying vibration rejection on an active-suspension benchmark," *European Journal of Control*, vol. 19, no. 4, pp. 300 – 312, 2013. [Online]. Available: http://dx.doi.org/10.1016/j.ejcon.2013.04.002

[8] X. Chen, A. Oshima, and M. Tomizuka, "Inverse based local loop shaping for vibration rejection in precision motion control," in *The 6th IFAC Symposium on Mechatronic Systems, Hangzhou, China, April 10-12*, 2013, pp. 490–497.

[9] X. Chen and M. Tomizuka, "Unknown multiple narrow-band disturbance rejection in hard disk drives–an adaptive notch filter and perfect disturbance observer approach," in *Proceedings of 2010 ASME Dynamic Systems and Control Conference*, 2010, pp. 936–970.

[10] ——, "Decoupled disturbance observers for dual-input-single-output systems with application to vibration rejection in dual-stage hard disk drives," in *Proceedings of 2012 ASME Dynamic Systems and Control Conference, and 2012 Motion & Vibration Conference*, 2012, pp. 1544–1554.

[11] ——, "Spiral servo writing in hard disk drives using iterative learning based tracking control," in *the 18th World Congress of the International Federation of Automatic Control (IFAC), Aug. 28-Sept. 2, 2011, Milano, Italy*, 2011, pp. 5279–5285.

[12] A. Oshima, X. Chen, and M. Tomizuka, "Control design for cancellation of unnatural reaction torque and vibrations in variable-gear-ratio steering system," in *2013 ASME Dynamic Systems and Control Conference, Stanford University, Palo Alto, CA, Oct. 21-23*, 2013.

[13] X. Chen and M. Tomizuka, "Add-on loop shaping via youla parameterization for precision motion control," in *American Society for Precision Engineering 2013 Spring Topical Meeting and the MIT Laboratory for Manufacturing and Productivity Annual Summit, Cambridge, MA, Apr. 21-23*, 2013, pp. 80–83.

[14] ——, "Control methodologies for precision positioning systems," in *American Control Conference (tutorial paper)*, 2013, pp. 3710–3717.

[15] ——, "Reduced-complexity and robust youla parameterization for discrete-time dual-input-single-output systems," in *IEEE/ASME International Conference on Advanced Intelligent Mechatronics, Wollongong, Australla, July 9-12*, 2013, pp. 490–497.

[16] ——, "Selective model inversion and adaptive disturbance observer for rejection of time-varying vibrations on an active suspension," in *2013 European Control Conf., July 17-19 (to appear)*.

[17] ——, "Optimal plant shaping for high bandwidth disturbance rejection in discrete disturbance observers," in *American Control Conference (ACC)*, 2010, pp. 2641–2646.

[18] L. Sievers and A. von Flotow, "Comparison and extensions of control methods for narrow-band disturbance rejection," *IEEE Transactions on Signal Processing*, vol. 40, no. 10, pp. 2377–2391, 1992.

[19] I. D. Landau, A. Constantinescu, and M. Alma, "Adaptive regulation-Rejection of unknown multiple narrow band disturbances," in *17th Mediterranean Conference on Control & Automation*. IEEE, 2009, pp. 1056–1065.

[20] R. de Callafon and C. E. Kinney, "Robust estimation and adaptive controller tuning for variance minimization in servo systems," *Journal of Advanced Mechanical Design, Systems, and Manufacturing*, vol. 4, no. 1, pp. 130–142, 2010.

[21] F. Ben-Amara, P. T. Kabamba, and a. G. Ulsoy, "Adaptive sinusoidal disturbance rejection in linear discrete-time systems–part I: Theory," *ASME Journal of Dynamic Systems, Measurement, and Control*, vol. 121, no. 4, pp. 648–654, 1999.

[22] ——, "Adaptive sinusoidal disturbance rejection in linear discrete-time systems–part II: Experiments," *ASME Journal of Dynamic Systems, Measurement, and Control*, vol. 121, no. 4, pp. 655–659, 1999.

[23] J. C. Doyle, B. A. Francis, and A. Tannenbaum, *Feedback control theory.* Macmillan, 1992, vol. 134.

[24] B. D. Anderson, "From youla kucera to identification, adaptive and nonlinear control," *Automatica*, vol. 34, no. 12, pp. 1485 – 1506, 1998.

[25] I. D. Landau, A. C. Silva, T.-B. Airimitoaie, G. Buche, and M. Noe, "Benchmark on adaptive regulationąłrejection of unknown/time-varying multiple narrow band disturbances," *European Journal of Control*, vol. 19, no. 4, pp. 237 – 252, 2013. [Online]. Available: http://dx.doi.org/10.1016/j.ejcon.2013.05.007

[26] K. Ohnishi, "Robust motion control by disturbance observer," *Journal of the Robotics Society of Japan*, vol. 11, no. 4, pp. 486–493, 1993.

[27] C. J. Kempf and S. Kobayashi, "Disturbance observer and feedforward design for a high-speed direct-drive positioning table," *IEEE Transactions on Control Systems Technology*, vol. 7, no. 5, pp. 513–526, 1999.

[28] M. White, M. Tomizuka, and C. Smith, "Improved track following in magnetic disk drives using a disturbance observer," *IEEE/ASME Transactions on Mechatronics*, vol. 5, no. 1, pp. 3–11, Mar. 2000.

[29] K. S. Eom, I. H. Suh, and W. K. Chung, "Disturbance observer based path tracking control of robot manipulator considering torque saturation," *Mechatronics*, vol. 11, no. 3, pp. 325 – 343, 2001.

[30] K. K. Tan, T. H. Lee, H. F. Dou, S. J. Chin, and S. Zhao, "Precision motion control with disturbance observer for pulsewidth-modulated-driven permanent-magnet linear motors," *IEEE Transactions on Magnetics*, vol. 39, no. 3, pp. 1813–1818, 2003.

[31] K. Yang, Y. Choi, and W. K. Chung, "On the tracking performance improvement of optical disk drive servo systems using error-based disturbance observer," *IEEE Transactions on Industrial Electronics*, vol. 52, no. 1, pp. 270–279, Feb. 2005.

[32] B. Widrow, J. Glover Jr, J. McCool, J. Kaunitz, C. Williams, R. Hearn, J. Zeidler, E. Dong Jr, and R. Goodlin, "Adaptive noise cancelling: principles and applications," *Proceedings of the IEEE*, vol. 63, no. 12, pp. 1692–1716, 1975.

[33] M. Bodson, "Adaptive algorithms for the rejection of sinusoidal disturbances with unknown frequency," *Automatica*, vol. 33, no. 12, pp. 2213–2221, Dec. 1997.

[34] B. Francis and W. Wonham, "The internal model principle of control theory," *Automatica*, vol. 12, no. 5, pp. 457–465, 1976.

[35] I. D. Landau, A. Constantinescu, and D. Rey, "Adaptive narrow band disturbance rejection applied to an active suspension–an internal model principle approach," *Automatica*, vol. 41, no. 4, pp. 563–574, Apr. 2005.

[36] Q. Zhang and L. Brown, "Noise analysis of an algorithm for uncertain frequency identification," *IEEE Transactions on Automatic Control*, vol. 51, no. 1, pp. 103–110, 2006.

[37] W. Kim, H. Kim, C. Chung, and M. Tomizuka, "Adaptive output regulation for the rejection of a periodic disturbance with an unknown frequency," *IEEE Transactions on Control Systems Technology*, vol. 19, no. 5, pp. 1296–1304, 2011.

[38] F. Ben-Amara, P. T. Kabamba, and a. G. Ulsoy, "Adaptive sinusoidal disturbance rejection in linear discrete-time systems–part I: Theory," *ASME Journal of Dynamics Systems, Measurement, and Control*, vol. 121, no. 4, pp. 648–654, 1999.

[39] ——, "Adaptive sinusoidal disturbance rejection in linear discrete-time systems–part II: Experiments," *ASME Journal of Dynamic Systems, Measurement, and Control*, vol. 121, no. 4, pp. 655–659, 1999.

[40] C. Kinney, R. de Callafon, E. Dunens, R. Bargerhuff, and C. Bash, "Optimal periodic disturbance reduction for active noise cancelation," *Journal of Sound and Vibration*, vol. 305, no. 1-2, pp. 22–33, Aug. 2007.

[41] Y. Kim, C. Kang, and M. Tomizuka, "Adaptive and optimal rejection of non-repeatable disturbance in hard disk drives," in *Proc. 2005 IEEE/ASME International Conf. on Advanced Intelligent Mechatronics*, vol. 1, 2005, pp. 1–6.

