2010

IIR Digital Filter Design Using Convex Optimization

Aimin Jiang
University of Windsor

IIR Digital Filter Design Using Convex Optimization

by

Aimin Jiang

A Dissertation
Submitted to the Faculty of Graduate Studies
through the Department of Electrical and Computer Engineering
in Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy at the
University of Windsor

Windsor, Ontario, Canada
2009

IIR Digital Filter Design Using Convex Optimization

By

Aimin Jiang

APPROVED BY:

Dr. Andreas Antoniou, External Examiner
Department of Electrical and Computer Engineering
University of Victoria

Dr. Fazle Baki, Outside Department Reader
Odette School of Business

Dr. Jonathan Wu, First Department Reader
Department of Electrical and Computer Engineering

Dr. Huapeng Wu, Second Department Reader
Department of Electrical and Computer Engineering

Dr. Hon Keung Kwan, Advisor
Department of Electrical and Computer Engineering

Dr. Afsaneh Edrisy, Chair of Defense
Faculty of Graduate Studies

AUTHOR'S DECLARATION OF ORIGINALITY

I hereby certify that I am the sole author of this dissertation and that no part of this dissertation has been published or submitted for publication.

I certify that, to the best of my knowledge, my dissertation does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my dissertation, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in my dissertation and have included copies of such copyright clearances to my appendix.

I declare that this is a true copy of my dissertation, including any final revisions, as approved by my dissertation committee and the Graduate Studies office, and that this dissertation has not been submitted for a higher degree to any other University or Institution.

ABSTRACT

Digital filters play an important role in digital signal processing and communication. From the 1960s, a considerable number of design algorithms have been proposed for finite-duration impulse response (FIR) digital filters and infinite-duration impulse response (IIR) digital filters. Compared with FIR digital filters, IIR digital filters have better approximation capabilities under the same specifications. Nevertheless, due to the presence of the denominator in its rational transfer function, an IIR filter design problem cannot be easily formulated as an equivalent convex optimization problem. Furthermore, for stability, all the poles of an IIR digital filter must be constrained within a stability domain, which, however, is generally nonconvex. Therefore, in practical designs, optimal solutions cannot be definitely attained.

In this dissertation, we focus on IIR filter design problems under the weighted least-squares (WLS) and minimax criteria. Convex optimization will be utilized as the major mathematical tool to formulate and analyze such IIR filter design problems. Since the original IIR filter design problem is essentially nonconvex, some approximation and convex relaxation techniques have to be deployed to achieve convex formulations of such design problems. We first consider the stability issue. A sufficient and necessary stability condition is derived from the argument principle. Although the original stability condition is in a nonconvex form, it can be appropriately approximated by a quadratic constraint and readily combined with sequential WLS design procedures. Based on the sufficient and necessary stability condition, this approximate stability constraint can achieve an improved description of the nonconvex stability domain. We also address the nonconvexity issue of minimax design of IIR digital filters. Convex relaxation techniques are applied to obtain relaxed design problems, which are formulated, respectively, as second-order cone programming (SOCP) and semidefinite programming (SDP) problems. By solving these relaxed design problems, we can estimate lower bounds of minimum approximation errors, which are useful in subsequent design procedures to achieve real minimax solutions. Since the relaxed design problems are independent of local information, compared with many prevalent design methods which employ local

search, the proposed design methods using the convex relaxation techniques have an increased chance to obtain an optimal design.

Dedicated to my wife Yanping Zhu, my parents Yulin Jiang and Gaizhen Wu for their constant love and support

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere appreciation to my advisor Prof. Hon Keung Kwan for his thorough guidance, valuable advices, and generous support throughout my research work. I could not complete my research work as reported in this dissertation without his help.

I am grateful to Prof. Jonathan Wu, Prof. Huapeng Wu, and Prof. Fazle Baki for their valuable suggestions and comments on my research work. I would also like to thank Prof. Andreas Antoniou, University of Victoria, for providing me lots of valuable comments on my research work reported in this dissertation.

I would like to thank all my lab-mates in the ISPLab for their help during the last few years. A special thank to Ms. Swarna Bai Arniker for her kind discussion with me and her encouragement.

Last but not least, I would like to express my appreciation to my older sister Aiping Jiang and my brother-in-law Cheng Li for their constant love and generous help. A special thank to my handsome nephew, Yang Li, and my adorable niece, Helen Li. Whenever I get into difficult situations, I can always gain strength from their shining smiles.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

BIBO	Bounded Input and Bounded Output
DSP	Digital Signal Processing
FIR	Finite-duration Impulse Response
IDFT	Inverse Discrete Fourier Transform
IIR	Infinite-duration Impulse Response
LMI	Linear Matrix Inequality
LP	Linear Programming
PSD	Positive Semi-Definite
QP	Quadratic Programming
SDP	Semi-Definite Programming
SOC	Second-Order Cone
SOCP	Second-Order Cone Programming
WISE	Weighted Integral of the Squared Error
WLS	Weighted Least-Squares

CHAPTER I

INTRODUCTION

A digital filter is a computational tool to extract useful information and remove undesired components from input sequences, and simultaneously generate output sequences. Digital filters can be implemented on general-purpose computers or some specific hardware. Some advantages of digital filters over analog filters are listed below:

1. Digital filters are programmable, which means that the characteristics of digital filters can be easily modified leaving the hardware unchanged.
2. Digital filters can be conveniently designed, tested and implemented on general-purpose computers.
3. Compared with analog filters, the characteristics of digital filters are much more consistent with respect to time and temperature.
4. Digital filters are very versatile in their ability to process signals in a variety of ways, which includes the ability of some types of digital filters to adapt to the changes of input signals.

As one of important and fundamental areas in digital signal processing (DSP), the research work on digital filter designs started in the 1960s. Although many design methods have been proposed so far, nowadays the research on digital filter designs is still active. More efficient and robust design techniques are being proposed with the advances of DSP and mathematical theories. On the other hand, the emergence of new classes of digital filters also stimulates the development of digital filter designs.

In general, digital filters can be classified into two categories according to the duration of their impulse responses, *finite-duration impulse response* (FIR) and *infinite-duration impulse response* (IIR). Note that some people prefer an alternative terminology, in which an FIR digital filter is known as a *nonrecursive* digital filter, and an IIR digital filter is referred as a *recursive* digital filter.

The characteristics of a digital filter can be described by its transfer function. The transfer function of an FIR digital filter is a polynomial function of z^{-1}, *i.e.*,

$$F(z) = \sum_{l=0}^{L} f_l z^{-l} = \boldsymbol{f}^T \boldsymbol{\varphi}_L(z) \tag{1.1}$$

where

$$\boldsymbol{f} = [f_0 \quad f_1 \quad \dots \quad f_L]^T \tag{1.2}$$

$$\boldsymbol{\varphi}_l(z) = [1 \quad z^{-1} \quad \dots \quad z^{-l}]^T \tag{1.3}$$

Here, the superscript T represents the transpose of a vector or matrix. For an IIR digital filter, its transfer function is a rational function of z^{-1}, *i.e.*,

$$H(z) = \frac{P(z)}{Q(z)} = \frac{\sum_{n=0}^{N} p_n z^{-n}}{1 + \sum_{m=1}^{M} q_m z^{-m}} = \frac{\boldsymbol{p}^T \boldsymbol{\varphi}_N(z)}{\boldsymbol{q}^T \boldsymbol{\varphi}_M(z)} \tag{1.4}$$

where

$$\boldsymbol{p} = [p_0 \quad p_1 \quad \dots \quad p_N]^T \tag{1.5}$$

$$\boldsymbol{q} = [1 \quad q_1 \quad \dots \quad q_M]^T \tag{1.6}$$

The frequency responses of digital filters are calculated by evaluating their transfer functions on the unit circle, that is, $F\left(e^{j\omega}\right) = F(z)|_{z=e^{j\omega}}$ and $H\left(e^{j\omega}\right) = H(z)|_{z=e^{j\omega}}$. From (1.1), it can be found that all poles of an FIR digital filter are located on the origin of the z plane. However, all poles of an IIR digital filter must be constrained inside the unit circle of the z plane for stability.

In this dissertation, we mainly study IIR filter design problems. Generally speaking, an IIR filter design problem can be stated as follows:

> *Given a set of design specifications, e.g., filter orders, ideal frequency response and so forth, find an IIR digital filter with coefficients* $\boldsymbol{p}$ *and* $\boldsymbol{q}$*,*

whose frequency response can best approximate the given ideal frequency response under some design criterion.

In the proposed design methods, we assume that all the numerator and denominator coefficients are real values. Nevertheless, all the design methods presented in this dissertation can be readily extended to IIR filter designs with complex coefficients. It is noteworthy that besides the models in the direct form of (1.1) and (1.4), there are some other useful models, such as zero-pole, lattice, and state-space. However, in this dissertation, we only consider the direct form due to its simplicity in formulating design problems.

Because of the close relationship between FIR and IIR digital filters, in this chapter we shall first introduce FIR digital filter designs. Then, the history of IIR digital filter designs will be briefly reviewed. Motivations and objectives of the research work reported in this dissertation will be described later. The organization of the rest of the dissertation and main contributions will be finally presented in this chapter.

1.1 Introduction to FIR Digital Filter Design

Compared with IIR digital filters, FIR digital filters have several advantages:

1. Since all poles of an FIR digital filter are fixed at the origin of the z plane, the frequency response of an FIR digital filter is determined by its zeroes. Thereby, no stability concern exists for FIR digital filter designs.
2. By utilizing (anti-)symmetric structures, FIR digital filters with exactly linear phase over the whole frequency band can be easily achieved. However, except for some special cases, it is difficult to design an IIR digital filter, which has exactly linear phase over the whole frequency band.
3. Generally speaking, an FIR digital filter design can be equivalently formulated as a convex optimization problem in a finite-dimensional linear space. Accordingly, its globally optimal solution can be achieved using various optimization techniques. However, when magnitude and phase responses are both under consideration, in general, it is hard to transform an IIR filter design problem into an equivalent

convex optimization problem. Hence, globally optimal solutions cannot be definitely attained.

From the 1960s, a large part of efforts have been devoted to develop efficient approaches to design linear-phase FIR digital filters [1]-[2]. As mentioned above, linear-phase FIR digital filter coefficients demonstrate (anti-)symmetric structures. Thus, the number of free variables of design problems can be reduced by about one half. Furthermore, besides a constant-delay component, the frequency response of a linear-phase FIR digital filter can be expressed by a trigonometric function of filter coefficients.

The first well-known design technique is the *Fourier series method* [1]-[2], in which a desired frequency response is first expanded as its Fourier series and then truncated to a finite length. This method suffers from Gibbs' oscillations due to the discontinuity of the desired frequency responses. In order to reduce Gibbs' oscillations near the cutoff frequencies, a smooth time-limited window, such as the Hamming window and the Kaiser window, is multiplied with the coefficients of the Fourier series. This method has two obvious drawbacks: First of all, FIR digital filters designed by this window method are not optimal in any optimization sense. Moreover, the frequency band edges of the designed FIR filters cannot be the same as specified.

The second design technique is called the *frequency sampling method* [1]-[2]. The desired frequency response is specified on a set of discrete frequency points, and then the inverse discrete Fourier transform (IDFT) is used to obtain the discrete-time impulse response. Despite its easy implementation, the performance of this method is not good enough compared with the design methods using optimization techniques.

The use of *optimization methods* for designing FIR digital filters is most prevalent in recent years. The most well-known design method was proposed by Parks and McClellan [3], where a linear-phase FIR digital filter design is translated to a weighted minimax approximation problem. By virtue of the alternation theorem, there exists an optimal design with equiripple magnitude response for the weighted minimax design problem. Using the Remez exchange algorithm, the optimal design can be efficiently attained. In [4], a linear programming (LP) method was proposed as an alternative to

designing linear-phase FIR digital filters in the minimax sense. Some other linear constraints can be further incorporated in this LP design method.

In order to achieve the linear phase over the whole frequency band, linear-phase FIR filter coefficients should be (anti-)symmetric, and the group delay can only be set equal to $L/2$, where L denotes the filter order. If one wants to achieve a lower group delay, the filter length has to be correspondingly reduced. However, sometimes this is impracticable because of the strict design specifications. On the other hand, FIR digital filters with nonlinear phase responses are useful in many applications. Therefore, we are also interested in general FIR digital filter designs, where the ideal frequency responses can be arbitrarily selected.

It can be observed that the transfer function $F(z)$ in (1.1) is a linear function of filter coefficients $\boldsymbol{f}$. In general, an FIR filter design problem can be expressed as an equivalent convex optimization problem [5]. The techniques of transforming an FIR design problem into an equivalent convex optimization problem are very useful in the latter discussion of IIR digital filter designs. Let $D(\omega)$ represent the desired frequency response to be approximated. In the WLS sense, the approximation error can be defined by

$$\begin{aligned} E_{WLS}(\boldsymbol{f}) &= \int_{\Omega_I} W(\omega)\left|F\left(e^{j\omega}\right) - D(\omega)\right|^2 d\omega \\ &= \boldsymbol{f}^T\boldsymbol{A}\boldsymbol{f} - 2\boldsymbol{f}^T\boldsymbol{b} + \text{constant} \end{aligned} \tag{1.7}$$

where $W(\omega) \geq 0$ denotes a given weighting function, and Ω_I is the union of frequency bands of interest. In (1.7), the matrix $\boldsymbol{A}$ and vector $\boldsymbol{b}$ are defined as follows

$$\boldsymbol{A} = \int_{\Omega_I} W(\omega) \cdot \text{Re}\left\{\boldsymbol{\varphi}_L(e^{j\omega})\boldsymbol{\varphi}_L^H(e^{j\omega})\right\} d\omega \tag{1.8}$$

$$\boldsymbol{b} = \int_{\Omega_I} W(\omega) \cdot \text{Re}\left\{\boldsymbol{\varphi}_L(e^{j\omega})D^*(\omega)\right\} d\omega \tag{1.9}$$

In (1.8) and (1.9), Re{·} represents the real part of a complex value, and the superscripts H and $*$ denote, respectively, the conjugate transpose of a vector or matrix and the conjugate value of a complex number. Since the matrix $\boldsymbol{A}$ in (1.8) is symmetric and

positive definite, the WLS approximation error $E_{WLS}(\boldsymbol{f})$ is a convex quadratic function of $\boldsymbol{f}$. If no other constraints need to be incorporated in the WLS design problem, the optimal filter coefficients $\boldsymbol{f}_{opt}$ can be readily obtained by solving the linear equation $\boldsymbol{A}\boldsymbol{f}_{opt} = \boldsymbol{b}$. Some numerical methods, *e.g.*, Newton's method, can be utilized here to find $\boldsymbol{f}_{opt}$. If only linear constraints are incorporated, the design problem can be formulated as a quadratic programming (QP) problem. The approximation error $E_{WLS}(\boldsymbol{f})$ can also be expressed by

$$E_{WLS}(\boldsymbol{f}) = \left\|\boldsymbol{A}^{1/2}\boldsymbol{f} - \boldsymbol{A}^{-1/2}\boldsymbol{b}\right\|_2^2 + \text{constant} \tag{1.10}$$

where $\boldsymbol{A}^{1/2}$ denotes the square root of $\boldsymbol{A}$, and $\|\boldsymbol{x}\|_2$ represents the Euclidean norm of a vector $\boldsymbol{x}$. By introducing an auxiliary variable δ, the WLS design problem can be equivalently expressed by

$$\min \quad \delta \tag{1.11}$$

$$\text{s.t.} \quad \left\|\boldsymbol{A}^{1/2}\boldsymbol{f} - \boldsymbol{A}^{-1/2}\boldsymbol{b}\right\|_2 \le \delta \tag{1.11.a}$$

It is known that (1.11.a) is a second-order cone (SOC) constraint, and the above design problem is essentially an SOCP optimization problem. Some other linear or (convex) quadratic constraints can be further incorporated in (1.11).