[42] Q. Jia and Z. Wang, "A new adaptive method for identification of multiple unknown disturbance frequencies in hdds," *IEEE Transactions on Magnetics*, vol. 44, no. 11 Part 2, pp. 3746–3749, 2008.

[43] Q. Zheng and M. Tomizuka, "Compensation of Dominant Frequency Components of Nonrepeatable Disturbance in Hard Disk Drives," *IEEE Transactions on Magnetics*, vol. 43, no. 9, pp. 3756–3762, 2007.

[44] R. Roy and T. Kailath, "Esprit-estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 7, pp. 984 –995, Jul. 1989.

[45] R. Schmidt, “Multiple emitter location and signal parameter estimation,” *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276 – 280, Mar. 1986.

[46] M. Hayes, *Statistical digital signal processing and modeling.* Wiley-India, 2009.

[47] A. Nehorai, “A minimal parameter adaptive notch filter with constrained poles and zeros,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no. 4, pp. 983–996, Aug. 1985.

[48] G. Obregon-Pulido, B. Castillo-Toledo, and A. Loukianov, “A globally convergent estimator for n-frequencies,” *IEEE Transactions on Automatic Control*, vol. 47, no. 5, pp. 857 –863, May. 2002.

[49] L. Hsu, R. Ortega, and G. Damm, “A globally convergent frequency estimator,” *IEEE Transactions on Automatic Control*, vol. 44, no. 4, pp. 698 –713, Apr. 1999.

[50] M. Mojiri and A. Bakhshai, “An adaptive notch filter for frequency estimation of a periodic signal,” *IEEE Transactions on Automatic Control*, vol. 49, no. 2, pp. 314 – 318, Feb. 2004.

[51] X. Xia, “Global frequency estimation using adaptive identifiers,” *IEEE Transactions on Automatic Control*, vol. 47, no. 7, pp. 1188 –1193, Jul. 2002.

[52] B. Wu and M. Bodson, “A magnitude/phase-locked loop approach to parameter estimation of periodic signals,” *IEEE Transactions on Automatic Control*, vol. 48, no. 4, pp. 612 – 618, Apr. 2003.

[53] R. Marino and P. Tomei, “Global estimation of n unknown frequencies,” *IEEE Transactions on Automatic Control*, vol. 47, no. 8, pp. 1324 – 1328, Aug. 2002.

[54] R. Marino, G. L. Santosuosso, and P. Tomei, “Robust adaptive compensation of biased sinusoidal disturbances with unknown frequency,” *Automatica*, vol. 39, no. 10, pp. 1755 – 1761, 2003.

[55] L. Ljung, *System Identification: Theory for the User*, 2nd ed. Prentice Hall PTR, 1999.

[56] G. Goodwin and K. Sin, *Adaptive filtering prediction and control.* Prentice-Hall, 1984.

[57] C. Mohtadi, “Bode’s integral theorem for discrete-time systems,” *Control Theory and Applications, IEE Proceedings D*, vol. 137, no. 2, pp. 57–66, 1990.

[58] J. B. Hoagg and D. S. Bernstein, “Nonminimum-phase zeros,” *IEEE Control Systems Magazine*, vol. 27, no. 3, pp. 45–57, 2007.

[59] E. Davison, "The robust control of a servomechanism problem for linear time-invariant multivariable systems," *IEEE Transactions on Automatic Control*, vol. 21, no. 1, pp. 25–34, 1976.

[60] S. Boyd and L. Vandenberghe, *Convex optimization.* Cambridge university press, 2004.

[61] G. Stein, "Respect the unstable," *IEEE Control Systems*, vol. 23, no. 4, pp. 12 – 25, Aug. 2003.

[62] S. Skogestad and I. Postlethwaite, *Multivariable Feedback Control: Analysis and Design*, 2nd ed. Wiley Chichester, UK, 2005.

[63] A. Packard, "Multivariable Control Systems," *Class Notes for ME234 at University of California, Berkeley.*

[64] S. Mishra, "Fundamental issues in iterative learning controller design: Convergence, robustness, and steady state performance," Ph.D. dissertation, 2009.

[65] D. Y. Abramovitch and G. F. Franklin, "A brief history of disk drive control," *IEEE Control Systems Magazine*, vol. 22, no. 3, pp. 28–42, 2002.

[66] A. Al Mamun and S. S. Ge, "Precision control of hard disk drives," *IEEE Control Systems Magazine*, vol. 25, no. 4, pp. 14–19, 2005.

[67] IEEJ, Technical Commitee for Novel Nanoscale Servo Control, "NSS benchmark problem of hard disk drive systems," http://mizugaki.iis.u-tokyo.ac.jp/nss/, 2007.

[68] J. R. Deller, J. H. L. Hansen, and J. G. Proakis, *Discrete-time processing of speech signals.* Wiley-IEEE Press, Sep. 1999.

[69] G. Lin and Y.-J. Chen, "Disk flutter and its impact on HDD servo performance," *IEEE Transactions on Magnetics*, vol. 37, no. 2, pp. 866–870, 2001.

[70] R. Ehrlich and D. Curran, "Major HDD TMR sources and projected scaling with TPI," *IEEE Transactions on Magnetics*, vol. 35, no. 2, pp. 885–891, Mar 1999.

[71] J. McAllister, "The effect of disk platter resonances on track misregistration in 3.5 inch disk drives," *IEEE Transactions on Magnetics*, vol. 32, no. 3, pp. 1762–1766, May 1996.

[72] I. D. Landau, T.-B. Airimitoaie, A. C. Silva, and G. Buche, "An active vibration control system as a benchmark on adaptive regulation," in *2013 European Control Conf., Zurich, Switzerland, July 17-19 (to appear).*

[73] S. P. Boyd, C. Baratt, and S. Norman, “Linear controller design: limits of performance via convex optimization,” *Proceedings of the IEEE*, vol. 78, no. 3, pp. 529–574, Mar. 1990.

[74] J. Hespanha, “Switching between stabilizing controllers,” *Automatica*, vol. 38, no. 11, pp. 1905–1917, Nov. 2002.

[75] C.-C. Wang and M. Tomizuka, “Design of robustly stable disturbance observers based on closed loop consideration using h-infinity optimization and its applications to motion control systems,” in *Proceedings of the 2004 American Control Conference*, vol. 4, july 2004, pp. 3764 –3769.

[76] C. C. Kemp and S. Kobayashi, “Discrete-time disturbance observer design for systems with time delay,” in *Advanced Motion Control, 1996. AMC ’96-MIE. Proceedings., 1996 4th International Workshop on*, vol. 1, 1996, pp. 332–337 vol.1.

[77] Y. Choi, K. Yang, W. K. Chung, H. R. Kim, and I. H. Suh, “On the robustness and performance of disturbance observers for second-order systems,” *IEEE Transactions on Automatic Control*, vol. 48, no. 2, pp. 315–320, Feb. 2003.

[78] T. Murakami and K. Ohnishi, “Advanced motion control in mechatronics-A tutorial,” vol. 1, pp. 9–17.

[79] M. T. White, M. Tomizuka, and C. Smith, “Rejection of disk drive vibration and shock disturbances with a disturbance observer,” *Proceedings of the 1999 American Control Conference*, vol. 6, pp. 4127–4131, 1999.

[80] C. Bohn, V. Cortabarria, A andHärtel, and K. Kowalczyk, “Active control of engine-induced vibrations in automotive vehicles using disturbance observer gain scheduling,” *Control Engineering Practice*, vol. 12, no. 8, pp. 1029 – 1039, 2004.

[81] P. Regalia, S. Mitra, and P. Vaidyanathan, “The digital all-pass filter: A versatile signal processing building block,” *Proceedings of the IEEE*, vol. 76, no. 1, pp. 19–37, 1988.

[82] C. Scherer, P. Gahinet, and M. Chilali, “Multiobjective output-feedback control via LMI optimization,” *IEEE Transactions on Automatic Control*, vol. 42, no. 7, pp. 896–911, Jul. 1997.

[83] S. P. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear matrix inequalities in system and control theory.* Society for Industrial Mathematics, 1994, vol. 15.

[84] M. Grant and S. Boyd, “CVX: Matlab software for disciplined convex programming,” http://cvxr.com/cvx, Sep. 2012.