In the minimax sense, the FIR filter design problem is defined by

$$\min_{\boldsymbol{f}} \quad \max_{\omega\in\Omega_I} \quad |E(\boldsymbol{f})| \tag{1.12}$$

where the (weighted) complex approximation error is defined by

$$E(\boldsymbol{f}) = W(\omega)\left[F\left(e^{j\omega}\right) - D(\omega)\right], \qquad \forall\omega \in \Omega_I \tag{1.13}$$

Even without any other constraint, the minimax design problem (1.12) does not have a closed-form solution. Thereby, we need to resort to numerical optimization methods to find the optimal designs. Fortunately, we can still transform (1.12) into an equivalent convex optimization problem. By introducing an auxiliary variable δ as the error limit of

$|E(\boldsymbol{f})|$ over Ω_I, the original minimax design problem (1.12) can be equivalently written by

$$\min \quad \delta \tag{1.14}$$

$$\text{s.t.} \quad |E(\boldsymbol{f})| \le \delta, \qquad \forall \omega \in \Omega_I \tag{1.14.a}$$

By reformulating $|E(\boldsymbol{f})|$, the constraint (1.14.a) can be transformed to the following SOC constraint

$$W(\omega)\left|F\left(e^{j\omega}\right) - D(\omega)\right| = \|\boldsymbol{G}(\omega)\boldsymbol{f} - \boldsymbol{g}(\omega)\|_2 \le \delta \tag{1.15}$$

where

$$\boldsymbol{G}(\omega) = W(\omega)\begin{bmatrix}\mathrm{Re}\left\{\boldsymbol{\varphi}_L^T(e^{j\omega})\right\} \\ \mathrm{Im}\left\{\boldsymbol{\varphi}_L^T(e^{j\omega})\right\}\end{bmatrix} \tag{1.16}$$

$$\boldsymbol{g}(\omega) = W(\omega)\begin{bmatrix}\mathrm{Re}\{D(\omega)\} \\ \mathrm{Im}\{D(\omega)\}\end{bmatrix} \tag{1.17}$$

In (1.16) and (1.17), Im{·} represents the imaginary part of a complex value. For simplicity, the constraint (1.15) can be enforced on a set of discrete frequency points densely sampled over Ω_I. Obviously, using the SOC constraint (1.15), the minimax design problem (1.14) can be converted to an SOCP problem.

As a generalization of the WLS and minimax criteria, the L_p-norm error criterion is also widely used in FIR filter designs as well. If $p \ge 1$, the corresponding FIR filter design problem is still convex in essence, although it may not be transformed to a convex optimization problem in some commonly used form, such as LP, QP, SOCP and SDP. In practical designs, some other linear and/or nonlinear constraints, for instance, magnitude and group delay flatness, peak error, and zero constraints, can be further incorporated in these design problems to improve the performances of the obtained FIR digital filters or make the design results satisfy some specific requirements.

1.2 Introduction to IIR Digital Filter Design

Compared with FIR digital filters, IIR digital filters can achieve much better performance under the same set of design specifications. However, IIR filter designs face more challenges due to the presence of the denominator $Q(z)$ in (1.4). The major difficulties we encounter are as follows:

1. Since the poles of an IIR digital filter can be anywhere in the z plane, in general, IIR filter design problems are nonconvex optimization problems. Accordingly, there exist many local optima on error performance surfaces, and globally optimal solutions cannot be definitely achieved or even verified.
2. If phase (or group delay) responses are also of concern, stability constraints must be incorporated in design procedures. However, when the denominator order M is larger than 2, the stability domain cannot be expressed as a convex set with respect to denominator coefficients $\boldsymbol{q}$.

The techniques of invariant impulse response, matched-z transformation, and bilinear transformation are widely used to achieve an IIR digital filter from a given analog filter [1]-[2]. These design techniques are straightforward, and can naturally guarantee the stability of obtained IIR digital filters. However, these techniques can only be applied to transform standard analog filters, such as lowpass, highpass, bandpass and bandstop filters, into digital counterparts.

Nowadays, IIR filter designs can be performed directly on the discrete time or frequency domain. If only the magnitude response is of concern, an IIR filter design problem can be simplified to some extent, since the stability can always be achieved by flipping the poles outside the unit circle into the inside without changing the magnitude response of the obtained IIR digital filter. So far, the minimax design for magnitude response approximation has been widely studied. One of most often used techniques is to approximate the squared ideal magnitude response by $H(z)H(z^{-1})$ [6]. This is mainly because in the form of squared magnitude, the design problem can be simplified to a quasi-convex optimization problem.

If phase (or group delay) responses are also under consideration, IIR filter design problems become more complicated. As in FIR filter design problems, the WLS and minimax criteria are also widely used in practical IIR filter designs. Like (1.7), the WLS approximation error of an IIR filter design can be defined by

$$\begin{aligned} E_{WLS}(\boldsymbol{x}) &= \int_{\Omega_I} W(\omega)\left|H\left(e^{j\omega}\right) - D(\omega)\right|^2 d\omega \\ &= \int_{\Omega_I} W(\omega)\left|\frac{P\left(e^{j\omega}\right)}{Q(e^{j\omega})} - D(\omega)\right|^2 d\omega \end{aligned} \tag{1.18}$$

where

$$\boldsymbol{x} = [\boldsymbol{q}^T \quad \boldsymbol{p}^T]^T \tag{1.19}$$

As in (1.7) and (1.13), $W(\omega)$ and $D(\omega)$ represent the given weighting function and the desired frequency response, respectively. Similarly, the minimax approximation error is expressed by

$$E_{MM}(\boldsymbol{x}) = \max_{\omega\in\Omega_I} |E(\omega)| \tag{1.20}$$

where the (weighted) complex approximation error is given by

$$\begin{aligned} E(\omega) &= W(\omega)\left[H\left(e^{j\omega}\right) - D(\omega)\right] \\ &= W(\omega)\left[\frac{P\left(e^{j\omega}\right)}{Q(e^{j\omega})} - D(\omega)\right] \end{aligned} \tag{1.21}$$

The objective of our design problems is to minimize these approximation errors subject to some other constraints. It is worth noting that although the complex approximation error $E(\omega)$ is differentiable over Ω_I, the minimax approximation error $E_{MM}(\boldsymbol{x})$ is a nondifferentiable function of $\boldsymbol{x}$. Therefore, it is inconvenient to directly manipulate $E_{MM}(\boldsymbol{x})$ in practical designs. Besides the WLS and minimax criteria, some other design criteria, such as the L_p-norm error criterion, where the approximation error is defined by $E_p(\boldsymbol{x}) = \int_{\Omega_I} W(\omega)\left|H\left(e^{j\omega}\right) - D(\omega)\right|^p d\omega$, are also adopted to formulate design problems.

In general, IIR filter design methods can be classified into two groups: direct and indirect ways. It should be mentioned here that direct design methods are often referred to as those methods that are carried out directly in the z domain and indirect design methods are generally considered to be those methods based on analog filters [2]. In this dissertation, however, we adopt somewhat different definitions for direct and indirect design methods. In the direct design strategy, the best approximation to a given ideal frequency response is found without any intermediate step. In the indirect design strategy, a design problem is first transformed to an FIR filter design problem. Then, model reduction techniques can be deployed to achieve an IIR digital filter, which can best approximate the FIR digital filter. As presented before, in general, FIR filter design problems can be equivalently cast as convex optimization problems and then efficiently solved. Therefore, the performances of indirect design methods are mainly determined by the second step, *i.e.*, FIR approximation by IIR digital filters. In this dissertation, we mainly study IIR filter designs using the direct design strategy. But it should be mentioned that the proposed design methods can be straightforwardly applied in indirect IIR filter designs by replacing the desired frequency response $D(\omega)$ by a well-defined FIR frequency response $F(e^{j\omega})$ and the frequency bands of interest Ω_I by the whole frequency band $[0, \pi]$.

As mentioned earlier, if the phase response is also under consideration, stability is an important issue to be addressed. On the other hand, the sensitivity of pole locations to coefficient quantization increases with decreasing distances of poles to the unit circle. The poles close to the unit circle may also cause considerable noise due to signal quantization. Thus, in practical designs, it is desirable to specify a maximum pole radius, which should be less than 1. Generally speaking, the stability issue can be overcome in two different ways: explicit and implicit descriptions. The explicit description of stability requirements, which is widely used in a variety of design methods, is to construct constraints or barrier functions on denominator coefficients to keep all poles inside the stability domain. Bounded input and bounded output (BIBO) is the classical definition of system stability. All the known stability constraints follow from this definition. Generally speaking, explicit stability constraints can be categorized into two groups, *i.e.*, time-domain stability constraints and frequency-domain stability constraints. Many time-

domain stability constraints try to control the l_2-norm of denominator coefficients $\boldsymbol{q}$ at a reasonable level or force the impulse responses $\hat{q}_m$ of the inverse filter $\hat{Q}(z) = 1/Q(z)$ to approach 0 as $m \to \infty$. The frequency-domain stability constraints are mainly derived from complex analysis. Compared with time-domain stability constraints, frequency-domain stability constraints are much more tractable. Many frequency-domain stability constraints are formulated in convex forms, such that they can be readily incorporated in optimization-based design methods. However, these convex frequency-domain stability constraints are only sufficient conditions for stability. This means that some stable IIR filters could be excluded from the set of admissible solutions. For the implicit description, the stability of designed IIR filters can be automatically guaranteed by design procedures. For example, by adjusting the step size at each iteration to keep all the updated poles staying inside the stability domain, some sequential design methods can always obtain stable designs without any explicit stability constraint.

1.3 Motivations and Objectives

This dissertation focuses on general IIR digital filter designs, in which the design requirements on magnitude and phase (or group delay) responses are both considered. In essence, IIR filter design problems are *nonconvex* optimization problems. Thereby, globally optimal solutions cannot be definitely attained, especially for those design methods in which local searches are utilized to gradually reduce approximation errors. On the other hand, even if a global design were obtained, it would be indeed difficult to confirm its optimality. In this dissertation, one of our major aims is to overcome the nonconvexity of design problems. We shall try to directly transform design problems into commonly used convex optimization models, such as SOCP and SDP. Convex relaxation techniques are to be introduced to achieve this goal. Since the feasible sets of the relaxed design problems are essentially larger than the ones of the original design problems, the global optima cannot be excluded from the convex formulations of these design problems. In the subsequent design procedures, we can gradually screen out unqualified solutions to approach the optimal designs. When a design problem is cast as a convex optimization problem, it can be solved reliably and efficiently using numerical algorithms

developed for convex optimization. Actually, many well-developed mathematical tools are available for solving these convex optimization problems.

So far, a large number of IIR filter design methods have been proposed. Although the effectiveness of these methods has been demonstrated by many examples in the literature, their design performances could be impaired by insufficient stability constraints adopted by these design methods, or their practical applications could be restricted by the unguaranteed convergence of these design methods. These issues will also be addressed in this dissertation.

Although it is difficult to completely resolve the nonconvexity and stability issues of IIR filter design problems, in this dissertation we shall try to alleviate these difficulties to some extent, such that the proposed design methods have more chances to approach optimal designs than traditional design methods.

1.4 Organization of the Dissertation

The rest of this dissertation is organized as follows: In Chapter II, some important IIR digital filter design methods will be briefly reviewed. Their advantages and disadvantages will be discussed. In Chapter III, a sufficient and necessary stability condition is to be derived from the argument principle of complex analysis, which can be combined with a sequential SOCP design method proposed in the WLS sense. In Chapter IV, another sequential SOCP design method is to be developed but in the minimax sense. Relaxation technique is to be introduced in this design method to achieve a relaxed design problem in convex form. A real minimax solution can be further attained by a sequential procedure based on the relaxed design problem. In Chapter V, a novel design method using SDP relaxation technique will be presented in the minimax sense. As in Chapter IV, convex relaxation technique will be utilized to formulate a relaxed SDP feasibility problem, which will be solved sequentially in a bisection search procedure. To achieve a real minimax design, an inner bisection search procedure is to be further introduced. The stability of designed IIR filters can also be guaranteed by the inner bisection search procedure. Conclusions and suggestions for future study will be presented in Chapter VI.

1.5 Main Contributions

In this dissertation, we are mainly studying IIR filter design problems under the WLS and minimax criteria. All the proposed design methods are primarily devoted to tackle the nonconvexity and stability issues of design problems. The main contributions of the research work reported in this dissertation are summarized as follows:

Firstly, a novel stability condition is derived from the argument principle of complex analysis. Compared with some other frequency-domain stability conditions, it is both sufficient and necessary. In practice, however, this stability condition is still nonconvex. Thereby, some approximation techniques need to be employed to achieve an approximate stability condition in a quadratic form, such that it can be readily combined with the sequential WLS design procedure. This approximate stability condition can guarantee the stability of designed IIR digital filters, if the sequential design method is convergent and a regularization parameter is appropriately selected.

Secondly, convex relaxation techniques are introduced in minimax IIR filter designs. The major idea of this design strategy is to relax the original nonconvex design problems so as to achieve design problems in convex forms, which can be efficiently and reliably solved. Furthermore, by solving these relaxed design problems, we can obtain some important information about optimal solutions of the original nonconvex design problems, *e.g.*, lower and upper bounds of the minimum approximation error. In this dissertation, two different types of convex relaxation techniques are used in minimax designs. The resulting relaxed design problems are formulated, respectively, as SOCP and SDP optimization problems. In the SDP formulation, a sufficient condition for an optimal design of the original design problem is presented, which can be used to detect the optimality of IIR filters designed by the proposed design method.

Finally, in conjunction with convex relaxation techniques, novel sequential design methods are presented for minimax designs. Since generally we cannot achieve real minimax designs by only solving the relaxed design problems, these sequential procedures are proposed to gradually reduce the discrepancy between the original and

relaxed design problems. Due to the essential nonconvexity of IIR filter design problems, some approximation techniques have to be further employed to achieve this goal.

CHAPTER II

REVIEW OF IIR DIGITAL FILTER DESIGN METHODS

Compared with an FIR filter design problem, an IIR filter design problem is more challenging due to its nonconvex nature. As mentioned before, the nonconvexity is mainly incurred by the denominator $Q(z)$ whose roots can be anywhere in the z plane. Recently, a number of design methods [7]-[43] have been proposed to solve various IIR filter design problems. These methods can be roughly classified into three groups: sequential design methods [7]-[27], nonsequential design methods [28]-[32], and model reduction methods [33]-[43]. We shall briefly review some important design methods in this chapter. It is worth emphasizing that this classification is not unique, since strictly some methods can be classified into two groups. For example, some model reduction methods also involve sequential procedures. We group these methods based on their basic design strategies. Another point, which should be mentioned here, is that many design methods depend on a variety of optimization methods [44]-[48] (*e.g.*, quasi-Newton methods, sequential quadratic programming method, simplex method, and interior-point methods) to solve these design problems. Essentially speaking, these optimization methods involve iterations. Nevertheless, in this dissertation, we shall focus on convex formulations and analyses of IIR filter design problems. Thereby, these optimization methods can be viewed as black-box subroutines, which can be invoked to solve practical problems formulated by designers. These optimization methods have been provided by many well-developed software.

2.1 Sequential Design Methods

The most prevalent design strategy is to employ sequential procedures [7]-[27] to gradually approach optimal solutions. At each iteration, original design problems are reformulated through some approximation techniques. These approximate design problems can then be more efficiently solved than the original design problems.

The Steiglitz-McBride (SM) scheme [49] is adopted in many sequential design methods [7]-[13] under various design criteria. At each iteration, the denominator of an approximation error is replaced by its counterpart obtained at the previous iteration and combined with a prescribed weighting function. Then, the original objective functions can be approximated by convex functions of filter coefficients. Accordingly, the IIR filter design problems can be transformed to convex optimization problems. Different stability constraints are utilized in these design methods, such as the positive realness [7]-[8], [10]-[11], the Lyapunov theory [12], and the argument principle [13] based stability constraints. Although the SM scheme does not completely tackle the nonconvexity of IIR filter design problems, compared with classical descent techniques, it can avoid being stuck at local minima near the initial points. Its effectiveness has been demonstrated by many examples reported in the literature. The major drawback of the SM design approaches is that the convergence of these sequential methods cannot be definitely guaranteed.

A design strategy similar to the SM scheme is used by the design method proposed in [14]. By introducing an inverse filter $\hat{Q}(z)$ corresponding to the denominator $Q(z)$, *i.e.*, $\hat{Q}(z)Q(z) = 1$, numerator and denominator designs can be decoupled into two separate optimization problems. The optimal numerators can be explicitly expressed in terms of coefficients of the inverse filter. The denominator design can be simplified as a QP problem by adopting an approximation technique similar to the SM scheme. The stability of designed filters can be ensured by flipping the poles outside the unit circle into the inside at each iteration. A variant of the design method [14] has been presented in the time domain by [15]. Instead of the approximation error $E_{WLS}(\boldsymbol{x})$ defined by (1.18), the design objective is to minimize the model-fitting error between the desired impulse responses and significant samples of an IIR digital filter system, *i.e.*, $\|\boldsymbol{h} - \boldsymbol{h}_d\|_2^2$ where $\boldsymbol{h} = [h(0)\ h(1)\ \ldots\ h(L)]^T$ denotes the impulse responses of $H(z)$ and $\boldsymbol{h}_d = [h_d(0)\ h_d(1)\ \ldots\ h_d(L)]^T$ represents the desired impulse responses.