[85] M. Tomizuka, "Zero phase error tracking algorithm for digital control," *ASME Journal of Dynamic Systems, Measurements, and Control*, vol. 109, no. 1, pp. 65–68, 1987.

[86] J. a. Butterworth, L. Y. Pao, and D. Y. Abramovitch, "The effect of nonminimum-phase zero locations on the performance of feedforward model-inverse control techniques in discrete-time systems," *2008 American Control Conference*, pp. 2696–2702, Jun. 2008.

[87] S. Devasia, "Should model-based inverse inputs be used as feedforward under plant uncertainty?" *IEEE Transactions on Automatic Control*, vol. 47, no. 11, pp. 1865–1871, 2002.

[88] B. Widrow and E. Walach, *Adaptive Inverse Control, Reissue Edition: A Signal Processing Approach*. Wiley-IEEE Press, Nov. 2007.

[89] K. Astrom, P. Hagander, and J. Sternby, "Zeros of sampled systems," *Automatica*, vol. 20, no. 1, pp. 31 – 38, 1984.

[90] I. D. Landau and G. Zito, *Digital control systems: design, identification and implementation*. Springer Verlag, Dec. 2006.

[91] K. Ohnishi, M. Shibata, and T. Murakami, "Motion control for advanced mechatronics," *IEEE/ASME Trans. Mechatronics*, vol. 1, no. 1, pp. 56 –67, march 1996.

[92] B. A. Francis and W. M. Wonham, "The internal model principle for linear multivariable regulators," *Applied Mathematics & Optimization*, vol. 2, no. 2, pp. 170–194, Jun. 1975.

[93] T. K. S. M. T. Inoue, M. Nakano and H. Baba, "High accuracy control of a proton synchrotron magnet power supply," in *Proceedings of the IFAC World Congress*, 1981, pp. 3137–3142.

[94] S. Hara, Y. Yamamoto, T. Omata, and M. Nakano, "Repetitive control system: a new type servo system for periodic exogenous signals," *IEEE Transactions on Automatic Control*, vol. 33, no. 7, pp. 659 –668, Jul 1988.

[95] M. Tomizuka, T.-C. Tsao, and K.-K. Chew, "Analysis and synthesis of discrete-time repetitive controllers," *ASME Journal of Dynamic Systems, Measurement, and Control*, vol. 111, no. 3, pp. 353–358, 1989.

[96] G. Hillerström and K. Walgama, "Repetitive control theory and applications-a survey," in *Proceedings of 13th IFAC World Congress*, 1996, pp. 1–6.

[97] C. Li, D. Zhang, and X. Zhuang, "A survey of repetitive control," in *Proceedings of 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 16, no. 7, Jan. 2004, pp. 1160 – 1166.

[98] K. Chew and M. Tomizuka, "Digital control of repetitive errors in disk drive systems," *IEEE Control Systems Magazine*, vol. 10, no. 1, pp. 16–20, 1990.

[99] J. Moon, M. Lee, and M. Chung, "Repetitive control for the track-following servo system of an optical disk drive," *IEEE Transactions on Control Systems Technology*, vol. 6, no. 5, pp. 663–670, 2002.

[100] Y. Onuki and H. Ishioka, "Compensation for repeatable tracking errors in hard drives using discrete-time repetitive controllers," in *Proc. 6th IEEE International Workshop on Advanced Motion Control*, 2000, pp. 490–495.

[101] L. Guo, "Reducing the manufacturing costs associated with hard disk drives-a new disturbance rejection control scheme," *IEEE/ASME Transactions on Mechatronics*, vol. 2, no. 2, pp. 77–85, 2002.

[102] C. Cosner, G. Anwar, and M. Tomizuka, "Plug in repetitive control for industrial robotic manipulators," in *Proceedings of 1990 IEEE International Conference on Robotics and Automation,*, May 1990, pp. 1970 –1975 vol.3.

[103] R. W. Longman, "Iterative learning control and repetitive control for engineering practice," *International Journal of Control*, vol. 73, no. 10, pp. 930–954, 2000.

[104] T. Doh, J. Ryoo, and M. Chung, "Design of a repetitive controller: an application to the track-following servo system of optical disk drives," in *IEE Proceedings of Control Theory and Applications*, vol. 153, no. 3. IET, 2006, pp. 323–330.

[105] T. Inoue, "Practical repetitive control system design," in *Proceedings of the 29th IEEE Conference on Decision and Control*, Dec 1990, pp. 1673 –1678 vol.3.

[106] K. Srinivasan and F.-R. Shaw, "Analysis and design of repetitive control systems using the regeneration spectrum," in *Proceedings of American Control Conference*, May 1990, pp. 1150 –1155.

[107] M.-C. Tsai and W.-S. Yao, "Design of a plug-in type repetitive controller for periodic inputs," *IEEE Transactions on Control Systems Technology*, vol. 10, no. 4, pp. 547 –555, jul 2002.

[108] M. Tomizuka, "Dealing with periodic disturbances in controls of mechanical systems," *Annual Reviews in Control*, vol. 32, no. 2, pp. 193 – 199, 2008.

[109] M. Steinbuch, S. Weiland, and T. Singh, "Design of noise and period-time robust high-order repetitive control, with application to optical storage," *Automatica*, vol. 43, no. 12, pp. 2086–2095, 2007.

[110] G. Pipeleers, B. Demeulenaere, J. De Schutter, and J. Swevers, "Robust high-order repetitive control: optimal performance trade-offs," *Automatica*, vol. 44, no. 10, pp. 2628–2634, 2008.

[111] D. Liberzon and A. Morse, “Basic problems in stability and design of switched systems,” *IEEE Control Systems Magazine*, vol. 19, no. 5, pp. 59–70, 1999.

[112] H. Dotsch, H. Smakman, P. Van den Hof, and M. Steinbuch, “Adaptive repetitive control of a compact disc mechanism,” in *Proc. 34th IEEE Conf. on Decision and Control, 1995*, vol. 2. IEEE, 1995, pp. 1720–1725.

[113] Q.-W. Jia, “Disturbance rejection through disturbance observer with adaptive frequency estimation,” *IEEE Transactions on Magnetics*, vol. 45, no. 6, pp. 2675–2678, June 2009.

[114] K. Zhou and J. C. Doyle, *Essentials of robust control.* Prentice Hall New Jersey, Oct. 1998.

[115] R. Horowitz, Y. Li, K. Oldham, S. Kon, and X. Huang, “Dual-stage servo systems and vibration compensation in computer hard disk drives,” *Control Engineering Practice*, vol. 15, no. 3, pp. 291–305, Mar. 2007.

[116] A. Al Mamun, I. Mareels, T. Lee, and A. Tay, “Dual stage actuator control in hard disk drive-a review,” in *The 29th Annual Conference of the IEEE Industrial Electronics Society*, vol. 3, 2003, pp. 2132–2137.

[117] L. Guo, D. Martin, and D. Brunnett, “Dual-stage actuator servo control for high density disk drives,” *IEEE/ASME International Conference on Advanced Intelligent Mechatronics*, pp. 132–137, 1999.

[118] I. D. Landau, R. Lozano, and M. M’Saad, *Adaptive Control*, J. W. Modestino, A. Fettweis, J. L. Massey, M. Thoma, E. D. Sontag, and B. W. Dickinson, Eds. Springer-Verlag New York, Inc., 1998.

[119] W. Rudin, *Principles of Mathematical Analysis.* McGraw-Hill New York, 1976, vol. 275.

[120] A. Oppenheim, R. Schafer, and J. Buck, *Discrete-time signal processing.* Prentice hall Englewood Cliffs, NJ, 1989.

[121] B.-S. Chen, T.-Y. Yang, and B.-H. Lin, “Adaptive notch filter by direct frequency estimation,” *Signal Processing*, vol. 27, no. 2, pp. 161 – 176, 1992. [Online]. Available: http://www.sciencedirect.com/science/article/B6V18-48TD1MD-4H/2/16039e61516b8163661a103112613bdf

[122] G. Li, “A stable and efficient adaptive notch filter for direct frequency estimation,” *IEEE Transactions on Signal Processing*, vol. 45, no. 8, pp. 2001–2009, Aug. 1997.

[123] S. Dasgupta and A. Bhagwat, “Conditions for designing strictly positive real transfer functions for adaptive output error identification,” *IEEE Transactions on Circuits and Systems*, vol. 34, no. 7, pp. 731 – 736, Jul. 1987.

[124] L. Ljung, "On positive real transfer functions and the convergence of some recursive schemes," *IEEE Transactions on Automatic Control*, vol. 22, no. 4, pp. 539 – 551, Aug. 1977.

[125] P. Ioannou and G. Tao, "Frequency domain conditions for strictly positive real functions," *IEEE Transactions on Automatic Control*, vol. 32, no. 1, pp. 53 – 54, Jan. 1987.

[126] B. Anderson, S. Dasgupta, P. Khargonekar, F. Kraus, and M. Mansour, "Robust strict positive realness: characterization and construction," *IEEE Transactions on Circuits and Systems*, vol. 37, no. 7, pp. 869 –876, Jul. 1990.

[127] A. Betser and E. Zeheb, "Design of robust strictly positive real transfer functions," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 40, no. 9, pp. 573 –580, Sep. 1993.

[128] B. Anderson and I. Landau, "Least squares identification and the robust strict positive real property," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 41, no. 9, pp. 601 –607, Sep. 1994.

[129] H. Marquez and P. Agathoklis, "On the existence of robust strictly positive real rational functions," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 45, no. 9, pp. 962 –967, Sep. 1998.

[130] G. Bianchini, A. Tesi, and A. Vicino, "Synthesis of robust strictly positive real systems with l2 parametric uncertainty," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 48, no. 4, pp. 438 –450, Apr. 2001.

[131] H. Marquez and C. Damaren, "On the design of strictly positive real transfer functions," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 42, no. 4, pp. 214 –218, Apr. 1995.

[132] A. Tesi, G. Zappa, and A. Vicino, "Enhancing strict positive realness condition on families of polynomials by filter design," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 40, no. 1, pp. 21 –32, Jan. 1993.

[133] A. Tesi, A. Vicino, and G. Zappa, "Design criteria for robust strict positive realness in adaptive schemes," *Automatica*, vol. 30, no. 4, pp. 643 – 654, 1994.

[134] L. Wang and W. Yu, "Complete characterization of strictly positive real regions and robust strictly positive real synthesis method," *Science in China Series E: Technological Sciences*, vol. 43, no. 1, pp. 97–112, 2000.

[135] M. Fu and S. Dasgupta, "New extreme point results on robust strict positive realness," in *Proceedings of the 35th IEEE Decision and Control Conference, 1996*, vol. 3, Dec. 1996, pp. 2872 –2877 vol.3.

[136] D. Henrion, "Linear matrix inequalities for robust strictly positive real design," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 49, no. 7, pp. 1017 –1020, Jul. 2002.

[137] L. Ljung and T. Soderstrom, *Theory and Practice of Recursive Identification.* Mass: MIT Press, 1983.

[138] J. Wen, "Time domain and frequency domain conditions for strict positive realness," *IEEE Transactions on Automatic Control*, vol. 33, no. 10, pp. 988 –992, Oct. 1988.

[139] H. Khalil and J. Grizzle, *Nonlinear systems*, 3rd ed. Prentice hall, 2002.

[140] B. Barmish, "Invariance of the strict hurwitz property for polynomials with perturbed coefficients," *IEEE Transactions on Automatic Control*, vol. 29, no. 10, pp. 935 – 936, Oct. 1984.

[141] W. Karl, J. Greschak, and G. Verghese, "Comments on «a necessary and sufficient condition for the stability of interval matrices»," *International Journal of Control*, vol. 39, no. 4, pp. 849–851, 1984.

[142] A. Bartlett, C. Hollot, and H. Lin, "Root locations of an entire polytope of polynomials: It suffices to check the edges," *Mathematics of Control, Signals, and Systems (MCSS)*, vol. 1, no. 1, pp. 61–71, 1988.

[143] C. Hollot and A. Bartlett, "Some discrete-time counterparts to kharitonov's stability criterion for uncertain systems," *IEEE Transactions on Automatic Control*, vol. 31, no. 4, pp. 355 – 356, Apr. 1986.

[144] L. Lee and J. L. Chen, "Strictly positive real lemma and absolute stability for discrete-time descriptor systems," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 50, no. 6, pp. 788 – 794, 2003.

[145] W. M. Haddad and D. S. Bernstein, "Explicit construction of quadratic Lyapunov functions for the small gain, positivity, circle, and popov theorems and their application to robust stability. part ii: Discrete-time theory," *International Journal of Robust and Nonlinear Control*, vol. 4, no. 2, pp. 249–265, 1994.

[146] M. Mahmoud and L. Xie, "Positive real analysis and synthesis of uncertain discrete time systems," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 47, no. 3, pp. 403 –406, Mar. 2000.

[147] T. Yamaguchi, K. Shishida, S. Tohyama, and H. Hirai, "Mode switching control design with initial value compensation and its application to head positioning control on magnetic disk drives," *IEEE Transactions on Industrial Electronics*, vol. 43, no. 1, pp. 65–73, 1996.

[148] G. Herrmann, M. C. Turner, and I. Postlethwaite, "Discrete-time and sampled-data anti-windup synthesis: stability and performance," *International Journal of Systems Science*, vol. 37, no. 2, pp. 91–113, 2006.

[149] P. F. Weston and I. Postlethwaite, "Linear conditioning for systems containing saturating actuators," *Automatica*, vol. 36, no. 9, pp. 1347–1354, 2000.

[150] A. Zheng, M. V. Kothare, and M. Morari, "Anti-windup design for internal model control," *International Journal of Control*, vol. 60, no. 5, pp. 1015–1024, 1994.

[151] M. Steinbuch, "Repetitive control for systems with uncertain period-time," *Automatica*, vol. 38, no. 12, pp. 2103–2109, 2002.

[152] M. Yamada, Z. Riadh, and Y. Funahashi, "Design of robust repetitive control system for multiple periods," in *Proceedings of the 39th IEEE Conference on Decision and Control*, vol. 4, 2000, pp. 3739–3744 vol.4.

[153] B.-S. Kim and T.-C. Tsao, "A performance enhancement scheme for robust repetitive control system," *Journal of dynamic systems, measurement, and control*, vol. 126, no. 1, pp. 224–229, 2004.

[154] M. M. Seron, J. H. Braslavsky, and G. C. Goodwin, *Limitations in Filtering and Control*. London: Springer-Verlag, 1997.

Appendix A

Proof of Theorems

A.1 Proof of Bode's Integral Formula

Theorem. *Let $L(s)$ be a proper, scalar rational transfer function, of relative degree larger than 1. Let $S(s) = (1 + L(s))^{-1}$ and assume that $S(s)$ has no poles in the right half plane, and has $q \geq 0$ zeros in the closed right half plane, at locations p_1, p_2, ..., p_q. Then*

$$\int_0^{\infty} \ln |S(j\omega)| \, d\omega = \pi \sum_{k=1}^{q} p_k$$

Proof. (key steps only) Consider the simple case where we have a real unstable pole in $L(s)$. We construct the complex integral with s shown in the contour in Figure A.1. Here $R \to \infty$. Since $\ln S(s)$ is analytic within the contour, the whole contour integral sums up to zero. This is the result of Cauchy Integral Theorem. It is not difficult to show that the part of the integral along the arc with radius R is zero under the assumption of relative degree larger than 1.[1] Therefore the integral along the imaginary axis (which is the quantity that we want to compute) plus the integral along the contour C (consisting of the path $I \to II \to III$) in Figure A.2 is zero, namely, when $R \to \infty$

$$\int_{-j\infty}^{0} \ln S(s) \, ds + \int_0^{j\infty} \ln S(s) \, ds + \lim_{\epsilon \to 0} \int_C \ln S(s) \, ds = 0 \tag{A.1}$$

Now we focus on the contour C in Figure A.2. Decompose first

[1]To see this, note that when $L(s)$ is small, a Taylor expansion for $\ln(1 + L(s))$ gives

$$\int_R \ln S(s) \, ds = -\int_R \ln(1 + L(s)) \, ds \approx -\int_R (\ln 1 + L(s)) \, ds \approx -\int_R L(s) ds$$

Since $L(s)$ decays to zero at a rate that is at least as fast as $1/s^2$ for large s, the above integral goes to zero when the radius of the circle goes to infinity.