Another design method employing the reweighting technique has been proposed by [16], in which a minimax design can be achieved by taking advantage of WLS designs. At each iteration, a new weighting function is determined by the magnitude envelope of

the complex approximation error of the IIR filter obtained at the previous iteration. Then, by solving a WLS design problem constructed by the new weighting function, the minimax error can be simultaneously reduced. The major drawback of this design method is that stability constraints cannot be directly incorporated into the design procedure. Thus, the resulting filters may be unstable. A similar strategy is also used by the minimax design method proposed in [17]. However, the magnitude of the complex approximation error of the IIR filter obtained at the previous iteration is directly employed to determine the weighting function.

Since the frequency response $H(e^{j\omega})$ is a nonlinear function of denominator coefficients, many design methods use its Taylor series to simplify design problems. Based on this idea, a minimax design method has been developed by [18]. At each iteration, given a denominator the optimal numerator design is first obtained. By fixing the numerator, $H(e^{j\omega})$ is then approximated by its first-order Taylor series with respect to denominator coefficients, *i.e.*, $H^{(k+1)}\left(e^{j\omega}\right) \approx H^{(k)}\left(e^{j\omega}\right) + \Delta\boldsymbol{q}^{(k+1)T}\nabla_{\boldsymbol{q}}H^{(k)}\left(e^{j\omega}\right)$, where k denotes the iteration index and $\Delta\boldsymbol{q}^{(k+1)}$ represents a descent direction of denominator coefficients to be determined. Using this linearized frequency response, the design problem at each iteration can be formulated as a convex optimization problem. Line search is employed to guarantee the convergence of this sequential design method. Provided the initial design is stable, the stability of a designed IIR filter can be guaranteed by adjusting the step size α at each iteration, such that the updated denominator coefficients $\boldsymbol{q}^{(k+1)} = \boldsymbol{q}^{(k)} + \alpha\Delta\boldsymbol{q}^{(k+1)}$ is always within the stability domain. Generally, the computational complexity of this design method is relatively low. However, since the descent direction is determined based on the local information, the design performance is sensitive to the selection of initial points.

Taylor series approximation is also utilized by the SOCP method [19] under the minimax criterion and the Gauss-Newton (GN) method [20] under the WLS criterion. Instead of separating the numerator and denominator designs, these two design methods approximate $H(e^{j\omega})$ by its first-order Taylor series with respect to both numerator and denominator coefficients. In [19], while the numerator still adopts the direct form as in (1.4), the denominator polynomial is factorized as a product of second-order sections and

a first-order section if the denominator order M is odd. Then, the resulting stability constraints can be expressed by a set of linear inequality constraints in terms of these factorized denominator coefficients. The advantage of using the factorized denominator is that the corresponding stability constraints can be easily expressed by a set of linear inequality constraints, which are sufficient and almost necessary for stability. Different from the SOCP method [19], the GN design method [20] adopts numerator and denominator polynomials both in the direct form. The Rouché's theorem based stability constraint is used in the GN design method, which is less restrictive than the positive realness based stability constraint [32]. Both the SOCP and GN design methods suffer the same drawback as SM design methods regarding nonguaranteed convergence. Another design method using a similar design strategy has been proposed by [21]. A linearized argument principle based stability constraint is employed to guarantee the stability of designed IIR filters.

By adopting linearized frequency responses, the approximation errors in [19]-[21] can generally be written as convex quadratic forms, *i.e.*, $\frac{1}{2}\Delta\boldsymbol{x}^T\boldsymbol{G}\Delta\boldsymbol{x} + \Delta\boldsymbol{x}^T\boldsymbol{g}$, where $\Delta\boldsymbol{x}$ denotes a descent direction of filter coefficients $\boldsymbol{x}$, $\boldsymbol{g}$ represents the gradient of the original approximation errors with respect to $\boldsymbol{x}$, and $\boldsymbol{G}$ is a positive definite matrix generally determined by the gradient. The matrix $\boldsymbol{G}$ can be viewed as an estimate of the Hessian of the original approximation errors. The real Hessian of the approximation error is utilized by the design method proposed in [22] under the L_p-norm error criterion. The modified Newton's method is employed to solve the design problem. The stability of designed IIR filters can be ensured by a similar strategy adopted in [18].

A multistage design method has been proposed by [23]. The SM [11], GN [20], and classical descent methods (*e.g.*, BFGS and Newton's method) are successively applied to achieve a better design in the WLS sense. A linear matrix inequality (LMI) stability constraint in terms of positive realness has been developed in [23]. It can be proved [23] that the stability domain defined by the Rouché's theorem based stability constraint [20] is contained in the one given by the LMI stability constraint. In order to incorporate this LMI stability constraint, all the design problems in [11] and [20] should be reformulated as equivalent SDP optimization problems. Starting from the WLS design obtained from

the multistage design method [23], a minimax design [24] can be obtained by successively optimizing numerators using the reweighting technique proposed by [16].

A special class of sequential design methods have been developed by [25]-[26] based on a sufficient condition for the optimal rational approximation, which states that the approximation error has a specific number of extremal points over the frequency bands of interest. The Remez exchange algorithm is employed to identify these extremal points. In order to achieve satisfactory designs, the initial point should be selected close enough to the optimal solution to guarantee the convergence of the sequential procedure. The Remez exchange algorithm is also employed by the minimax design method proposed by [27]. However, the transfer function of an IIR filter in [27] is in the form of a parallel connection of two allpass filters.

2.2 Nonsequential Design Methods

In practice, optimal designs cannot be definitely achieved even using the sequential design methods described earlier. In practice, if an obtained solution satisfies the prescribed specifications, it can be taken as a successful design. On the other hand, as mentioned before, the convergence of some sequential design methods cannot be always ensured. Therefore, some design methods [28]-[30] abandon the sequential design strategy and try to strictly formulate design problems as unconstrained optimization problems, which are then solved by a variety of efficient and robust unconstrained optimization methods. In [28]-[29], the objective functions of the WLS design problems consist of two components. The first part reflects the WLS approximation error, while the second one serves as a barrier function to control poles' positions for stability. Gradient-based optimization methods can be applied to solve these unconstrained optimization problems. In general, designers should provide at least the gradients of the objective functions. Satisfactory designs can be obtained by repeating the design procedures from different initial points.

In [30], the IIR filter design problem is formulated as a nonlinear optimization problem, whose objective function is expressed as a weighted sum of magnitude and group delay approximation errors. Instead of the direct form, the transfer function in [30]

is decomposed as a cascade of second-order sections. The Fletcher-Powell algorithm [50] is employed in [30] to solve this nonlinear design problem. The stability of designed filters can be ensured by the same technique used in [18].

In [31], the design problem is first formulated as a multiple-criterion optimization problem, in which both magnitude and group delay approximation errors are simultaneously minimized. This multiple-criterion design problem can be further transformed to a constrained nonlinear programming problem and then solved by sequential quadratic programming method. In [30] and [31], the design problems are both formulated under the L_p-norm error criterion.

An LP design method has been proposed by [32] under the minimax criterion. In order to simplify the design problem, the denominator of the complex approximation error $E(\omega)$ defined by (1.21) is neglected, such that the peak error constraint $|E(\omega)| \leq \delta$ is transformed to a quadratic form, which can be further approximated by a set of linear inequality constraints. The stability of designed IIR digital filters can be assured by a positive realness based constraint. Despite its simplicity, it is hard to obtain a true minimax design by this method. However, in practice, we can use this method at the beginning of some sequential design methods to achieve initial designs [18].

2.3 Model Reduction Design Methods

Sequential and nonsequential design methods described above both belong to the category of direct design methods, that is, given a desired frequency response, we can directly obtain an IIR digital filter using these design methods. Another category of methods [33]-[43] design IIR digital filters through an indirect way. An FIR digital filter satisfying prescribed specifications are designed first, and then model reduction techniques are applied to approximate the FIR digital filter by a reduced-order IIR digital filter. Specifically, for the WLS and minimax designs, the desired frequency response $D(\omega)$ in (1.18) and (1.21) is replaced by the frequency response $F\left(e^{j\omega}\right)$ of an FIR digital filter, which is designed first to approximate the ideal frequency response $D(\omega)$ by any existing FIR design method.

The indirect design scheme has two advantages:

1. Since an FIR filter design problem can be conveniently formulated as a convex optimization problem in a finite-dimensional space, which has been extensively studied, the second step becomes the kernel of an IIR filter design problem. By contrast with direct IIR digital filter design methods, the FIR approximation by IIR digital filters is less complicated.
2. In most of indirect design methods, the FIR approximation by IIR digital filters can substantially guarantee the stability of designed IIR digital filters, which also facilitates the design procedures.

However, even though the optimal results can be obtained in each step of indirect design methods, it cannot be concluded that the optimal solutions of the original IIR filter design problems can be definitely attained by indirect design methods.

2.4 Filter Designs Using Convex Optimization

The mathematics of convex optimization [51]-[55] has been studied for about one century. However, new research interests in this topic have been rejuvenated due to the advances of interior-point methods developed in the 1980s. Recently, many applications of convex optimization have been discovered in various fields of applied science and engineering, such as automatic control system, signal processing, VLSI circuit design, mechanical structure design, statistics and probability, and finance. There are many advantages of utilizing convex optimization to solve practical engineering problems. The most important one is that when a problem is equivalently cast as a convex optimization problem, any local solution is also a *global* optimum. Furthermore, a convex optimization problem can be solved very efficiently and reliably, using interior-point methods [70]-[71].

Recently, convex optimization has been applied to FIR [4]-[5], [56]-[61], allpass [62]-[63], and IIR [6]-[8], [10]-[13], [19], [21], [23]-[24], [32], [35] digital filter designs. It has been shown in Chapter I that given a desired frequency response, the WLS and minimax FIR filter design problems can be cast as equivalent convex optimization

problems. Thus, the optimal designs can be definitely obtained. Compared with FIR filter designs, allpass filter designs face more challenges due to the same difficulties as encountered in IIR filter designs. An important property which can be exploited is the mirror symmetric relation between numerator and denominator, *i.e.*, $P(z) = z^{-M}Q(z^{-1})$. Note that if the transfer function of an allpass filter is still defined by (1.4) with $N = M$, this property can be described by a set of linear equality constraints $p_{M-m} = q_m$ for $m = 0, 1, ..., M$. Therefore, most of optimization-based IIR filter design methods described in the proceeding sections can also be used to design allpass filters. However, this design strategy does not make full use of the characteristics of allpass digital filters, and hence some computation resources will be wasted. Since allpass filters have the fullband unity magnitude responses, the design problems can also be formulated in terms of phase response approximation error. Let $\theta_d(\omega)$ and $\phi_Q(\omega)$ denote, respectively, the ideal phase response to be approximated and the phase response of the denominator $Q(z)$. Then, the phase response approximation error $\theta_e(\omega)$ can be calculated by $\theta_e(\omega) = -M\omega - \theta_d(\omega) - 2\phi_Q(\omega)$. Since the tangent function is an increasing function within $[-\pi/2, \pi/2]$, we can reduce the phase response approximation error by minimizing the error limit of $\tan\frac{\theta_e(\omega)}{2}$ over Ω_I, where $\tan\frac{\theta_e(\omega)}{2} = \frac{\sum_{m=0}^{M} q_m \sin\phi_m(\omega)}{\sum_{m=0}^{M} q_m \cos\phi_m(\omega)}$ and $\phi_m(\omega) = m\omega - \frac{M\omega+\theta_d(\omega)}{2}$. It can be seen that the approximation error is a linear fractional function of denominator coefficients. Accordingly, allpass filter design problems can be transformed into quasi-convex optimization problems.

As discussed in the previous sections, convex optimization has been widely used to solve IIR filter design problems, especially in a variety of sequential design methods. Since IIR filter designs are essentially nonconvex optimization problems, generally it is impossible or computationally costly to achieve optimal designs. Furthermore, even if an optimal design were given, it would be hard to confirm that it was indeed the global optimum. However, this difficulty can be alleviated to some extent, under the framework of convex optimization. For example, convex relaxation techniques can be applied to transform the original nonconvex design problems into convex forms. Then, lower bounds of optimal values of the original design problems can be obtained. These lower bounds provide us some important information regarding the globally optimal designs.

CHAPTER III

IIR DIGITAL FILTER DESIGN WITH NEW STABILITY CONSTRAINT BASED ON ARGUMENT PRINCIPLE

Stability is a critical concern in an IIR filter design problem. So far, many stability constraints have been proposed in frequency domain. However, some of these stability constraints are only sufficient conditions, which means stable filters could be excluded from the feasible sets of design problems. Recently, a stability constraint based on the argument principle of complex analysis has been developed in [21], which is both sufficient and necessary. By truncating the higher-order Taylor series components, the resulting stability constraint becomes a linear equality constraint. However, through a large number of simulations, it is found that this linearized constraint could be invalid in some situations. As an attempt to resolve this problem, a new stability constraint is proposed in this chapter, which is also based on the argument principle. Unlike the linearized stability constraint in [21], this new stability constraint is approximated in a quadratic form. The effectiveness of this approximate stability constraint can be demonstrated by theoretical analysis and many simulation examples.

This chapter is organized as follows. In Section 3.1, a sequential SOCP method without any stability constraint is first introduced to design IIR digital filters in the WLS sense. Then, peak error constraints are incorporated as SOC constraints. In Section 3.2, a novel stability constraint is developed from the argument principle of complex analysis, which is then combined with the sequential design method. Design examples are presented in Section 3.3 to illustrate the performance of the proposed method.

3.1 WLS Design of IIR Digital Filters

3.1.1 Sequential Design Procedure

In the WLS sense, the IIR filter design problem can be expressed by

$$\min_{\boldsymbol{x}=[\boldsymbol{q}^T\ \boldsymbol{p}^T]^T} E_{WLS}(\boldsymbol{x}) \tag{3.1}$$

where the approximation error $E_{WLS}(\boldsymbol{x})$ has been defined by (1.18). By introducing an auxiliary variable δ, (3.1) can be reformulated as

$$\min \quad \delta \tag{3.2}$$

$$\text{s.t.} \quad \int_{\Omega_I} W(\omega)\left|\frac{P(e^{j\omega})}{Q(e^{j\omega})} - D(\omega)\right|^2 d\omega \le \delta^2 \tag{3.2.a}$$

Because of the existence of denominator $Q(e^{j\omega})$ in the integrand, the constraint (3.2.a) cannot be cast as a convex form. Here, we employ the Steiglitz-McBride scheme [49] to simplify the above design problem. This strategy has been widely used by many design methods [7]-[13]. At the kth iteration, the constraint (3.2.a) is modified as

$$\begin{aligned} &\int_{\Omega_I} W^{(k-1)}(\omega)\left|P^{(k)}\left(e^{j\omega}\right) - D(\omega)Q^{(k)}\left(e^{j\omega}\right)\right|^2 d\omega \\ &= \int_{\Omega_I} W^{(k-1)}(\omega)\left|\boldsymbol{c}^T(\omega)\boldsymbol{x}^{(k)}\right|^2 d\omega \\ &\le \delta^2 \end{aligned} \tag{3.3}$$

where $\boldsymbol{x}^{(k)}$ denotes the current filter coefficients to be determined, and the vector $\boldsymbol{c}(\omega)$ is defined by

$$\boldsymbol{c}(\omega) = \begin{bmatrix} -D(\omega)\boldsymbol{\varphi}_M(e^{j\omega}) \\ \boldsymbol{\varphi}_N(e^{j\omega}) \end{bmatrix} \tag{3.4}$$

The major modification is on the weighting function, *i.e.*, $W^{(k-1)}(\omega)$, which is defined by

$$W^{(k-1)}(\omega) = \frac{W(\omega)}{|Q^{(k-1)}(e^{j\omega})|^2} \tag{3.5}$$

Here, the denominator obtained at the previous iteration is taken into (3.5) to construct a new weighting function. Obviously, the left hand side of the inequality (3.3) is in a convex quadratic form with respect to $\boldsymbol{x}^{(k)}$, which can be expressed by

$$\boldsymbol{x}^{(k)T}\boldsymbol{A}^{(k-1)}\boldsymbol{x}^{(k)} \leq \delta^2 \tag{3.6}$$

where

$$\boldsymbol{A}^{(k-1)} = \int_{\Omega_I} W^{(k-1)}(\omega) \cdot \mathrm{Re}\{\boldsymbol{c}(\omega)\boldsymbol{c}^H(\omega)\}d\omega \tag{3.7}$$

Since $\boldsymbol{A}^{(k-1)}$ is a symmetric and positive definite matrix, (3.6) can be further cast into an SOC constraint

$$\left\| \left[\boldsymbol{A}^{(k-1)}\right]^{1/2} \boldsymbol{x}^{(k)} \right\|_2 \leq \delta \tag{3.8}$$

where $\boldsymbol{A}^{1/2}$ denotes the square root of the matrix $\boldsymbol{A}$.