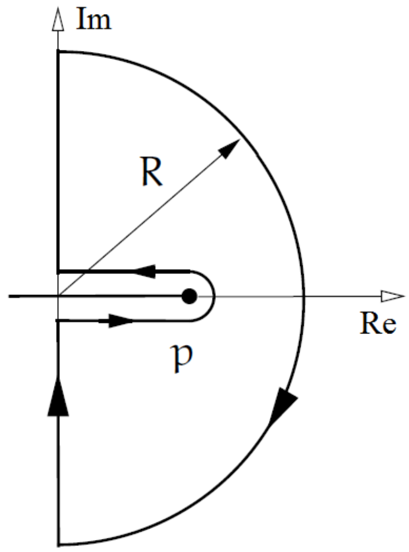


Figure A.1: Contour of s for Bode's Integral

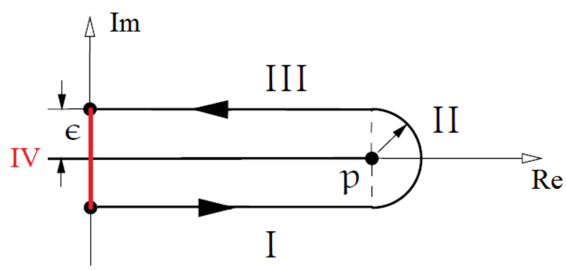


Figure A.2: Partial contour for Bode's Integral ($\epsilon \to 0$)

$$
\begin{aligned}
& S(s) = (s - p) S^*(s) \\
\Rightarrow \int_C \ln S(s)\, ds = & \int_C \ln (s - p)\, ds + \int_C \ln S^*(s)\, ds
\end{aligned} \tag{A.2}
$$

so that we can separate the analytic part of $\ln S(s)$ as $\ln S^*(s)$. We will show that as $\epsilon \to 0$, $\int_C \ln S^*(s)\, ds \to 0$ and $\int_C \ln(s-p)\, ds$ approaches to some constant value that will show up in Bode's Integral Formula. For the first part, if we add a path IV to make a closed contour $I \to II \to III \to IV$, we have

$$
\oint \ln S^*(s)\, ds = 0
$$

due to the fact that $\ln S^*(s)$ is analytic on and within the contour. Hence

$$\begin{aligned}
&\int_C \ln S^*(s)\,ds + \int_{IV} \ln S^*(s)\,ds = 0 \\
&\qquad\Rightarrow \int_C \ln S^*(s)\,ds = -\int_{IV} \ln S^*(s)\,ds = -\int_{\epsilon j}^{-\epsilon j} \ln S^*(s)\,ds \\
&\qquad\qquad = \int_{-\epsilon j}^{\epsilon j} \ln S^*(s)\,ds
\end{aligned}$$

In this way we need to just compute a line integral. The function $\ln S^*(s)$ is analytic so it is bounded by some finite value $f_m > 0$, therefore

$$\left|\int_{-\epsilon j}^{\epsilon j} \ln S^*(s)\,ds\right| \le \int_{-\epsilon j}^{\epsilon j} |\ln S^*(s)|\,ds \le \int_{-\epsilon j}^{\epsilon j} f_m ds = f_m 2\epsilon j \to 0 \tag{A.3}$$

Now switch to proving the second part. This needs just some small steps of algebra. As

$$\ln x dx = d\left[x\ln x - x\right]$$

we have

$$\begin{aligned}
\int_C \ln(s-p)\,ds &= \int_C d\left[(s-p)\ln(s-p) - (s-p)\right] \\
&= \left[(s-p)\ln(s-p) - (s-p)\right]\Big|_{-\epsilon j}^{\epsilon j} \\
&= \left[(s-p)\ln(s-p)\right]\Big|_{-\epsilon j}^{\epsilon j} + \left[-(s-p)\right]\Big|_{-\epsilon j}^{\epsilon j} \\
&= \left[s\ln(s-p)\right]\Big|_{-\epsilon j}^{\epsilon j} + \left[-p\ln(s-p)\right]\Big|_{-\epsilon j}^{\epsilon j} + \left[-(s-p)\right]\Big|_{-\epsilon j}^{\epsilon j}
\end{aligned}$$

The terms $\left[s\ln(s-p)\right]\Big|_{-\epsilon j}^{\epsilon j}$ and $\left[-(s-p)\right]\Big|_{-\epsilon j}^{\epsilon j}$ all go to zero as $\epsilon \to 0$, for the remaining term we use the property of log functions:

$$\ln x = \ln\left(|x|\,e^{j\angle x}\right) = \ln|x| + \ln e^{j\angle x} = \ln|x| + j\angle x$$

and have

$$\begin{aligned}
\lim_{\epsilon\to 0}\left[-p\ln(s-p)\right]\Big|_{-\epsilon j}^{\epsilon j} &= \lim_{\epsilon\to 0}\left[-p\ln|s-p| - pj\angle(s-p)\right]\Big|_{-\epsilon j}^{\epsilon j} \\
&= \lim_{\epsilon\to 0}\left[-pj\angle(s-p)\right]\Big|_{-\epsilon j}^{\epsilon j}
\end{aligned}$$

Draw a picture of the vector $s-p$ in Figure A.2. We will see that as s goes along the contour starting at $-\epsilon j$ and ending at ϵj, the angular change of $\angle(s-p)$ is 2π as $\epsilon \to 0$. Hence

$$\int_C \ln(s-p)\,ds = \left[-pj\angle(s-p)\right]\Big|_{-\epsilon j}^{\epsilon j} \to -2\pi pj \tag{A.4}$$

Combining (A.2) (A.3) and (A.4) we get

$$\int_C \ln S(s)\, ds \to -2\pi p j$$

as $\epsilon \to 0$. Using (A.1), we obtain

$$\int_{-j\infty}^{0} \ln S(s)\, ds + \int_{0}^{j\infty} \ln S(s)\, ds = 2\pi p j$$

When there are multiple unstable open-loop poles, the above analysis can be easily extended and we have

$$\int_{-j\infty}^{0} \ln S(s)\, ds + \int_{0}^{j\infty} \ln S(s)\, ds = 2j\pi \sum_k Re(p_k) = 2j\pi \sum_k p_k \tag{A.5}$$

In control engineering we prefer using ω instead of s in the left half side of the above equation. To make this happen, we note that

$$\begin{aligned}
\int_{-j\infty}^{0} \ln S(s)\, ds + \int_{0}^{j\infty} \ln S(s)\, ds &= j\int_{-\infty}^{0} \ln S(j\omega)\, d\omega + j\int_{0}^{\infty} \ln S(j\omega)\, d\omega \\
&= j\int_{0}^{\infty} \ln S(-j\omega)\, d\omega + j\int_{0}^{\infty} \ln S(j\omega)\, d\omega \\
&= j\int_{0}^{\infty} \left[\ln S(-j\omega) + \ln S(j\omega)\right] d\omega \\
&= j\int_{0}^{\infty} \ln\left[S(-j\omega)\, S(j\omega)\right] d\omega \\
&= 2j\int_{0}^{\infty} \ln|S(j\omega)|\, d\omega
\end{aligned}$$

Putting the above result to (A.5), we obtain the final conclusion

$$\boxed{\int_{0}^{\infty} \ln|S(j\omega)|\, d\omega = \pi \sum_k p_k}$$

□

A.2 Proof of Internal Model Principle

Consider the block diagram in Figure A.3. The relationship between the disturbance $d(k)$ and the output $y(k)$ is

$$y(k) = \frac{1}{1 + P(q^{-1})C(q^{-1})} d(k)$$

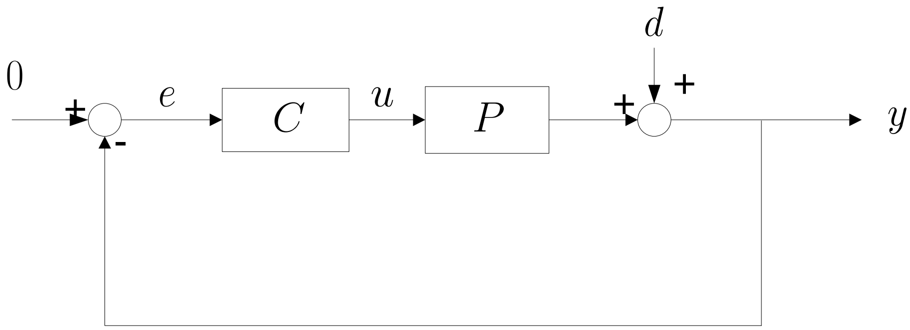


Figure A.3: Explanation of internal model principle

When

$$d(k) = \frac{B_d(q^{-1})}{A_d(q^{-1})}\delta(k)$$

we have

$$y(k) = \frac{1}{1 + P(q^{-1})C(q^{-1})}\frac{B_d(q^{-1})}{A_d(q^{-1})}\delta(k) \tag{A.6}$$

Let $P(q^{-1}) = B_p(q^{-1})/A_p(q^{-1})$, and $C(q^{-1}) = B_c(q^{-1})/A_c(q^{-1})$. (A.6) becomes

$$y(k) = \frac{A_p(q^{-1})A_c(q^{-1})}{A_p(q^{-1})A_c(q^{-1}) + B_p(q^{-1})B_c(q^{-1})}\frac{B_d(q^{-1})}{A_d(q^{-1})}\delta(k)$$

If $A_c(q^{-1}) = A_d(q^{-1})A_c'(q^{-1})$ then

$$y(k) = \frac{A_p(q^{-1})A_c'(q^{-1})B_d(q^{-1})}{A_p(q^{-1})A_c(q^{-1}) + B_p(q^{-1})B_c(q^{-1})}\delta(k)$$

which indicates that $y(k)$ converges to zero if the closed-loop system is asymptotically stable, hence the rejection of the disturbance. The convergence speed here depends on the closed-loop poles. Notice that the inclusion of $A_d(q^{-1})$ in the denominator of the controller builds an anti-disturbance signal generator for cancellation of $d(k)$.

A.3 Proof for Section 5.3: Zeros of an IIR Filter Structure

The Q-filter numerator in (5.14) can be partitioned into

$$B_Q(z) = b_n z^{-n} + \sum_{i=0}^{n-1}\left(b_i z^{-i} + b_{2n-i} z^{-2n+i}\right), \tag{A.7}$$

where $b_i = (\alpha^i - k)a_i$, $a_0 = 1$, and $a_i = a_{2n-i}$. Letting $k = \alpha^n$ gives $b_n = 0$ and

$$\begin{aligned} & b_i z^{-i} + b_{2n-i} z^{-2n+i} \\ =& \alpha^n a_i z^{-i} \left[\left(\alpha^{i-n} - 1 \right) + \left(\alpha^{-i+n} - 1 \right) z^{-2n+2i} \right] \end{aligned} \tag{A.8}$$

We claim that $b_i z^{-i} + b_{2n-i} z^{-2n+i}$ always contains the factor $1 - \alpha z^{-2}$. To see this, substituting in $z^{-2} = \alpha^{-1}$ to (A.8), we can observe that $\left[(\alpha^{i-n} - 1) + (\alpha^{-i+n} - 1) z^{-2n+2i} \right] = \left[(\alpha^{i-n} - 1) + (\alpha^{-i+n} - 1) \alpha^{i-n} \right] = 0$, which proves that $1 - \alpha z^{-2}$ is a common factor of $b_i z^{-i} + b_{2n-i} z^{-2n+i}$. Since $b_n = 0$, $B_Q(z)$ in (A.7) thus contains the common factor $1 - \alpha z^{-2}$.

A.4 Proof for Theorem 6.1: Waterbed Effect in Local Loop Shaping

We need first two fundamental results about complex integrals.

Theorem A.1. *(Poisson Integral Formula for the disk) If f is a function analytic[2] on and inside the unit disk, then for any interior points $s_0 = r_0 e^{j\theta_0}$, $r_0 < 1$, we have*

$$f\left(r_0 e^{j\theta_0}\right) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f\left(e^{j\theta}\right) \frac{1 - r_0^2}{1 - 2r_0 \cos\left(\theta - \theta_0\right) + r_0^2} d\theta \tag{A.9}$$

Proof. See any text book on complex analysis. One version that is more orientated to control engineering is in [154]. □

Corollary A.1. *If f is a function harmonic[3] outside the unit disk, then for any point exterior to the unit disk $s_0 = r_0 e^{j\theta_0}$, $r_0 > 1$, we have*

$$f\left(r_0 e^{j\theta_0}\right) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f\left(e^{j\theta}\right) \frac{r_0^2 - 1}{1 - 2r_0 \cos\left(\theta - \theta_0\right) + r_0^2} d\theta. \tag{A.10}$$

Proof. The proof follows from Theorem A.1 by taking the real part of (A.9) and considering $f\left(1/s\right)$. Interested readers can refer to [154] for more details. □

Now we formaly prove Theorem 6.1.

Proof. First, $1 - z^{-m} Q\left(z\right)$ is strictly stable. The roots of $A\left(z\right)$ are on the unit circle. If the stable $K\left(z\right)$ does not have zeros outside the unit circle, then $\ln\left(1 - z^{-m} Q\left(z\right)\right)$ is analytic

[2]One intuition for a function to be analytic in domain D is that it must not have singularities (poles) in D.

[3]A harmonic function is formally defined as a real-valued function $f\left(\sigma, \omega\right)$ that satisfies the Laplace equation $\frac{\partial^2 f}{\partial \sigma^2} + \frac{\partial^2 f}{\partial \omega^2} = 0$ (for the 2-dimentional case). A particular case that is interesting to us is that the real and the complex parts of analytic functions are harmonic.

outside the unit disk. Hence, the real part of $\ln(1 - z^{-m}Q(z))$, with $z = re^{j\omega}$, is a harmonic function. From Corollary A.1, letting $z_0 = r_0 e^{j\omega_0}$ with $r_0 > 1$, we have

$$Re\left\{\ln\left(1 - z_0^{-m}Q(z_0)\right)\right\} = \ln\left|1 - z_0^{-m}Q(z_0)\right| = \frac{1}{2\pi}\int_{-\pi}^{\pi}\ln\left|1 - e^{-mj\omega}Q\left(e^{j\omega}\right)\right|\frac{r_0^2 - 1}{1 - 2r_0\cos(\omega - \omega_0) + r_0^2}d\omega \tag{A.11}$$

Next, take the limit $r_0 \to \infty$. The right hand side of (A.11) tends to

$$\begin{aligned}
&\frac{1}{2\pi}\int_{-\pi}^{\pi}\ln\left|1 - e^{-mj\omega}Q\left(e^{j\omega}\right)\right|d\omega \\
=&\frac{1}{2\pi}\int_{0}^{\pi}\ln\left|1 - e^{-mj\omega}Q\left(e^{j\omega}\right)\right|d\omega + \frac{1}{2\pi}\int_{-\pi}^{0}\ln\left|1 - e^{-mj\omega}Q\left(e^{j\omega}\right)\right|d\omega \\
=&\frac{1}{2\pi}\int_{0}^{\pi}\ln\left|1 - e^{-mj\omega}Q\left(e^{j\omega}\right)\right|d\omega + \frac{1}{2\pi}\int_{-\pi}^{0}\ln\left|1 - e^{mj\omega}Q\left(e^{-j\omega}\right)\right|d\omega \\
=&\frac{1}{\pi}\int_{0}^{\pi}\ln\left|1 - e^{-mj\omega}Q\left(e^{j\omega}\right)\right|d\omega
\end{aligned} \tag{A.12}$$

Notice that for $m > 0$,

$$\lim_{z_0 \to \infty}\left(1 - z_0^{-m}Q(z_0)\right) = 1$$

Hence we have the following limiting situation of (A.11):

$$\frac{1}{\pi}\int_{0}^{\pi}\ln\left|1 - e^{-mj\omega}Q\left(e^{j\omega}\right)\right|d\omega = 0$$

Following the proof of Bode's Integral Formula (Appendix A.1), if $K(z)$ introduces additional unstable zeros, denoted as γ_i, $i = 1, \ldots, l$, we can perform all-pass factorization and let

$$1 - z^{-m}Q(z) = f(z)\prod_{i=1}^{l}\frac{\bar{\gamma}_i}{|\gamma_i|}\frac{z - \gamma_i}{1 - \bar{\gamma}_i z}$$

where $f(z)$ is minimum-phase. Then

$$\ln\left|1 - e^{-mj\omega}Q\left(e^{j\omega}\right)\right| = \ln\left|f\left(e^{j\omega}\right)\right|$$

and (A.11) becomes

$$\begin{aligned}
\ln|f(z_0)| &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\ln\left|f\left(e^{j\omega}\right)\right|\frac{r_0^2 - 1}{1 - 2r_0\cos(\omega - \omega_0) + r_0^2}d\omega \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi}\ln\left|1 - e^{-mj\omega}Q\left(e^{j\omega}\right)\right|\frac{r_0^2 - 1}{1 - 2r_0\cos(\omega - \omega_0) + r_0^2}d\omega
\end{aligned} \tag{A.13}$$