In practice, for the sake of robustness of the sequential design procedure, the filter coefficients are updated by

$$\boldsymbol{x}^{(k)} = \boldsymbol{x}^{(k-1)} + \gamma\boldsymbol{\eta}^{(k)}, \qquad 0 < \gamma < 1 \tag{3.9}$$

where $\boldsymbol{x}^{(k-1)}$ is the coefficient vector obtained at the previous iteration, γ is a fixed step size, and $\boldsymbol{\eta}^{(k)}$ is the updating vector at the current iteration. By specifying $x_0^{(k)} = 1$ or equivalently $\eta_0^{(k)} = 0$ for all $k \geq 0$, the design problem (3.2) with the SOC constraint (3.8) can be rewritten by

$$\min \quad \delta \tag{3.10}$$

$$\text{s.t.} \quad \eta_0^{(k)} = 0 \tag{3.10.a}$$

$$\left\| \boldsymbol{F}^{(k-1)}\boldsymbol{\eta}^{(k)} + \boldsymbol{g}^{(k-1)} \right\|_2 \leq \delta \tag{3.10.b}$$

where

$$\boldsymbol{F}^{(k-1)} = \left[\boldsymbol{A}^{(k-1)}\right]^{1/2} \tag{3.11}$$

$$\boldsymbol{g}^{(k-1)} = \left[\boldsymbol{A}^{(k-1)}\right]^{1/2} \boldsymbol{x}^{(k-1)} \tag{3.12}$$

The sequential design procedure continues until the following condition is satisfied

$$\left\|\boldsymbol{\eta}^{(k)}\right\|_2 \leq \varepsilon \tag{3.13}$$

where ε is a prescribed convergence tolerance, or k exceeds a specified maximum number of iterations. Although so far the convergence of the sequential procedure has not been definitely guaranteed, the effectiveness of the SM scheme has been demonstrated by many filter examples in a variety of papers.

3.1.2 Peak Error Constraint

In [7] and [8], linearized peak error constraints have been developed to control the peak errors. Here, we shall reformulate the peak error constraints as a set of SOC constraints, which can better approximate the true peak error constraints.

The peak error constraints can be strictly expressed by

$$\left|\frac{P(e^{j\omega_i})}{Q(e^{j\omega_i})} - D(\omega_i)\right| \leq \mu(\omega_i), \qquad \omega_i \in \Omega_I, \quad i = 1,2,\dots,K \tag{3.14}$$

where $\mu(\omega)$ denotes the prescribed peak error limit at a specific frequency ω. Like the difficulty encountered in formulating the design problem (3.2), the real peak error constraint also has the denominator on the left hand side of (3.14). Adopting a similar technique employed in (3.3) and rearranging terms, we obtain

$$\begin{aligned} &\left|P^{(k)}\left(e^{j\omega_i}\right) - Q^{(k)}\left(e^{j\omega_i}\right)D(\omega_i)\right| \\ &= \left\|\boldsymbol{B}(\omega_i)\boldsymbol{\eta}^{(k)} + \boldsymbol{B}(\omega_i)\boldsymbol{x}^{(k-1)}\right\|_2 \\ &\leq \mu(\omega_i) \cdot \left|Q^{(k-1)}\left(e^{j\omega_i}\right)\right|, \quad \omega_i \in \Omega_I, \quad i = 1,2,\dots,K \end{aligned} \tag{3.15}$$

where

$$\boldsymbol{B}(\omega) = [\mathrm{Re}\{\boldsymbol{c}(\omega)\boldsymbol{c}^H(\omega)\}]^{1/2} \tag{3.16}$$

Note that in [7] and [8] the IIR filter design problems are cast, respectively, into LP and QP problems, in which only linear constraints can be handled. Therefore, the approximation of a circle by a regular polygon is applied to linearize the constraint (3.14). Although this approximation is applicable when the edge number of a regular polygon is large enough, the total number of peak error constraints is rapidly increased.

3.2 Argument Principle Based Stability Constraint

A new stability constraint based on the argument principle is to be developed in this section. First of all, the argument principle is to be reviewed. The stability constraint derived from the argument principle is then to be approximated by a quadratic constraint and combined with the sequential design method described in Section 3.1.

3.2.1 Argument Principle

If $f(z)$ is analytic in a region R enclosed by a contour C in the z plane except at a finite number of poles, let N_z be the number of zeros and N_p be the number of poles of the function $f(z)$ in R, where each zero and pole is counted according to its multiplicity. Then we have

$$N_z - N_p = \frac{1}{2\pi j}\oint_C \frac{f'(z)}{f(z)}dz \tag{3.17}$$

This result is called the argument principle [64]-[65].

In order to develop a practical stability constraint for IIR digital filter designs, we consider the following monic polynomial function

$$\begin{aligned} f(z) &= z^M Q(z) \\ &= \sum_{m=0}^{M} q_m z^{M-m}, \ \ q_0 = 1 \end{aligned} \tag{3.18}$$

Obviously, $f(z)$ has M zeros and no poles in the finite z plane. The contour C is chosen as an origin-centered circle with a prescribed maximum pole radius r, *i.e.*, $C = \{z: |z| = r,\ r < 1\}$. Then, according to the argument principle described above, all zeros of $f(z)$ lie strictly in the region R enclosed by C, if and only if the following equality condition is satisfied

$$M = \frac{1}{2\pi j}\oint_C \frac{f'(z)}{f(z)}dz \tag{3.19}$$

The integral in (3.19) is carried out counterclockwise along C. Note that

$$\begin{aligned}\oint_C \frac{f'(z)}{f(z)}dz &= \oint_C d\ln f(z) \\ &= \oint_C d\ln|f(z)| + j\oint_C d\arg f(z)\end{aligned} \tag{3.20}$$

where $\arg f(z)$ denotes the argument of $f(z)$. The first term on the right-hand side of the second equation of (3.20) is always equal to zero, since the logarithmic function is single-valued and C is closed. According to (3.18), $\arg f(z)$ can be expanded as $M\omega + \arg Q(z)$ on C, and then the stability constraint (3.19) can be simplified as

$$\frac{1}{2\pi}\oint_C d\arg Q(z) = 0 \tag{3.21}$$

Thus, the stability constraint (3.21) of an IIR digital filter is stated as: An IIR digital filter with the denominator $Q(z)$ is stable, if and only if the total change in the argument of $Q(z)$ is equal to 0, when the integral is carried out along C counterclockwise.

3.2.2 Argument Principle Based Stability Constraint

The polynomial function $Q(z)$ can be expressed as

$$Q(z)|_{z=re^{j\omega}} = Q_R\left(re^{j\omega}\right) + jQ_I\left(re^{j\omega}\right) \tag{3.22}$$

where

$$Q_R\left(re^{j\omega}\right) = \mathrm{Re}\left\{Q\left(re^{j\omega}\right)\right\} = \mathrm{Re}\left\{\boldsymbol{\varphi}_M^T\left(re^{j\omega}\right)\right\}\boldsymbol{q} \tag{3.23}$$

$$Q_I\left(re^{j\omega}\right) = \mathrm{Im}\left\{Q\left(re^{j\omega}\right)\right\} = \mathrm{Im}\left\{\boldsymbol{\varphi}_M^T\left(re^{j\omega}\right)\right\}\boldsymbol{q} \tag{3.24}$$

The argument of $Q\left(re^{j\omega}\right)$ is then computed by

$$\begin{aligned} \arg Q\left(re^{j\omega}\right) &= \arctan\frac{Q_I\left(re^{j\omega}\right)}{Q_R\left(re^{j\omega}\right)} \\ &= \arctan\frac{\mathrm{Im}\left\{\boldsymbol{\varphi}_M^T\left(re^{j\omega}\right)\right\}\boldsymbol{q}}{\mathrm{Re}\left\{\boldsymbol{\varphi}_M^T\left(re^{j\omega}\right)\right\}\boldsymbol{q}} \end{aligned} \tag{3.25}$$

By taking differentials with respect to ω on both sides of (3.25) and rearranging terms, we have

$$\frac{d}{d\omega}\arg Q\left(re^{j\omega}\right) = -\frac{\boldsymbol{q}^T\boldsymbol{\Lambda}\boldsymbol{\Psi}\left(re^{j\omega}\right)\boldsymbol{q}}{\left|Q\left(re^{j\omega}\right)\right|^2} \tag{3.26}$$

where

$$\boldsymbol{\Lambda} = \mathrm{diag}\{0,1,\dots,M\} \tag{3.27}$$

$$\begin{aligned} &\boldsymbol{\Psi}\left(re^{j\omega}\right) \\ &= \mathrm{Re}\left\{\boldsymbol{\varphi}_M\left(re^{j\omega}\right)\boldsymbol{\varphi}_M^H\left(re^{j\omega}\right)\right\} \\ &= \begin{bmatrix} 1 & r^{-1}\cos\omega & \cdots & r^{-M}\cos M\omega \\ r^{-1}\cos\omega & r^{-2} & \cdots & r^{-(M+1)}\cos(M-1)\,\omega \\ \vdots & \vdots & \ddots & \vdots \\ r^{-M}\cos M\omega & r^{-(M+1)}\cos(M-1)\,\omega & \cdots & r^{-2M} \end{bmatrix} \end{aligned} \tag{3.28}$$

In (3.27), $\mathrm{diag}\{a_0, a_1, \dots, a_n\}$ represents a diagonal matrix with a_i on its ith diagonal. By taking (3.26) into (3.21) and computing the integral over $[0, \pi]$, the stability constraint (3.21) is transformed to

$$\tau(r, \boldsymbol{q}) = \boldsymbol{q}^T\boldsymbol{G}(r, \boldsymbol{q})\boldsymbol{q} = 0 \tag{3.29}$$

where

$$\boldsymbol{G}(r,\boldsymbol{q}) = \int_0^{\pi} \frac{\boldsymbol{\Lambda}\boldsymbol{\Psi}\left(re^{j\omega}\right) + \boldsymbol{\Psi}(re^{j\omega})\boldsymbol{\Lambda}}{2|Q(re^{j\omega})|^2} d\omega \tag{3.30}$$

If $Q(z)$ has L ($\leq M$) roots outside C and $M-L$ roots inside C, it can be verified that $\tau(r,\boldsymbol{q}) = L\pi$. Then, given a denominator $\boldsymbol{q}$, $\tau(r,\boldsymbol{q})$ has a stair shape with respect to r.

Unfortunately, the stability constraint (3.29) cannot be directly incorporated into the design problem (3.10), due to the following difficulties:

1. The stability condition (3.29) represents a nonlinear equality constraint.
2. The matrix $\boldsymbol{G}(r,\boldsymbol{q})$ is dependent on denominator coefficients $\boldsymbol{q}$.
3. The matrix $\boldsymbol{G}(r,\boldsymbol{q})$ is indefinite.

The first difficulty can be overcome by adopting the following inequality

$$\tau(r,\boldsymbol{q}) = \boldsymbol{q}^T\boldsymbol{G}(r,\boldsymbol{q})\boldsymbol{q} \leq \rho \tag{3.31}$$

Decreasing ρ makes more poles move inside the circle C. When $0 < \rho < \pi$, all poles will lie inside C. In order to tackle the second difficulty, we adopt a similar technique used in Section 3.1. At the kth iteration, $\tau(r,\boldsymbol{q})$ is modified by

$$\tau\left(r,\boldsymbol{q}^{(k)}\right) = \boldsymbol{q}^{(k)T}\boldsymbol{G}\left(r,\boldsymbol{q}^{(k-1)}\right)\boldsymbol{q}^{(k)} \tag{3.32}$$

Since $\boldsymbol{G}\left(r,\boldsymbol{q}^{(k-1)}\right)$ is an indefinite matrix, this explicit stability constraint cannot be directly transformed into an SOC constraint. Therefore, we combine the stability constraint with the constraint (3.6) and obtain

$$\boldsymbol{x}^{(k)T}\widehat{\boldsymbol{A}}^{(k-1)}\boldsymbol{x}^{(k)} \leq \delta^2 \tag{3.33}$$

where

$$\widehat{\boldsymbol{A}}^{(k-1)} = \alpha\boldsymbol{A}^{(k-1)} + (1-\alpha)\widehat{\boldsymbol{G}}\left(r,\boldsymbol{q}^{(k-1)}\right) \tag{3.34}$$

$$\widehat{\boldsymbol{G}}\left(r,\boldsymbol{q}^{(k-1)}\right) = \begin{bmatrix} \boldsymbol{G}\left(r,\boldsymbol{q}^{(k-1)}\right) & \boldsymbol{0}_{(M+1)\times(N+1)} \\ \boldsymbol{0}_{(N+1)\times(M+1)} & \boldsymbol{0}_{(N+1)\times(N+1)} \end{bmatrix} \tag{3.35}$$

In (3.35), $\mathbf{0}_{m\times n}$ denotes a zero matrix of size m-by-n. Accordingly, $\boldsymbol{A}^{(k-1)}$ in (3.11) and (3.12) is replaced by $\widehat{\boldsymbol{A}}^{(k-1)}$. If the sequential design procedure described in Section 3.1 converges, it follows that $s^{(k)}(\omega)=\frac{|Q^{(k)}(e^{j\omega})|}{|Q^{(k-1)}(e^{j\omega})|}\to 1$ for $\forall\omega\in[0,\pi]$ as $k\to+\infty$. Then, we can obtain that

$$\begin{aligned}&\boldsymbol{G}\left(r,\boldsymbol{q}^{(k-1)}\right)\\&=\int_0^{\pi}\frac{\boldsymbol{\Lambda\Psi}\left(re^{j\omega}\right)+\boldsymbol{\Psi}\left(re^{j\omega}\right)\boldsymbol{\Lambda}}{2|Q^{(k)}(re^{j\omega})|^2}\cdot\left[s^{(k)}(\omega)\right]^2d\omega\Big|_{k\to+\infty}\\&\approx\boldsymbol{G}\left(r,\boldsymbol{q}^{(k)}\right)\end{aligned}\tag{3.36}$$

In practice, we can decrease α to achieve lower $\tau\left(r,\boldsymbol{q}^{(k)}\right)$, which corresponds to decreasing ρ of (3.31) as $k\to+\infty$. Therefore, besides the prescribed maximum pole radius r, the regularization coefficient α also plays an important role of restricting poles' locations. It is noteworthy that decreasing α makes $\widehat{\boldsymbol{A}}^{(k-1)}$ approach an indefinite matrix, which cannot be used to formulate the SOC constraint in (3.10). Thus, α cannot be too small. Fortunately, generally α is large enough to guarantee the positive definiteness of $\widehat{\boldsymbol{A}}^{(k-1)}$. Simulation experience indicates that α is normally within the range [0.99, 0.999999]. The effects of α on the final design results will be illustrated by Example 2 in the next section.

Finally, the major steps of the proposed sequential design method are summarized below:

Step 1. Given an ideal frequency response $D(\omega)$, filter orders N and M, a weighting function $W(\omega)$, set $k=0$ and choose an initial guess $\boldsymbol{x}^{(0)}$.

Step 2. Set $k=k+1$, and compute $W^{(k-1)}(\omega)$ by (3.5), $\widehat{\boldsymbol{A}}^{(k-1)}$ by (3.34) and $\boldsymbol{B}(\omega)$ by (3.16). Then utilize $\widehat{\boldsymbol{A}}^{(k-1)}$ to calculate $\boldsymbol{F}^{(k-1)}$ by (3.11) and $\boldsymbol{g}^{(k-1)}$ by (3.12). Finally, solve for $\boldsymbol{\eta}^{(k)}$ the SOCP problem (3.10) with peak error constraints (3.15).

Step 3. Update coefficients $\boldsymbol{x}^{(k)}$ by (3.9). If the stopping condition (3.13) is satisfied,

or k exceeds a predetermined maximum number of iterations, terminate the sequential design procedure. Otherwise, go to Step 2 and continue.

3.3 Simulations

In this section, four examples are presented to demonstrate the effectiveness of the proposed design method. At each iteration, the SOCP problem (3.10) is to be solved by *SeDuMi* [66] in MATLAB environment. Besides the peak and L_2 errors of magnitude (MAG) and group delay (GD), we also adopt the WLS approximation error E_{WLS} defined by (1.18) to evaluate design performances. In our designs, the step size γ in (3.9), the convergence tolerance ε in (3.13), and the maximum number of iterations are always chosen as 0.8, 10^{-6}, and 200, respectively.

3.3.1 Example 1

The first example taken from [7] is to design a lowpass digital filter. The ideal frequency response is defined by

$$D(\omega) = \begin{cases} e^{-j15\omega} & 0 \le \omega \le 0.5\pi \\ e^{-186.6(0.5\pi-\omega)}e^{-j15\omega} & 0.5\pi < \omega < \pi \end{cases}$$

Filter orders are chosen as $N = M = 18$. The maximum pole radius is set to $r = 0.99$. The weighting function $W(\omega)$ is set equal to 1 over the entire frequency band. The regularization coefficient α of (3.34) is chosen as 0.999 in this example. All the initial numerator coefficients are chosen equal to 1, and the initial denominator coefficient vector is set to $[1\ 0\ ...\ 0]^T$. The sequential design procedure converges to the final solution after 49 iterations. The maximum pole radius of the obtained IIR filter is 0.9226. All the filter coefficients are given in Table 3.1. The magnitude and group delay responses are shown in Fig. 3.1. All the error measurements are summarized in Table 3.2. For comparison, we also design a lowpass filter using the least 4-power method [7] under the same set of specifications. The maximum pole radius of the IIR filter obtained by [7] is 0.9407. The design results are also shown in Fig. 3.1 as dashed curves. All the error measurements of the corresponding IIR filter are also given in Table 3.2 for comparison.

It can be seen that the proposed method can achieve much better performances in the WLS sense.

Table 3.1 Filter Coefficients (p_0 to p_N and q_0 to q_M) of IIR digital filters Designed in Example 1

Proposed WLS design	$p_0 \sim p_4$	-1.0713e-002	-1.3178e-002	8.9219e-003	9.5276e-004	-8.1651e-003
	$p_5 \sim p_9$	-1.9799e-003	1.0818e-002	1.7638e-003	-1.5605e-002	-8.9659e-004
	$p_{10} \sim p_{14}$	2.5278e-002	-1.2083e-003	-5.2009e-002	7.5360e-003	2.2659e-001
	$p_{15} \sim p_{18}$	4.4907e-001	4.7822e-001	3.0489e-001	1.1057e-001	
	$q_0 \sim q_4$	1.0000e+000	-2.5196e-001	9.3246e-001	-2.2941e-001	8.3066e-002
	$q_5 \sim q_9$	2.2442e-002	-1.6792e-002	-5.7370e-003	5.6841e-003	1.9017e-003
	$q_{10} \sim q_{14}$	-2.1488e-003	-5.2038e-004	5.4899e-004	-2.1759e-004	7.2785e-004
	$q_{15} \sim q_{18}$	8.4456e-004	-4.4200e-003	5.6961e-003	-3.3667e-003	
Proposed WLS design with peak error constraints	$p_0 \sim p_4$	-3.7274e-003	-2.4165e-003	4.6995e-003	-2.8594e-003	-1.9867e-003
	$p_5 \sim p_9$	5.2531e-004	5.0396e-003	-1.9890e-003	-8.0441e-003	4.2797e-003
	$p_{10} \sim p_{14}$	1.4556e-002	-9.6013e-003	-3.4800e-002	2.6837e-002	1.9091e-001
	$p_{15} \sim p_{18}$	3.4038e-001	3.4966e-001	2.1667e-001	8.0380e-002	
	$q_0 \sim q_4$	1.0000e+000	-8.3160e-001	1.6551e+000	-1.1865e+000	7.9640e-001
	$q_5 \sim q_9$	-3.1829e-001	3.8143e-002	3.6156e-002	-1.4048e-002	-9.7852e-003
	$q_{10} \sim q_{14}$	7.9478e-003	5.1406e-003	-1.0034e-002	-1.2727e-003	2.2218e-002
	$q_{15} \sim q_{18}$	-3.7153e-002	3.5275e-002	-2.0688e-002	6.5262e-003	

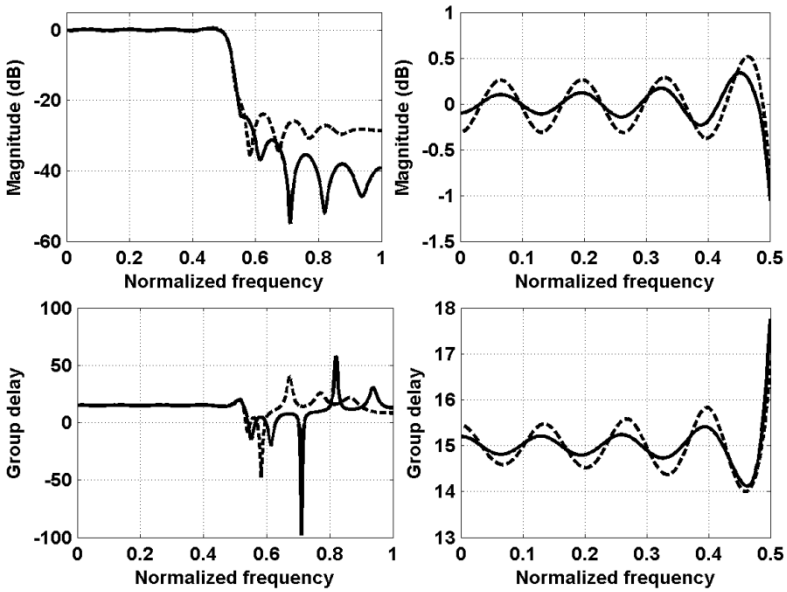

Fig. 3.1 Magnitude and group delay responses of IIR filters designed in Example 1. Solid curves: designed by the proposed method. Dashed curves: designed by the least 4-power method [7].

Table 3.2 Error Measurements of Design Results in Example 1

Method	WLS Error E_{WLS} (in dB)	Passband MAG (Peak/L_2 in dB)	Passband GD (Peak/ L_2)
Proposed	-48.586	-18.829/ -37.594	2.754/ 2.691e-1
Least 4-power [7]	-43.699	-20.308/ -33.920	1.849/ 3.364e-1

In order to illustrate the effectiveness of peak error constraints formulated in (3.15), we introduce a transition band into the original design, and then the ideal frequency response $D(\omega)$ is modified as

$$D(\omega) = \begin{cases} e^{-j15\omega} & 0 \le \omega \le 0.5\pi \\ 0 & 0.55\pi \le \omega < \pi \end{cases}$$

The regularization coefficient α is set to 0.99996 in this design. Then, we impose peak error constraints on 90 equally-spaced frequency points over the stopband $[0.55\pi, \pi]$ with $\mu(\omega_i) = 0.0178$ (−35 dB) for $\omega_i \in [0.55\pi, \pi]$ for $i = 1, 2, \ldots, 90$. The weighting function is set to 1 over the passband and stopband, and 0 over the transition band. After 65 iterations, the design procedure converges to the final solution. The maximum pole radius of the obtained filter is 0.9732. Both numerator and denominator coefficients of the obtained IIR filter are also listed in Table 3.1. The design results are shown in Fig. 3.2 as solid curves. We also adopt the WLS method [11] to design an IIR filter under the same set of specifications. Note that the WLS method [11] is essentially a special case of the least p-power method [7] with $p = 2$. The maximum pole radius of the IIR filter designed by [11] is 0.9620. The design results are also shown in Fig. 3.2 as dashed curves, and all the error measurements are summarized in Table 3.3 for comparison. In [11] and [7], the positive realness based stability constraint is employed to guarantee the stability of designed IIR filters, which is expressed by

$$\mathrm{Re}\{Q(e^{j\omega})\} = \boldsymbol{q}^T \cdot \mathrm{Re}\{\boldsymbol{\varphi}_M(e^{j\omega})\} \ge \nu, \qquad \forall \omega \in [0, \pi] \tag{3.37}$$

where ν is a small positive number. This stability constraint is only sufficient. Simulation results indicate that IIR filters designed by the proposed method do not always satisfy (3.37), whereas the obtained IIR filters are still stable.

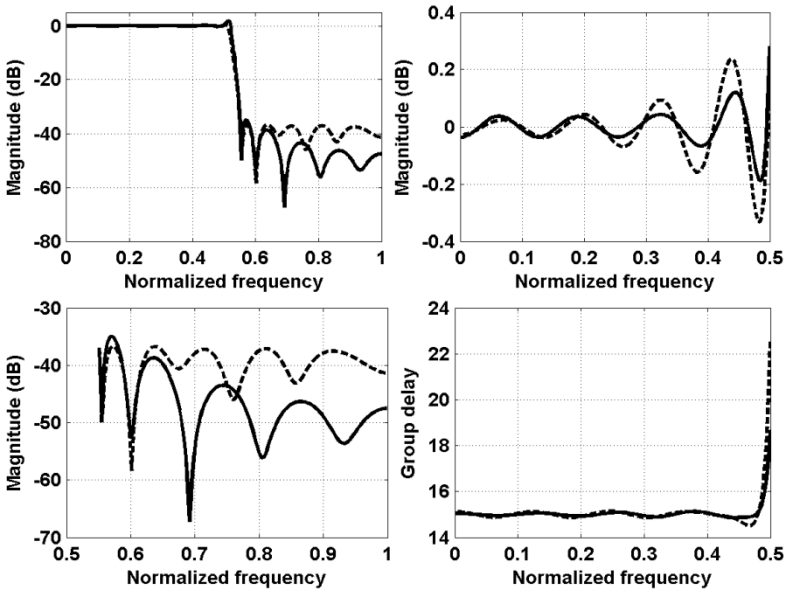

Fig. 3.2 Magnitude and group delay responses of IIR filters designed in Example 1 with peak error constraints. Solid curves: designed by the proposed method. Dashed curves: designed by the WLS method [11].

Table 3.3 Error Measurements of Design Results in Example 1 with Peak Error Constraints

Method	WLS Error E_{WLS} (in dB)	Passband MAG (Peak/L_2 in dB)	Passband GD (Peak/ L_2)	Stopband MAG (Peak/L_2 in dB)
Proposed	-74.162	-29.668/ -47.348	3.675/ 2.221e-1	-35.001/ -47.305
WLS [11]	-63.911	-28.466/ -42.058	7.528/ 4.687e-1	-36.467/ -42.709

3.3.2 Example 2

The second example is to design a halfband highpass filter [11], [28]. The ideal frequency response is given by

$$D(\omega) = \begin{cases} e^{-j12\omega} & 0.525\pi \le \omega < \pi \\ 0 & 0 \le \omega \le 0.475\pi \end{cases}$$

Numerator and denominator orders are chosen as $M = N = 14$. The prescribed maximum pole radius is set equal to $r = 1$. The weighting function is chosen as $W(\omega) = 1$ over the passband $[0.525\pi, \pi]$ and the stopband $[0, 0.475\pi]$, and 0 over the transition band $(0.475\pi, 0.525\pi)$. The regularization coefficient α is selected as 0.99996. The initial

numerator coefficients are all set equal to 1 as in Example 1. The initial poles are uniformly located on the unit circle, *i.e.*, $e^{\pm j\left(\frac{2\pi m}{M}-\frac{\pi}{M}\right)}$ for m = 1, 2, ..., M/2. Therefore, the initial denominator polynomial is chosen by

$$\begin{aligned} Q^{(0)}(z) &= \prod_{m=1}^{M/2}\left[1-z^{-1}e^{j\left(\frac{2\pi m}{M}-\frac{\pi}{M}\right)}\right]\cdot\left[1-z^{-1}e^{-j\left(\frac{2\pi m}{M}-\frac{\pi}{M}\right)}\right] \\ &= \prod_{m=1}^{M/2}\left[1-2z^{-1}\cos\left(\frac{2\pi m}{M}-\frac{\pi}{M}\right)+z^{-2}\right] \end{aligned} \tag{3.38}$$

Note that this initial IIR filter is unstable. In many sequential design methods (*e.g.*, the GN method [20]), unstable IIR filters cannot be used as initial designs. Otherwise, the stability constraints therein could become invalid. However, this is not required by the proposed design method. The stability of IIR filters designed by the proposed method can always be assured, provided the design procedure converges and the regularization parameter is appropriately selected. Starting from the initial point (3.38), the sequential design procedure reaches the final solution after 72 iterations. The maximum pole radius of the designed IIR filter is 0.9782. All the filter coefficients are listed in Table 3.4. The magnitude and group delay responses of the designed IIR filter are shown as solid curves in Fig. 3.3. For comparison, we also adopt the WLS method [28] proposed under the weighted integral of the squared error (WISE) criterion to design an IIR filter under the same specifications. The maximum pole radius of the obtained IIR filter is 0.9950. The magnitude and group delay responses of the corresponding IIR filter are also presented as dashed curves in Fig. 3.3. All the error measurements are given in Table 3.5. Apparently, the proposed method can achieve much better performances than the WISE method [28].

Table 3.4 Filter Coefficients (p_0 to p_N and q_0 to q_M) of IIR Digital Filter Designed in Example 2

p_0~p_4	6.8821e-005	8.6792e-003	1.3100e-002	6.2211e-003	-2.3882e-003
p_5~p_9	3.1138e-003	1.0164e-002	-5.8755e-003	-2.0533e-002	1.8923e-002
p_{10}~p_{14}	5.1744e-002	-1.5480e-001	2.1374e-001	-1.5531e-001	8.2951e-002
q_0~q_4	1.0000e+000	1.5137e+000	2.3726e+000	2.2287e+000	1.5549e+000
q_5~q_9	7.3344e-001	1.8059e-001	-2.7787e-002	-5.4524e-002	-4.7085e-002
q_{10}~q_{14}	-3.9418e-002	-2.0828e-002	7.1364e-005	7.5728e-003	3.8850e-003

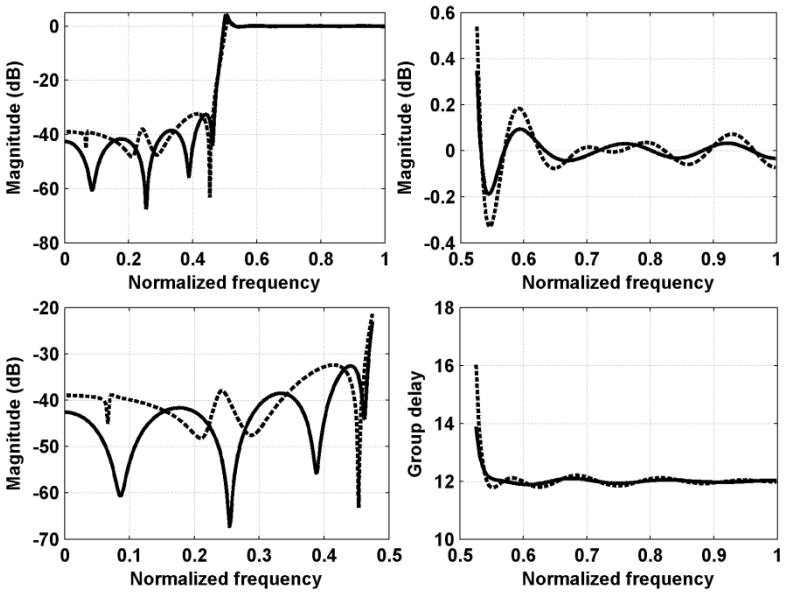

Fig. 3.3 Magnitude and group delay responses of IIR filters designed in Example 2. Solid curves: designed by the proposed method. Dashed curves: designed by the WISE method [28].

Table 3.5 Error Measurements of Design Results in Example 2

Method	WLS Error E_{WLS} (in dB)	Passband MAG (Peak/L_2 in dB)	Passband GD (Peak/ L_2)	Stopband MAG (Peak/L_2 in dB)
Proposed	-70.869	-27.769/ -47.505	1.887/ 1.234e-1	-23.064/ -42.801
WISE [28]	-64.096	-23.748/ -42.684	4.086/ 2.418e-1	-21.362/ -39.942

In order to demonstrate the effects of parameter α on final design results, we repeat the design procedure for 20 times by increasing α from 0.99 to 1. In all the designs, the admissible maximum pole radius is always set to 1. Fig. 3.4 shows the variation of maximum pole radii of the obtained IIR filters with respect to α. It can be observed that some poles approach the boundary of the prescribed stability domain when gradually augmenting α, which coincides with the previous discussion. Note that when α = 1, the design problem is essentially formulated without any stability constraint. We also plot the variation of total number of iterations in each design with respect to α in Fig. 3.5. When α = 1, the design procedure cannot converge within the specified maximum number of iterations. All the other design procedures converge to the final solutions within 40 iterations. Furthermore, it can be observed that with a smaller α the design procedure can converge to the final solution within a less number of iterations. However, the maximum

pole radius of the designed IIR filter can accordingly be reduced, which may degrade the design performance. Thus, in practical designs, the regularization coefficient α should be appropriately selected, such that we can achieve the balance between the design performance and the convergence speed. The simulation results presented in Fig. 3.4 and Fig. 3.5 suggest a way to choose α. First of all, given a maximum pole radius r, choose α = 1 and perform the design procedure. If the design procedure converges within the specified maximum number of iterations and all poles of the obtained IIR filter lie inside the prescribed stability domain, the design result can be accepted as the final solution. Otherwise, α should be gradually decreased until a satisfactory design is obtained. Actually, the values of α adopted in all the designs presented in this section are determined in this way.