Noting

$$\begin{aligned}\ln\left|1-z_0^{-m}Q(z_0)\right| &= \ln|f(z_0)|+\ln\left|\prod_{i=1}^{n}\frac{\bar{\gamma}_i}{|\gamma_i|}\frac{z_0-\gamma_i}{1-\bar{\gamma}_i z_0}\right| \\ &= \ln|f(z_0)|+\sum_{i=1}^{n}\ln\left|\frac{\bar{\gamma}_i}{|\gamma_i|}\frac{z_0-\gamma_i}{1-\bar{\gamma}_i z_0}\right|\end{aligned}$$

and taking the limit of $z_0 \to \infty$, we get

$$\lim_{z_0\to\infty}\ln|f(z_0)| = \sum_{i=1}^{l}\ln|\gamma_i| \tag{A.14}$$

Combining (A.13) and (A.14) gives

$$\frac{1}{\pi}\int_0^{\pi}\ln\left|1-e^{-mj\omega}Q\left(e^{j\omega}\right)\right|d\omega = \sum_{i=1}^{l}\ln|\gamma_i| > 0$$

□

A.5 Proof for Section 10.2: Stability of the Series-Parallel Predictor

The adaptation law is

$$\hat{\theta}(k) = \hat{\theta}(k-1) + \frac{F(k-1)\,\psi(k-1)\,e^o(k)}{1+\psi(k-1)^T F(k-1)\,\psi(k-1)} \tag{A.15}$$

or, equivalently

$$\hat{\theta}(k+1) = \hat{\theta}(k) + \frac{F(k)\,\psi(k)\,e^o(k+1)}{1+\psi(k)^T F(k)\,\psi(k)}. \tag{A.16}$$

For stability analysis, we first transform the PAA to the *a posteriori* form. We have:

- the ideal output for a tuned controller ($y(k)=0$):

$$\begin{aligned}0 = y(k+1) &= \psi(k)^T\theta + \left(w(k+1)+w(k-2n+1)-\alpha^{2n}y(k-2n+1)\right) \\ &= \psi(k)^T\theta + (w(k+1)+w(k-2n+1))\end{aligned}$$

- the *a posteriori* prediction of $y(k+1)$:

$$\hat{y}(k+1) = \psi(k)^T\hat{\theta}(k+1) + (w(k+1)+w(k-2n+1)) \tag{A.17}$$

- the *a posteriori* prediction error:

$$e(k+1) = y(k+1) - \hat{y}(k+1) = -\psi(k)^T \tilde{\theta}(k+1) \tag{A.18}$$

where
$$\tilde{\theta}(k+1) = \hat{\theta}(k+1) - \theta \tag{A.19}$$

- the *a priori* prediction error:

$$e^o(k+1) = -\psi(k)^T \tilde{\theta}(k) \tag{A.20}$$

Pre-multiplying $\psi^T(k)$ to (A.16) yields

$$\psi^T(k)\hat{\theta}(k+1) = \psi^T(k)\hat{\theta}(k) + \frac{\psi^T(k)F(k)\psi(k)}{1+\psi^T(k)F(k)\psi(k)} e^o(k+1). \tag{A.21}$$

Subtracting $\psi^T(k)\theta$ from each side in (A.21), and substituting in (A.18) and (A.20), we have
$$e(k+1) = \frac{e^o(k+1)}{1+\psi^T(k)F(k)\psi(k)}. \tag{A.22}$$

Substituting (A.22) back to (A.16), we arrive at the PAA in the *a posteriori* form:

$$\hat{\theta}(k+1) = \hat{\theta}(k) + F(k)\psi(k)e(k+1) \tag{A.23}$$
$$e(k+1) = -\psi(k)^T \tilde{\theta}(k+1) \tag{A.24}$$

From (A.23) and (A.19), we get

$$\tilde{\theta}(k+1) = \tilde{\theta}(k) + F(k)\psi(k)e(k+1). \tag{A.25}$$

Combining (A.22) and (A.25), we can construct the equivalent feedback loop for the adaptive system as shown in Figure A.4.

The nonlinear block NL in Figure A.4 is shown to be passive and satisfies the Popov Inequality (Section 3.3.4 of [118]). The linear block $L = 1 - 1/2$ is strictly positive real. Therefore, the parameter adaptation algorithm is asymptotically hyperstable. Applying Theorem 3.3.2 from [118], we have

$$\lim_{k\to 0} e(k) = 0. \tag{A.26}$$

Recalling that $e(k) = -\psi(k-1)^T \tilde{\theta}(k)$, $\theta = [a_1, a_2, \ldots, a_n]^T$, and

$$\psi_i(k-1) = w(k-i) + w(k-2n+i); \quad i = 1, ..., n-1, \tag{A.27}$$
$$\psi_n(k-1) = w(k-n), \tag{A.28}$$

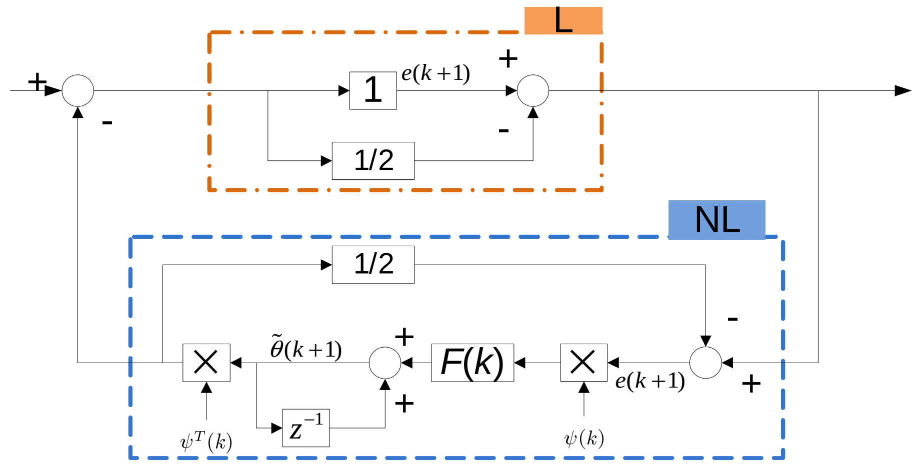


Figure A.4: Equivalent feedback loop of the adaptive system

(see Section 10.2), we have

$$\begin{aligned} e(k) =& \psi(k-1)^T \tilde{\theta}(k) \\ =& \sum_{i=1}^{n-1} \left(w(k-i) + w(k-2n+i) \right) \tilde{a}_i(k) + w(k-n)\, \tilde{a}_n(k) \\ =& \left(\sum_{i=1}^{n-1} \left(q^{-i} + q^{-2n+i} \right) \tilde{a}_i(k) + q^{-n} \tilde{a}_n(k) \right) w(k) \\ \to& 0 \qquad \text{as } k \to \infty. \end{aligned} \tag{A.29}$$

Based on the assumption that $w(k)$ has n independent frequency components, the Frequency Richness Condition for Parameter Convergence holds. Therefore, the only solution to the above equation is $\lim_{k\to\infty} \tilde{a}_i(k) = 0$, i.e., the parameters converge to their true values.

Appendix B

Mathematical Backgrounds

B.1 The Diophantine Equation

An important problem in algebra is to solve the polynomial equation:

$$AX + BY = F \tag{B.1}$$

where A, B, F are polynomials of real coefficients; X and Y are unknown polynomials to be solved. (B.1) is called the Diophantine equation, named after the Alexandrian Greek mathematician Diophantus (200 AD to 298 AD).

Remark B.1. One immediate application of the Diophantine equation in control theory, is to design feedback systems by pole placement, where F is the assigned closed-loop characteristic polynomial; and the transfer function of the plant is B/A.

The simplest case of the problem is when all polynomials are constrained to be scalar numbers, e.g. $3x + 5y = 1$ where x and y are integers. A particular solution pair is $(x, y) = (7, -4)$. The solution is not unique, as we can see that any integer Q in the following equations also yields a valid solution:

$$\begin{aligned} x &= 7 + 5Q \\ y &= -4 - 3Q. \end{aligned}$$

The solution set may also be empty. Consider, for example, the integer equation $4x+2y = 1$. The left-hand side of the equality is always even while the right-hand side is always odd.

We discuss the existence and computation of the solutions below. For control engineering, most commonly we use the matrix method for computation.

Existence of solutions

The Diophantine equation has solutions if and only if the greatest common factor (gcf) of A and B divides F. For example, the gcf of 3 and 5 are 1 in $3x + 5y = 1$, while in $4x + 2y = 1$, the gcf of 4 and 2 is 2, which does not divide 1.