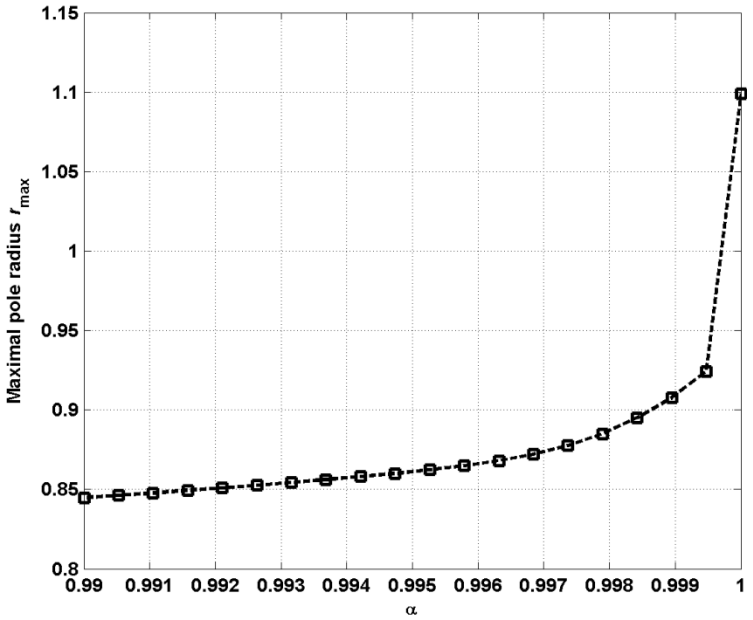

Fig. 3.4 Variation of maximum pole radii of designed IIR digital filters with respect to the regularization parameter α.

3.3.3 Example 3

Another lowpass digital filter with the following ideal frequency response is designed in this example

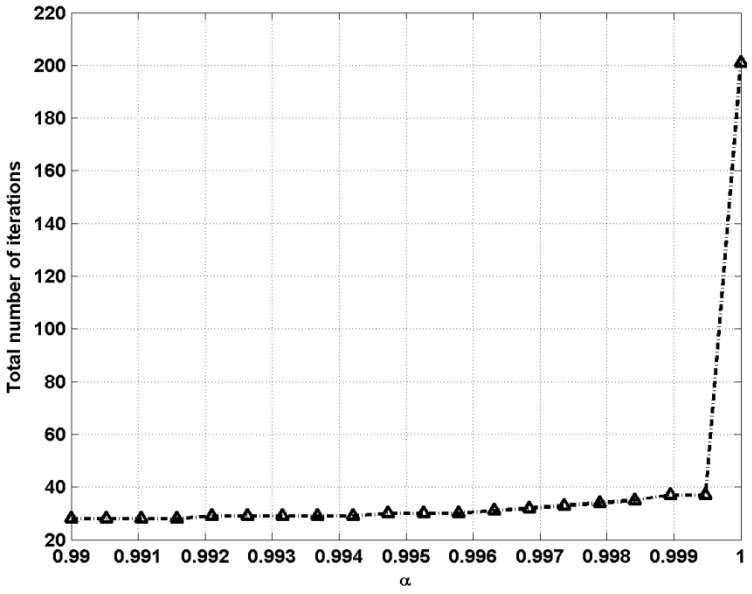

Fig. 3.5 Variation of total number of iterations with respect to the regularization parameter α.

$$D(\omega) = \begin{cases} e^{-j12\omega} & 0 \le \omega \le 0.4\pi \\ 0 & 0.56\pi \le \omega < \pi \end{cases}$$

The design specifications are exactly the same as those used by the first example in [21]. Filter orders are chosen as N = 15 and M = 4. The prescribed maximum pole radius is set to r = 0.84. The weighting function is specified as

$$W(\omega) = \begin{cases} 1 & 0 \le \omega \le 0.4\pi \\ 2.6 & 0.56\pi \le \omega < \pi \\ 0 & \text{otherwise} \end{cases}$$

The regularization coefficient α used in (3.34) is set to 0.999992. The same initial numerator and denominator coefficients are chosen as the same as in Example 1. After 12 iterations, the sequential design method converges to the final solution. The maximum pole radius of the designed IIR digital filter is 0.7896. Both numerator and denominator coefficients of the obtained IIR filter are summarized in Table 3.6. We also utilize the WLS method [21] to design an IIR filter under the same set of specifications. The

maximum pole radius of the corresponding filter is 0.7233. The magnitude and group delay responses of designed IIR filters are shown in Fig. 3.6. And all the error measurements are given in Table 3.7 for comparison. It can be observed that the proposed design method can achieve much reduction on the WLS approximation error E_{WLS}.

Table 3.6 Filter Coefficients (p_0 to p_N and q_0 to q_M) of IIR digital filter Designed in Example 3

p_0~p_4	-3.9873e-003	-1.4152e-003	6.1913e-003	3.7134e-003	-1.0342e-002
p_5~p_9	-8.8568e-003	1.6292e-002	1.9862e-002	-2.5760e-002	-4.9525e-002
p_{10}~p_{14}	4.6060e-002	2.3067e-001	3.4924e-001	3.0320e-001	1.5781e-001
p_{15}	4.1217e-002				
q_0~q_4	1.0000e+000	-5.3440e-001	7.9664e-001	-2.4615e-001	6.1287e-002

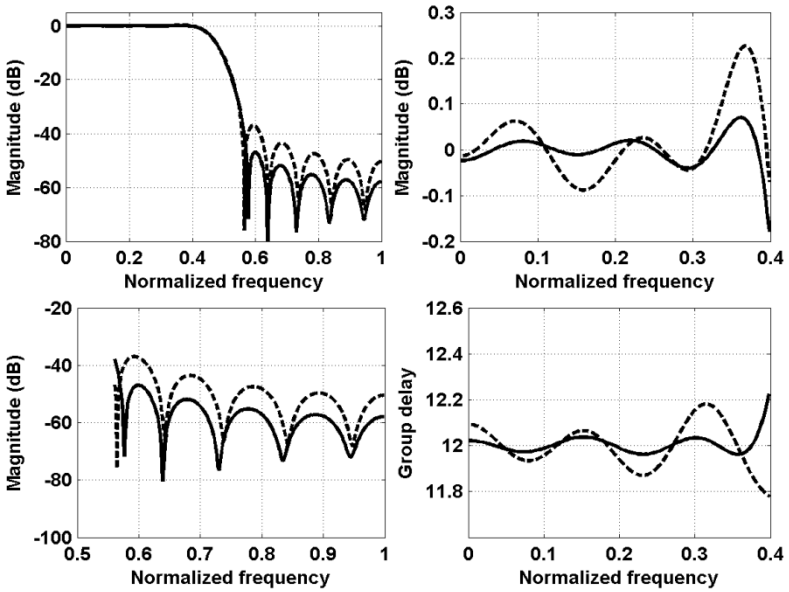

Fig. 3.6 Magnitude and group delay responses of IIR filters designed in Example 3. Solid curves: designed by the proposed method. Dashed curves: designed by the WLS method with linearized argument principle based stability constraint of [21].

Table 3.7 Error Measurements of Design Results in Example 3

Method	WLS Error E_{WLS} (in dB)	Passband MAG (Peak/L_2 in dB)	Passband GD (Peak/ L_2)	Stopband MAG (Peak/L_2 in dB)
Proposed	-89.138	-33.069/ -52.705	0.239/ 2.382e-2	-37.890/ -57.177
WLS [21]	-72.213	-31.547/ -44.707	0.223/ 5.946e-2	-36.990/ -49.421

A stability constraint based on the linearized argument principle is used by the WLS design method [21]. At the kth iteration, $\tau\left(r, \boldsymbol{q}^{(k)}\right)$ is approximated by its first-order Taylor series, and then the stability constraint (3.29) can be expressed by

$$\tau\left(r, \boldsymbol{q}^{(k)}\right) \approx \tau\left(r, \boldsymbol{q}^{(k-1)}\right) + \nabla^T \tau\left(r, \boldsymbol{q}^{(k-1)}\right) \boldsymbol{\eta}_q^{(k)} = 0 \tag{3.39}$$

where $\boldsymbol{\eta}_q^{(k)}$ is composed of the first M+1 elements of $\boldsymbol{\eta}^{(k)}$ to update the denominator coefficients. Assuming that at the previous iteration all poles lie inside C, then we have $\tau\left(r, \boldsymbol{q}^{(k-1)}\right) = 0$. Thus, the stability constraint (3.39) is simplified as

$$\nabla^T \tau\left(r, \boldsymbol{q}^{(k-1)}\right) \boldsymbol{\eta}_q^{(k)} = 0 \tag{3.40}$$

which is a linear equality constraint with respect to $\boldsymbol{\eta}_q^{(k)}$. The design procedures, which incorporate (3.40) as the stability constraint, have to start from a stable initial point. Fig. 3.7 shows the values of $\nabla^T \tau\left(r, \boldsymbol{q}^{(k-1)}\right) \boldsymbol{\eta}_q^{(k)}$ during the design procedure of the proposed method. It can be observed that the maximum pole radius of the designed IIR filter is still less than r, even though the linearized stability constraint (3.40) is not satisfied.

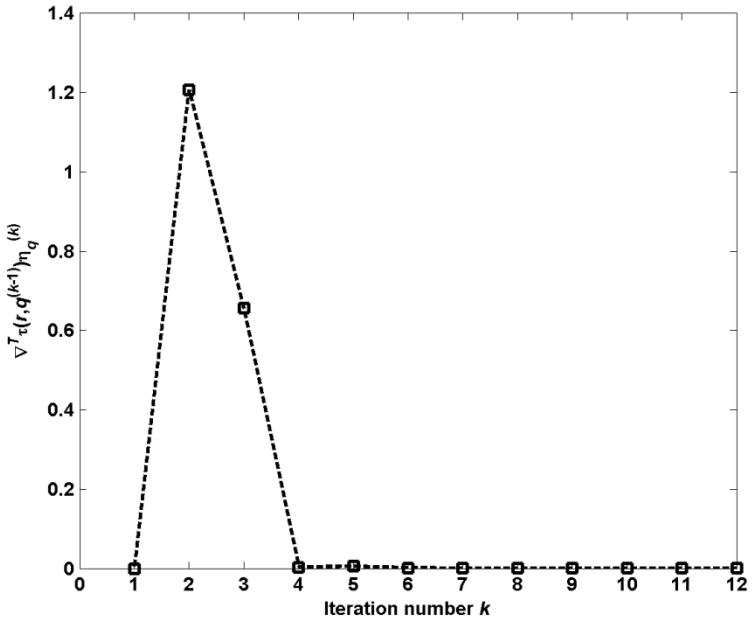

Fig. 3.7 Values of $\nabla^T \tau\left(r, \boldsymbol{q}^{(k-1)}\right) \boldsymbol{\eta}_q^{(k)}$ during the design procedure of the proposed method.

3.3.4 Example 4

The last example is to implement an equalization and anti-aliasing filter [20], [67], which follows an analog anti-aliasing filter and a sampler, to equalize the magnitude and phase (or group delay) responses of the analog filter in the passband and increase the attenuation in the stopband. The ideal frequency response of the cascaded system is defined by

$$D_c(\omega) = \begin{cases} e^{-j\omega\tau_d} & 0 \le \omega \le \dfrac{\pi}{16} \\ 0 & \dfrac{3\pi}{16} \le \omega < \pi \end{cases}$$

Then, the desired frequency response of the IIR equalization and anti-aliasing filter is $D_c(\omega)/H_a(j\omega/T)$, where $H_a(j\omega/T)$ is the frequency response of the analog filter and here T denotes the sampling period. The transfer function $H_a(s)$ of the analog filter has been given in [67]. The desired delay τ_d can be used as a free parameter to minimize the approximation error. In [67], the best FIR filter design according to the complex Chebyshev criterion has been presented with τ_d = 35, while an IIR filter with τ_d = 32 has been designed in [20] under the least-squares sense. In our designs, the best result can be obtained when τ_d = 34. Filter orders are chosen as N = 20 and M = 4. The prescribed maximum pole radius is chosen as r = 0.99. The regularization coefficient α is selected as 0.99994. In our design, the weighting function is chosen as

$$W(\omega) = \begin{cases} 100 & 0 \le \omega \le \dfrac{\pi}{16} \\ 1 & \dfrac{3\pi}{16} \le \omega < \pi \\ 0 & \text{otherwise} \end{cases}$$

The initial numerator and denominator coefficients are also chosen as $\boldsymbol{p}^{(0)} = [1\ 1\ \ldots\ 1]^T$, and $\boldsymbol{q}^{(0)} = [1\ 0\ \ldots\ 0]^T$. The proposed method converges to the final solution after 69 iterations. The maximum pole radius of the designed IIR filter is 0.9673. All the filter coefficients are given in Table 3.8. The magnitude responses, phase response errors, and group delays of analog filter, designed IIR equalization and anti-aliasing filter, and

cascaded system are all shown in Fig. 3.8. For comparison, we also design an IIR equalization and anti-aliasing filter under the same set of specifications using the GN method proposed by [20]. The maximum pole radius of the corresponding IIR filter is 0.9810. All the error measurements of IIR equalization and anti-aliasing filters are summarized in Table 3.9 for comparison. It can be seen that the proposed method can achieve better performances except the peak error of group delay on the passband.

Table 3.8 Filter Coefficients (p_0 to p_N and q_0 to q_M) of IIR Digital Filter Designed in Example 4

$p_0 \sim p_4$	4.4725e-003	-8.8140e-003	5.2339e-003	-1.5784e-003	-2.5343e-004
$p_5 \sim p_9$	-1.0631e-004	8.3618e-005	2.8017e-004	4.4110e-004	5.3299e-004
$p_{10} \sim p_{14}$	5.3437e-004	4.4628e-004	2.8717e-004	9.3356e-005	-9.4123e-005
$p_{15} \sim p_{19}$	-2.3516e-004	-3.0346e-004	-2.4784e-003	9.2375e-003	-1.4547e-002
p_{20}	8.2835e-003				
$q_0 \sim q_4$	1.0000e+000	-3.6483e+000	5.0585e+000	-3.1553e+000	7.4664e-001

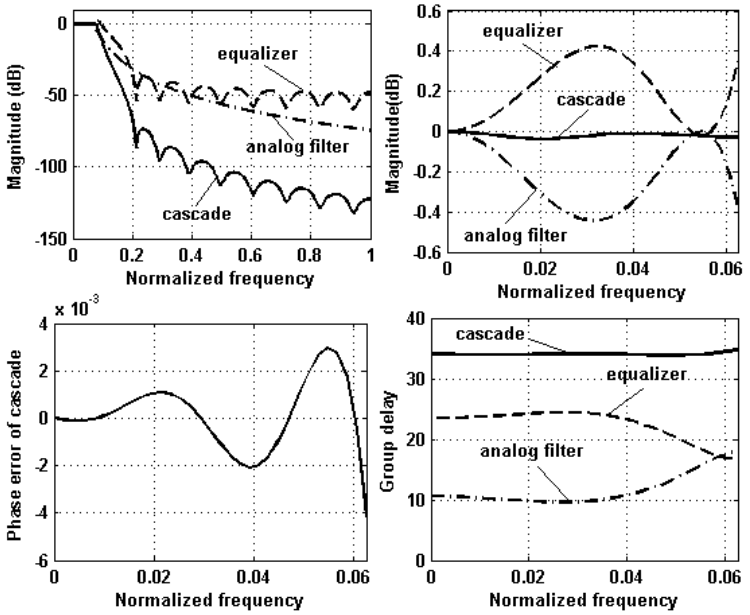

Fig. 3.8 Magnitude and group delay responses, and phase error of IIR filter designed in Example 4. Solid curves: cascaded system. Dashed curves: equalizer designed by the proposed method. Dash-dotted curves: analog filter.

Table 3.9 Error Measurements of Design Results in Example 4

Method	WLS Error E_{WLS} (in dB)	Passband MAG (Peak/L_2 in dB)	Passband GD (Peak/ L_2)	Stopband MAG (Peak/L_2 in dB)
Proposed	-77.786	-46.979/ -63.413	0.774/ 3.693e-2	-27.225/ -44.750
GN [20]	-74.334	-33.147/ -60.577	0.677/ 6.532e-2	-26.166/ -44.356

In [20], the Rouché's theorem is employed to develop a stability constraint: Given an initial denominator $Q^{(0)}(z)$ chosen with all its roots inside C, then all denominators $Q^{(k)}(z)$ (k = 1, 2, ...) have their roots inside C if the denominator updates $\Delta^{(k)}(z) = Q^{(k)}(z) - Q^{(k-1)}(z)$ satisfy

$$\begin{aligned} \left|\Delta^{(k)}\left(re^{j\omega}\right)\right| &= \left|\boldsymbol{\varphi}_M^T\left(re^{j\omega}\right)\boldsymbol{\eta}_q^{(k)}\right| \\ &\leq \left|Q^{(k-1)}\left(re^{j\omega}\right)\right|, \qquad \forall \omega \in [0, \pi] \end{aligned} \tag{3.41}$$

Like (3.37), the Rouché's theorem based stability constraint is only a sufficient condition to ensure stability. Moreover, these two constraints must be satisfied for $\forall \omega \in [0, \pi]$. A traditional way to incorporate these constraints is to impose them on a set of frequency points densely sampled over $[0, \pi]$, which, however, greatly increases the number of constraints. Another efficient way is to employ a multiple exchange algorithm to keep tracking the active constraints, such that only a finite number of stability constraints need to be incorporated. Unlike (3.37) and (3.41), the proposed stability constraint is realized over the whole frequency band $[0, \pi]$ instead of at each specific frequency. Thus, we do not need to enforce the stability constraint on a large number of frequency points or employ an inner iterative procedure for the multiple exchange algorithm.

CHAPTER IV

MINIMAX DESIGN OF IIR DIGITAL FILTERS USING SEQUENTIAL SOCP

In this chapter, we shall develop a new sequential design method in the minimax sense. Compared with some other sequential design methods, the most important advantage of this design method is that the convergence of the design procedure can be guaranteed. In order to tackle the nonconvexity of the original design problem, convex relaxation technique is to be introduced, such that the original design problem can be transformed to a relaxed SOCP design problem. By solving this relaxed design problem, lower and upper bounds of the minimum approximation error can be further estimated. By reducing the discrepancy between the original and relaxed design problems, a real minimax design can be finally obtained.

This chapter is organized as follows. The original design problem is first presented in Section 4.1. Then, convex relaxation technique is introduced to transform the original nonconvex design problem into a convex form. A sequential design procedure is presented in Section 4.1. Some practical issues are discussed in Section 4.2. Several design examples are presented in Sections 4.3.

4.1 Minimax Design Method

4.1.1 Problem Formulation

Using the complex approximation error $E(\omega)$ defined by (1.21) and the minimax approximation error $E_{MM}(\boldsymbol{x})$ defined by (1.20), the design problem of an IIR digital filter in the (weighted) minimax sense can be strictly expressed by

$$\min_{\boldsymbol{x}} E_{MM}(\boldsymbol{x}) = \min_{\boldsymbol{x}} \max_{\omega \in \Omega_I} |E(\omega)| \tag{4.1}$$

By introducing an auxiliary variable δ, the design problem (4.1) can be formulated as

$$\min \quad \delta \tag{4.2}$$

$$\text{s.t.} \quad x_0 = 1 \tag{4.2.a}$$

$$\begin{aligned} &\left|E(\omega)Q\left(e^{j\omega}\right)\right|^2 = \|\boldsymbol{C}(\omega)\boldsymbol{x}\|_2^2 \\ &\leq \delta \cdot \left|Q\left(e^{j\omega}\right)\right|^2 = \delta \cdot \|\boldsymbol{F}(\omega)\boldsymbol{q}\|_2^2, \qquad \omega \in \Omega_I \end{aligned} \tag{4.2.b}$$

where

$$\boldsymbol{C}(\omega) = W(\omega)\begin{bmatrix} \mathrm{Re}\left\{D(\omega)\boldsymbol{\varphi}_M^T(e^{j\omega})\right\} & -\mathrm{Re}\left\{\boldsymbol{\varphi}_N^T(e^{j\omega})\right\} \\ \mathrm{Im}\left\{D(\omega)\boldsymbol{\varphi}_M^T(e^{j\omega})\right\} & -\mathrm{Im}\left\{\boldsymbol{\varphi}_N^T(e^{j\omega})\right\} \end{bmatrix} \tag{4.3}$$

$$\boldsymbol{F}(\omega) = \begin{bmatrix} \mathrm{Re}\left\{\boldsymbol{\varphi}_M^T(e^{j\omega})\right\} \\ \mathrm{Im}\left\{\boldsymbol{\varphi}_M^T(e^{j\omega})\right\} \end{bmatrix} \tag{4.4}$$

Note that the term $|E(\omega)|$ in the original problem (4.1) has been replaced by its squared value in (4.2). The variable δ can be viewed as the (squared) approximation error limit in (4.2). It is obvious that the solution of (4.1) is also optimal to (4.2) and vice versa. Thereby, these two design problems are essentially equivalent to each other. In the design problems (4.1) and (4.2), there is an implicit constraint on the denominator $Q(z)$, that is, all roots of $Q(z)$ should lie inside the unit circle. For ease of discussion, we shall first describe the design method without any stability constraint. Then, the stability issue will be addressed in Section 4.2.

4.1.2 Convex Relaxation

It can be noticed that only the magnitude of the denominator is required on the right hand side of the inequality constraint (4.2.b), which will be used to develop the design method.

By introducing another polynomial $R(z)$ with coefficients d_m ($m = -M, -M+1, \ldots, M-1, M$), it can be verified that

$$
\begin{aligned}
Q(z)Q(z^{-1}) &= \left(\sum_{m=0}^{M} q_m z^{-m}\right)\left(\sum_{m=0}^{M} q_m z^{m}\right) \\
&= \sum_{m=-M}^{M} d_m z^{-m} \\
&= R(z)
\end{aligned}
\tag{4.5}
$$

where the polynomial coefficients of $R(z)$ can be computed by

$$
d_m = d_{-m} = \sum_{i=0}^{M-m} q_i q_{i+m}, \qquad m = 0,1,\ldots,M \tag{4.6}
$$

It is well known that $\{d_m, m = -M, \ldots, -1, 0, 1, \ldots, M\}$ is an autocorrelation sequence. Some important properties can be directly derived from (4.6):

1. $d_0 = \sum_{m=0}^{M} q_m^2 = \|\boldsymbol{q}\|_2^2$
2. $d_{\pm M} = q_0 q_M = q_M$, where $q_0 = 1$
3. $|d_{\pm m}| = |\boldsymbol{q}_{m,1}^T \boldsymbol{q}_{m,2}| \le \|\boldsymbol{q}_{m,1}\|_2 \cdot \|\boldsymbol{q}_{m,2}\|_2 \le \|\boldsymbol{q}\|_2^2 = d_0$ ($m = 0, 1, \ldots, M$), where $\boldsymbol{q}_{m,1} = [q_0 \ \ldots \ q_{M-m}]^{\mathrm{T}}$ and $\boldsymbol{q}_{m,2} = [q_m \ \ldots \ q_M]^{\mathrm{T}}$. The first inequality follows from the Cauchy-Schwartz inequality.

By defining $\boldsymbol{d} = [d_0 \ d_1 \ \ldots \ d_M]^T$ and $\boldsymbol{s}(\omega) = [1 \ 2\cos\omega \ \ldots \ 2\cos M\omega]^T$, and evaluating (4.5) on the unit circle, we have

$$
\begin{aligned}
\left|Q\left(e^{j\omega}\right)\right|^2 &= \|\boldsymbol{F}(\omega)\boldsymbol{q}\|_2^2 \\
&= \boldsymbol{d}^T \boldsymbol{s}(\omega) \\
&= R(e^{j\omega})
\end{aligned}
\tag{4.7}
$$

Using (4.7), the constraint (4.2.b) can be cast as a hyperbolic constraint by replacing $\left|Q(e^{j\omega})\right|^2$ by $\boldsymbol{d}^T \boldsymbol{s}(\omega)$ on the right hand side of the inequality. It is known that a hyperbolic constraint can be further transformed to an equivalent SOC constraint [68]. On the other hand, the feasible $\boldsymbol{q}$ and $\boldsymbol{d}$ should satisfy (4.7) for $\forall\omega \in [0, \pi]$. Then, the design problem (4.2) can be reformulated as

$$\min \quad \delta \tag{4.8}$$

$$\text{s.t.} \quad x_0 = 1 \tag{4.8.a}$$

$$\|\boldsymbol{C}(\omega)\boldsymbol{x}\|_2^2 \le \delta \cdot \boldsymbol{d}^T \boldsymbol{s}(\omega), \qquad \forall \omega \in \Omega_I \tag{4.8.b}$$

$$\|\boldsymbol{F}(\omega)\boldsymbol{q}\|_2^2 = \boldsymbol{d}^T \boldsymbol{s}(\omega), \qquad \forall \omega \in [0, \pi] \tag{4.8.c}$$

Note that the hyperbolic constraint (4.8.b) is enforced over Ω_I, while the quadratic equality constraint (4.8.c) must be satisfied for $\forall \omega \in [0, \pi]$. Although the trigonometric polynomial coefficients $\boldsymbol{d}$ are introduced as auxiliary variables in (4.8), they are closely related to the denominator coefficients $\boldsymbol{q}$ through (4.6). In order to establish the equivalence between the design problems (4.2) and (4.8), the constraint (4.6) should be incorporated in (4.8). However, the equality constraint (4.8.c) implies that the polynomial $R(z)$ is nonnegative on the unit circle. Then, based on the theorem of spectral factorization [69], we can find a causal polynomial $F(z) = \sum_{m=0}^{M} f_m z^{-m}$ with real coefficients such that $F(z)F(z^{-1}) = R(z)$. Although the spectral factorization is not unique, among all the possible spectral factorizations, there is only one minimum-phase polynomial. Then, in view of the stability requirement, $F(z)$ should be the unique minimum-phase polynomial. Thereby, $Q(z)$ is equivalent to $F(z)$, and (4.6) becomes a redundant constraint.

Due to the existence of the quadratic equality constraint (4.8.c), the design problem (4.8) is still nonconvex. However, we can relax it into a convex problem by replacing (4.8.c) by another hyperbolic inequality constraint $\|\boldsymbol{F}(\omega)\boldsymbol{q}\|_2^2 \le \boldsymbol{d}^T \boldsymbol{s}(\omega)$. Then, the design problem (4.8) is transformed to

$$\min \quad \delta \tag{4.9}$$

$$\text{s.t.} \quad x_0 = 1 \tag{4.9.a}$$

$$x_M - d_M = 0 \tag{4.9.b}$$

$$\|\boldsymbol{C}(\omega)\boldsymbol{x}\|_2^2 \le \delta \cdot \boldsymbol{d}^T \boldsymbol{s}(\omega), \qquad \forall \omega \in \Omega_I \tag{4.9.c}$$

$$\|\boldsymbol{F}(\omega)\boldsymbol{q}\|_2^2 \le \boldsymbol{d}^T\boldsymbol{s}(\omega), \qquad \forall\omega \in [0,\pi] \tag{4.9.d}$$

Since the equality constraint (4.8.c) is replaced by the SOC constraint (4.9.d), the variables $\boldsymbol{q}$ and $\boldsymbol{d}$ in (4.9) may not satisfy the equality constraint (4.6) any longer. For ease of later discussion, we represent the resulting difference between $\left|Q(e^{j\omega})\right|^2$ and $\boldsymbol{d}^T\boldsymbol{s}(\omega)$ by

$$\begin{aligned}\lambda(d_0,\boldsymbol{q}) &= \frac{1}{\pi}\int_0^{\pi}\left[\left|Q\left(e^{j\omega}\right)\right|^2 - \boldsymbol{d}^T\boldsymbol{s}(\omega)\right]d\omega \\ &= \|\boldsymbol{q}\|_2^2 - d_0\end{aligned} \tag{4.10}$$

From (4.9.d), we have $\lambda(d_0,\boldsymbol{q}) \le 0$. Hence, by introducing the relaxed constraint (4.9.d), the Property 1 of (4.6) has been accordingly relaxed to $\|\boldsymbol{q}\|_2^2 \le d_0$. Although the equality constraint (4.8.c) has been replaced by (4.9.d), the nonnegativity of $R(z)$ on the unit circle is still guaranteed. According to the theorem of spectral factorization, $\{d_m, m = -M, ..., -1, 0, 1, ..., M\}$ in (4.9) is still an autocorrelation sequence. Thus, the Property 3 of (4.6) can be automatically satisfied. The Property 2 of (4.6) is ensured by the constraint (4.9.b), which can pre-filter out unqualified x_M and d_M.

Let δ^* denote the optimal value of the original design problem (4.8), and δ_{rel}^* be the optimal value of the relaxed design problem (4.9). Since the feasible set defined by the relaxed constraint (4.9.d) is larger than that of (4.8.c), we always have $\delta_{rel}^* \le \delta^*$, which means a lower bound on the optimal value of the original design problem (4.8) can be obtained by solving (4.9). However, due to the existence of the relaxed constraint (4.9.d), δ_{rel}^* is not equal to the real (squared) minimax error of the IIR filter obtained by (4.9), which is denoted by $\delta_{mm}^* = \max_{\omega\in\Omega_I}|E(\omega)|^2$. Furthermore, δ_{mm}^* serves as an upper bound of δ^*, *i.e.*, $\delta_{rel}^* \le \delta^* \le \delta_{mm}^*$. By reducing the discrepancy between δ_{rel}^* and δ_{mm}^*, satisfactory designs can be achieved.

4.1.3 Sequential Design Procedure

In general, by solving the relaxed design problem (4.9), the obtained optimal value δ_{rel}^* is less than the real (squared) minimax error δ_{mm}^*, and the corresponding $\boldsymbol{x}$ and $\boldsymbol{d}$

cannot exactly satisfy the quadratic equality constraint (4.8.c) over the whole frequency band $[0, \pi]$. Hence, they are not the true solution for the minimax design problem (4.8). In this section, we will develop a design procedure, in which a sequence of SOCP problems based on (4.9) are subsequently solved so as to gradually reduce the discrepancy between $\boldsymbol{d}^T\boldsymbol{s}(\omega)$ and $\|\boldsymbol{F}(\omega)\boldsymbol{q}\|_2^2$ in (4.9.d). At the kth iteration, the filter coefficients $\boldsymbol{x}$ and the trigonometric polynomial coefficients $\boldsymbol{d}$ are updated by

$$\boldsymbol{x}^{(k)} = \boldsymbol{x}^{(k-1)} + \alpha\Delta\boldsymbol{x}^{(k)} \tag{4.11}$$

$$\boldsymbol{d}^{(k)} = \boldsymbol{d}^{(k-1)} + \alpha\Delta\boldsymbol{d}^{(k)} \tag{4.12}$$

where the step size α is chosen within the range of (0, 1), $\boldsymbol{x}^{(k-1)}$ and $\boldsymbol{d}^{(k-1)}$ are obtained at the previous iteration, and the search direction $\Delta\boldsymbol{x}^{(k)} = \left[\boldsymbol{u}^{(k)T}\ \boldsymbol{v}^{(k)T}\right]^T$ and $\Delta\boldsymbol{d}^{(k)}$ are determined at the current iteration. In $\Delta\boldsymbol{x}^{(k)}$, subvectors $\boldsymbol{u}^{(k)}$ and $\boldsymbol{v}^{(k)}$ are used to update the denominator and numerator coefficients, respectively. Suppose $x_0^{(k)} = 1$ for $k \geq 0$. Then, (4.9.a) can be replaced by another linear equality constraint $\Delta x_0^{(k)} = 0$. Since the integrand of (4.10) is always non-positive, $\left|\lambda(d_0^{(k)}, \boldsymbol{q}^{(k)})\right|$ can be regarded as the total discrepancy between $\|\boldsymbol{F}(\omega)\boldsymbol{q}\|_2^2$ and $\boldsymbol{d}^T\boldsymbol{s}(\omega)$ over $[0, \pi]$ at the kth iteration. Based on this observation, the proposed sequential design procedure attempts to gradually reduce $\left|\lambda(d_0^{(k)}, \boldsymbol{q}^{(k)})\right|$ as $k \to +\infty$. When $\left|\lambda(d_0^{(k)}, \boldsymbol{q}^{(k)})\right|$ is reduced to 0, the relaxed inequality constraint (4.9.d) will become the equality constraint (4.8.c). Let $\delta_{rel}^{(k)}$ denote the optimal value of the relaxed design problem (4.9) to be solved at the kth iteration, and $\delta_{mm}^{(k)}$ represent the corresponding squared minimax error of the obtained IIR filter. Then, according to the above analysis, we have $\lim_{k\to+\infty}\left(\delta_{mm}^{(k)} - \delta_{rel}^{(k)}\right) = 0$ if $\lim_{k\to+\infty}\left|\lambda(d_0^{(k)}, \boldsymbol{q}^{(k)})\right| = 0$. This property implies that a real minimax design can be attained by decreasing $\left|\lambda(d_0^{(k)}, \boldsymbol{q}^{(k)})\right|$.