Formal proof*[1] A formal proof of the result naturally follows during finding the gcf by the Euclid's algorithm. This is an iterative algorithm initialized by $A_0 = A$ and $B_0 = B$. At every step, let $A_{n+1} = B_n$ and B_{n+1} be the remainder of A_n divided by B_n. The iteration ends when $B_{n+1} = 0$. The gcf is then $G = B_n$. For instance, consider the gcf of 48 and 18. The solution is tabulated as follows

iteration number	0	1	2	3
A_i	48	18	12	$\boxed{6 \text{ (gcf)}}$
B_i	18	$12\,(= 48 - 18 \times 2)$	$6\,(= 18 - 12)$	$0\,(= 12 - 6 \times 2)$

The procedure is the same in solving polynomial Diophantine equations. Consider the following example

iteration number	0	1	2
A_i	$1 + q^{-1}$	$1 - q^{-2}$	$\boxed{1 + q^{-1} \text{ (gcf)}}$
B_i	$1 - q^{-2}$	$1 + q^{-1}(= 1 + q^{-1} - (1 - q^{-2}) \times 0)$	$0(= 1 - q^{-2} - (1 + q^{-1}) \times (1 - q^{-1}))$

There are no division steps in the solution tables. We can actually backtrack to find

$$AX' + BY' = B_n \triangleq G. \tag{B.2}$$

If G divides F then we have $F = GF'$, yielding

$$A\underbrace{X'F'}_{X} + B\underbrace{Y'F'}_{Y} = F. \tag{B.3}$$

Computation of solutions

Polynomial method*

When the gcf satisfies

$$F = GF' \tag{B.4}$$

a particular solution can be found from (B.2) and (B.3). Solving additionally the minimum-order solutions N and D for

$$AN + BD = 0$$

we can get the general solution

$$\tilde{X} = X'F + NQ \tag{B.5}$$

$$\tilde{Y} = Y'F - DQ \tag{B.6}$$

[1]Materials marked by * are supplementary.

The above can be combined in the following polynomial matrix:

$$\begin{bmatrix} X' & Y' \\ N & D \end{bmatrix} \begin{bmatrix} A & 1 & 0 \\ B & 0 & 1 \end{bmatrix} = \begin{bmatrix} G & X' & Y' \\ 0 & N & D \end{bmatrix}$$

Apply elementary row operations to

$$\begin{bmatrix} A & 1 & 0 \\ B & 0 & 1 \end{bmatrix}$$

so that

$$\begin{bmatrix} A & 1 & 0 \\ B & 0 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} G & X' & Y' \\ 0 & N & D \end{bmatrix} \tag{B.7}$$

The solution elements in (B.4)–(B.6) then immediately appear on the right hand side of (B.7).

For example, consider performing pole placement to a plant

$$a\frac{1+q^{-1}}{\left(1-q^{-1}\right)^2}$$

and the target closed loop has two poles at 0.8 and the remaining poles at the origin. We need

$$a\left(1+q^{-1}\right)X\left(q^{-1}\right) + \left(1-q^{-1}\right)^2 Y\left(q^{-1}\right) = \left(1-0.8q^{-1}\right)^2 \tag{B.8}$$

The greatest common factor of $a\left(1+q^{-1}\right)$ and $\left(1-q^{-1}\right)^2$ is 1. Forming

$$\begin{bmatrix} a\left(1+q^{-1}\right) & 1 & 0 \\ \left(1-q^{-1}\right)^2 & 0 & 1 \end{bmatrix}$$

we can perform the row operations

$$\begin{aligned}
\begin{bmatrix} a\left(1+q^{-1}\right) & 1 & 0 \\ \left(1-q^{-1}\right)^2 & 0 & 1 \end{bmatrix} &\rightarrow \begin{bmatrix} 1+q^{-1} & 1/a & 0 \\ 1-2q^{-1}+q^{-2} & 0 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 1+q^{-1} & 1/a & 0 \\ -4q^{-1} & -\frac{1}{a}\left(1+q^{-1}\right) & 1 \end{bmatrix} \\
&\rightarrow \begin{bmatrix} 1+q^{-1} & 1/a & 0 \\ -q^{-1} & -\frac{1}{4a}\left(1+q^{-1}\right) & \frac{1}{4} \end{bmatrix} \rightarrow \begin{bmatrix} 1 & \frac{1}{a}-\frac{1}{4a}\left(1+q^{-1}\right) & \frac{1}{4} \\ -q^{-1} & -\frac{1}{4a}\left(1+q^{-1}\right) & \frac{1}{4} \end{bmatrix} \\
&\rightarrow \begin{bmatrix} 1 & \frac{1}{a}-\frac{1}{4a}\left(1+q^{-1}\right) & \frac{1}{4} \\ 0 & -\frac{1}{4a}\left(1+q^{-1}\right)+q^{-1}\left[\frac{1}{a}-\frac{1}{4a}\left(1+q^{-1}\right)\right] & \frac{1}{4}\left(1+q^{-1}\right) \end{bmatrix} \\
&= \begin{bmatrix} 1 & \frac{1}{4a}\left(3-q^{-1}\right) & \frac{1}{4} \\ 0 & \frac{1}{4a}\left[-\left(1-q^{-1}\right)^2\right] & \frac{1}{4}\left(1+q^{-1}\right) \end{bmatrix}
\end{aligned}$$

Thus, after simplification, the solutions of (B.8) is

$$\begin{aligned}
X\left(q^{-1}\right) &= \frac{1}{4a}\left(3-q^{-1}\right)\left(1-0.8q^{-1}\right)^2 + Q\left(q^{-1}\right) \times \frac{1}{4a}\left[-\left(1-q^{-1}\right)^2\right] \\
Y\left(q^{-1}\right) &= \frac{1}{4}\left(1-0.8q^{-1}\right)^2 - Q\left(q^{-1}\right) \times \frac{1}{4}\left(1+q^{-1}\right)
\end{aligned}$$

where $Q\left(q^{-1}\right)$ is any polynomial of q^{-1}.

Matrix method

If the orders of the polynomials are fixed, then the polynomial equation can be solved by solving a matrix equality. Let q^{-1} be the polynomial variable and

$$\begin{aligned}
A\left(q^{-1}\right) &= a_0 + a_1 q^{-1} + a_2 q^{-2} + \cdots + a_n q^{-n} \\
B\left(q^{-1}\right) &= b_0 + b_1 q^{-1} + \ldots b_n q^{-n} \\
X\left(q^{-1}\right) &= x_0 + x_1 q^{-1} + \ldots x_{n-1} q^{-n+1} \\
Y\left(q^{-1}\right) &= y_0 + y_1 q^{-1} + \ldots y_{n-1} q^{-n+1} \\
F\left(q^{-1}\right) &= f_0 + f_1 q^{-1} + \ldots f_{2n-1} q^{-2n+1}
\end{aligned}$$

Matching the coefficients in $A(q^{-1})X(q^{-1}) + B(q^{-1})Y(q^{-1}) = F(q^{-1})$, we have

$$\underbrace{\begin{bmatrix}
a_0 & 0 & 0 & \ldots & 0 & b_0 & 0 & 0 & \ldots & 0 \\
a_1 & a_0 & 0 & \ldots & 0 & b_1 & b_0 & 0 & \ldots & 0 \\
a_2 & a_1 & a_0 & \ldots & 0 & b_2 & b_1 & b_0 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
a_n & a_{n-1} & a_{n-2} & \ldots & a_0 & b_n & b_{n-1} & b_{n-2} & \ldots & b_0 \\
0 & a_n & a_{n-1} & \ldots & a_1 & 0 & b_n & b_{n-1} & \ldots & b_1 \\
0 & 0 & a_n & \ldots & a_2 & 0 & 0 & b_n & \ldots & b_2 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & & a_n & 0 & 0 & 0 & \ldots & b_n
\end{bmatrix}}_{S}
\begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_{n-1} \\ y_0 \\ y_1 \\ \vdots \\ y_{n-1} \end{bmatrix}
=
\begin{bmatrix} f_0 \\ f_1 \\ \vdots \\ f_n \\ f_{n+1} \\ f_{n+2} \\ \vdots \\ f_{2n-1} \end{bmatrix}$$

The matrix S is called the Sylvester matrix and is non-singular if and only if $A(q^{-1})$ and $B(q^{-1})$ are coprime.