Define a ratio $\gamma^{(k)}$ by

$$\gamma^{(k)} = \frac{\left|\lambda(d_0^{(k)}, \boldsymbol{q}^{(k)})\right|}{\left|\lambda(d_0^{(k-1)}, \boldsymbol{q}^{(k-1)})\right|} \tag{4.13}$$

At the kth iteration, we can impose the constraint $\gamma^{(k)} \leq \gamma < 1$ on $\lambda(d_0^{(k)}, \boldsymbol{q}^{(k)})$, which is equivalent to

$$\begin{aligned} &\lambda\left(d_0^{(k)}, \boldsymbol{q}^{(k)}\right) = \left\|\boldsymbol{q}^{(k)}\right\|_2^2 - d_0^{(k)} \\ &\geq \gamma \cdot \lambda\left(d_0^{(k-1)}, \boldsymbol{q}^{(k-1)}\right) = \gamma \cdot \left(\left\|\boldsymbol{q}^{(k-1)}\right\|_2^2 - d_0^{(k-1)}\right) \end{aligned} \tag{4.14}$$

Applying the above inequality recursively, we have

$$\lambda(d_0^{(k)}, \boldsymbol{q}^{(k)}) \geq \gamma^k \cdot \lambda(d_0^{(0)}, \boldsymbol{q}^{(0)}) \tag{4.15}$$

As $k \to +\infty$, the right-hand side of the above inequality will approach 0. Combined with $\lambda(d_0^{(k)}, \boldsymbol{q}^{(k)}) \leq 0$, it can be concluded that $\lim_{k\to+\infty}\left|\lambda(d_0^{(k)}, \boldsymbol{q}^{(k)})\right| = 0$. The major obstacle to incorporate the inequality constraint (4.14) into the relaxed SOCP design problem (4.9) is that (4.14) is still nonconvex. Here, the first-order Taylor series approximation is employed to linearize $\lambda(d_0^{(k)}, \boldsymbol{q}^{(k)})$ at $d_0^{(k-1)}$ and $\boldsymbol{q}^{(k-1)}$. The constraint $\lambda(d_0^{(k)}, \boldsymbol{q}^{(k)}) \geq \gamma \cdot \lambda(d_0^{(k-1)}, \boldsymbol{q}^{(k-1)})$ is then approximated by

$$\begin{aligned} &\lambda\left(d_0^{(k)}, \boldsymbol{q}^{(k)}\right) - \lambda\left(d_0^{(k-1)}, \boldsymbol{q}^{(k-1)}\right) \\ &\approx \nabla^T \lambda\left(d_0^{(k-1)}, \boldsymbol{q}^{(k-1)}\right) \begin{bmatrix} d_0^{(k)} - d_0^{(k-1)} \\ \boldsymbol{q}^{(k)} - \boldsymbol{q}^{(k-1)} \end{bmatrix} \\ &= -\Delta d_0^{(k)} + 2\boldsymbol{q}^{(k-1)T}\boldsymbol{u}^{(k)} \\ &\geq (\gamma - 1) \cdot \lambda\left(d_0^{(k-1)}, \boldsymbol{q}^{(k-1)}\right) \end{aligned} \tag{4.16}$$

Note that $\lambda(d_0^{(k)}, \boldsymbol{q}^{(k)})$ is a (convex) quadratic function of $d_0^{(k)}$ and $\boldsymbol{q}^{(k)}$. Then, the first-order Taylor series approximation serves as a global under-estimator of $\lambda(d_0^{(k)}, \boldsymbol{q}^{(k)})$.

Therefore, by imposing (4.16) on $d_0^{(k)}$ and $\boldsymbol{q}^{(k)}$ (or, equivalently, $\Delta d_0^{(k)}$ and $\boldsymbol{u}^{(k)}$), the inequality (4.14) can be definitely ensured. However, since the search direction is restricted in the halfspace defined by (4.16) instead of the original nonconvex set defined by (4.14), it cannot be guaranteed that the globally optimal solution will be certainly achieved by the proposed sequential design procedure.

Incorporating (4.16) into the relaxed design problem (4.9), then at the kth iteration the design problem (4.9) can be reformulated as

$$\min \quad \delta^{(k)} \tag{4.17}$$

$$\text{s.t.} \quad \Delta x_0^{(k)} = 0 \tag{4.17.a}$$

$$\Delta x_M^{(k)} - \Delta d_M^{(k)} = 0 \tag{4.17.b}$$

$$-\Delta d_0^{(k)} + 2\boldsymbol{q}^{(k-1)T}\boldsymbol{u}^{(k)} \geq (\gamma - 1) \cdot \lambda(d_0^{(k-1)}, \boldsymbol{q}^{(k-1)}) \tag{4.17.c}$$

$$\left\|\boldsymbol{C}(\omega_i)\boldsymbol{x}^{(k-1)} + \boldsymbol{C}(\omega_i)\Delta\boldsymbol{x}^{(k)}\right\|_2^2 \leq \delta^{(k)} \cdot \left(\boldsymbol{d}^{(k-1)} + \Delta\boldsymbol{d}^{(k)}\right)^T \boldsymbol{s}(\omega_i) \quad \omega_i \in \Omega_I, \qquad i = 1, 2, \dots, L \tag{4.17.d}$$

$$\left\|\boldsymbol{F}(\omega_j)\boldsymbol{q}^{(k-1)} + \boldsymbol{F}(\omega_j)\boldsymbol{u}^{(k)}\right\|_2^2 \leq \left(\boldsymbol{d}^{(k-1)} + \Delta\boldsymbol{d}^{(k)}\right)^T \boldsymbol{s}(\omega_j) \quad \omega_j \in [0, \pi], \qquad j = 1, 2, \dots, K \tag{4.17.e}$$

For simplicity, both (4.9.c) and (4.9.d) are imposed on a set of grid frequency points as (4.17.d) and (4.17.e), respectively. In (4.17), the decision variables are $\delta^{(k)}$, $\Delta\boldsymbol{x}^{(k)}$ (or $\boldsymbol{u}^{(k)}$ and $\boldsymbol{v}^{(k)}$), and $\Delta\boldsymbol{d}^{(k)}$. After solving (4.17), the obtained $\Delta\boldsymbol{x}^{(k)}$ and $\Delta\boldsymbol{d}^{(k)}$ are used to update the filter coefficients $\boldsymbol{x}^{(k-1)}$ and the trigonometric polynomial coefficients $\boldsymbol{d}^{(k-1)}$ through (4.11) and (4.12), respectively.

The sequential design procedure continues until the following condition is satisfied

$$\left|\lambda(d_0^{(k)}, \boldsymbol{q}^{(k)})\right| \leq \varepsilon \tag{4.18}$$

where ε is a prescribed convergence tolerance. Based on the previous analysis, which has shown that $\lim_{k\to+\infty}\left|\lambda(d_0^{(k)}, \boldsymbol{q}^{(k)})\right| = 0$, the convergence of the design procedure can be definitely assured. From (4.15), it can be further deduced that the design procedure will be terminated at the kth iteration if $\gamma^k \cdot \left|\lambda(d_0^{(0)}, \boldsymbol{q}^{(0)})\right| \le \varepsilon$. By taking logarithm on both sides of this inequality, an estimated maximum number $k_{\max}$ of iterations required by the sequential procedure can be obtained by

$$k_{\max} = \left\lfloor \frac{\ln \varepsilon - \ln \left|\lambda\left(d_0^{(0)}, \boldsymbol{q}^{(0)}\right)\right|}{\ln \gamma} \right\rfloor + 1 \tag{4.19}$$

where $\lfloor x \rfloor$ denotes the largest integer less than or equal to x. Moreover, if k becomes large enough, owing to $0 \le \lambda(d_0^{(k)}, \boldsymbol{q}^{(k)}) - \lambda(d_0^{(k-1)}, \boldsymbol{q}^{(k-1)}) \le -\lambda(d_0^{(k-1)}, \boldsymbol{q}^{(k-1)})$, we have

$$\begin{aligned} &\lambda\left(d_0^{(k)}, \boldsymbol{q}^{(k)}\right) - \lambda\left(d_0^{(k-1)}, \boldsymbol{q}^{(k-1)}\right) \\ &= \left\|\boldsymbol{u}^{(k)}\right\|_2^2 + 2\boldsymbol{q}^{(k-1)T}\boldsymbol{u}^{(k)} - \Delta d_0^{(k)} \\ &\approx 0 \end{aligned} \tag{4.20}$$

The constraint (4.16) indicates that $2\boldsymbol{q}^{(k-1)T}\boldsymbol{u}^{(k)} - \Delta d_0^{(k)} \ge 0$. Then, it follows from (4.20) that $\left\|\boldsymbol{u}^{(k)}\right\|_2 \approx 0$, which means there is no significant change on $\boldsymbol{q}^{(k)}$ as $k \to +\infty$.

4.2 Practical Considerations

4.2.1 Convergence Speed

As discussed in Section 4.1.3, the convergence of the sequential design procedure can be guaranteed if the linear inequality constraint (4.16) is incorporated. Obviously, a larger γ yields a larger feasible set for the search direction. Therefore, it is reasonable to choose γ as close to 1 as possible in order to achieve a satisfactory design. However, if γ is too close to 1, (4.19) shows that the total number of iterations required by the proposed design procedure could be too large. As an attempt to resolve this dilemma, we introduce

a new variable $\beta^{(k)} \geq 0$ to replace the term $(\gamma - 1) \cdot \lambda(d_0^{(k-1)}, \boldsymbol{q}^{(k-1)})$ on the right hand side of (4.16). Then, (4.16) is rewritten by

$$-\Delta d_0^{(k)} + 2\boldsymbol{q}^{(k-1)T}\boldsymbol{u}^{(k)} \geq \beta^{(k)} \tag{4.21}$$

In (4.21), $\beta^{(k)}$ serves as a soft threshold at each iteration. Apparently, in order to achieve the fastest convergence speed, we want to maximize $\beta^{(k)}$ (or minimize $-\beta^{(k)}$) at each iteration, while minimizing $\delta^{(k)}$ to reduce the approximation error. A common way to solve this bi-objective optimization problem is to minimize the weighted sum of these two objectives. By introducing a relative weight $\zeta > 0$, the design problem (4.17) is expressed by

$$\min \quad \delta^{(k)} - \zeta \cdot \beta^{(k)} \tag{4.22}$$

$$\text{s.t.} \quad \Delta x_0^{(k)} = 0 \tag{4.22.a}$$

$$\Delta x_M^{(k)} - \Delta d_M^{(k)} = 0 \tag{4.22.b}$$

$$-\Delta d_0^{(k)} + 2\boldsymbol{q}^{(k-1)T}\boldsymbol{u}^{(k)} - \beta^{(k)} \geq 0 \tag{4.22.c}$$

$$\beta^{(k)} \geq 0 \tag{4.22.d}$$

$$\left\|\boldsymbol{C}(\omega_i)\boldsymbol{x}^{(k-1)} + \boldsymbol{C}(\omega_i)\Delta\boldsymbol{x}^{(k)}\right\|_2^2 \leq \delta^{(k)} \cdot \left(\boldsymbol{d}^{(k-1)} + \Delta\boldsymbol{d}^{(k)}\right)^T \boldsymbol{s}(\omega_i) \quad \omega_i \in \Omega_I, \qquad i = 1, 2, \dots, L \tag{4.22.e}$$

$$\left\|\boldsymbol{F}(\omega_j)\boldsymbol{q}^{(k-1)} + \boldsymbol{F}(\omega_j)\boldsymbol{u}^{(k)}\right\|_2^2 \leq \left(\boldsymbol{d}^{(k-1)} + \Delta\boldsymbol{d}^{(k)}\right)^T \boldsymbol{s}(\omega_j) \quad \omega_j \in [0, \pi], \qquad j = 1, 2, \dots, K \tag{4.22.f}$$

The selection of parameter ζ is a tradeoff between the convergence speed and the design performance. The convergence of the sequential design procedure can be accelerated by increasing ζ, while the better performance can be attained by decreasing ζ. It seems that the effects of ζ used in the regularized design problem (4.22) are similar to those of γ used in (4.17). However, it should be emphasized that given γ the ratio $\gamma^{(k)}$ is confined at

each iteration by the constraint (4.16), and hence the convergence speed cannot be further improved. In contrast, the restriction on $\gamma^{(k)}$ has been removed in (4.21) by introducing the soft threshold $\beta^{(k)}$. Thereby, the modified design method can achieve faster convergence speed, which has been verified by a large number of simulation examples. In practice, when ζ is small enough, decreasing ζ contributes less to the performance improvement, and the convergence speed of the design procedure could be too slow for practical designs. If at each iteration the following constraint is still valid for some $\gamma < 1$

$$\beta^{(k)} \geq (\gamma - 1) \cdot \lambda(d_0^{(k-1)}, \boldsymbol{q}^{(k-1)}) \tag{4.23}$$

the convergence of the modified design method can also be strictly guaranteed. However, it should be noticed that (4.23) is only a sufficient condition for the convergence of the sequential design procedure, which implies that even without (4.23), the sequential procedure can still converge to the final solution when ζ is appropriately selected. The effects of ζ on final design results will be illustrated in Example 1 to be presented in the next section.

4.2.2 Stability Constraint

A sufficient condition for the stability of IIR filters in terms of positive realness has been proposed by [23], which can be stated as: If $Q^{(k-1)}(z)$ is a Schur polynomial, *i.e.*, all roots of $Q^{(k-1)}(z)$ lie inside the unit circle of the z plane, and the transfer function $G^{(k)}(z) = 1 + \frac{u^{(k)}(z)}{Q^{(k-1)}(z)}$ is strictly positive real (SPR), *i.e.*,

$$\mathrm{Re}\{G^{(k)}(e^{j\omega})\} > 0, \qquad \forall \omega \in [0, \pi] \tag{4.24}$$

where $u^{(k)}(z) = \boldsymbol{u}^{(k)T}\boldsymbol{\varphi}_M(z)$ ($u_0^{(k)} = 0$), then the weighted sum of $Q^{(k-1)}(z)$ and $u^{(k)}(z)$, *i.e.*, $Q_\alpha^{(k)}(z) = Q^{(k-1)}(z) + \alpha u^{(k)}(z)$ for $\forall \alpha \in [0, 1]$, is also a Schur polynomial. According to this condition, a stability domain with an interior point $\boldsymbol{q}^{(k-1)}$ can be defined by $D_s = \{\boldsymbol{u}^{(k)}: G^{(k)}(z) \text{ is SPR}\}$. The condition that $G^{(k)}(z)$ is SPR is equivalent to requiring that

$$G^{(k)}(z) + G^{(k)}(z^{-1})$$
$$= \frac{2Q^{(k-1)}(z)Q^{(k-1)}(z^{-1}) + u^{(k)}(z)Q^{(k-1)}(z^{-1}) + Q^{(k-1)}(z)u^{(k)}(z^{-1})}{Q^{(k-1)}(z)Q^{(k-1)}(z^{-1})} \tag{4.25}$$

is real and positive on the unit circle. Since the denominator of (4.25) is positive on the unit circle, it follows that the symmetric numerator polynomial of (4.25) must be positive on the unit circle, which is further cast in [23] as an LMI constraint independent of frequency ω. It has been proved [23] that this stability constraint defines a larger feasible domain than the Rouché's theorem based stability constraint [20].

Since SOCP problems cannot cope with LMI constraints, we express the stability constraint $G^{(k)}\left(e^{j\omega}\right)+G^{(k)}\left(e^{-j\omega}\right) > 0$ as the following linear inequality constraints:

$$\begin{aligned} &\mathrm{Re}\left\{Q^{(k-1)}\left(e^{-j\omega_j}\right)\boldsymbol{\varphi}_M^T\left(e^{j\omega_j}\right)\right\}\cdot \boldsymbol{u}^{(k)} \geq \mu - \left|Q^{(k-1)}\left(e^{j\omega_j}\right)\right|^2 \\ &\omega_j \in [0,\pi], \qquad j = 1, 2, \ldots, K \end{aligned} \tag{4.26}$$

where μ is a specified small positive number. If all poles of the designed IIR filters are required to lie inside a prescribed circle of radius $\rho < 1$ for robust stability, $\boldsymbol{\varphi}_M(e^{j\omega})$ and $Q^{(k-1)}(e^{j\omega})$ in (4.26) should be replaced by $\boldsymbol{\varphi}_M(\rho e^{j\omega})$ and $Q^{(k-1)}(\rho e^{j\omega})$, respectively. In general, parameter μ can be selected within [10^{-3}, 10^{-6}]. Simulation results show that generally design results are not very sensitive to the selection of μ.

4.2.3 Selection of Initial IIR Digital Filter

For sequential design methods, the selection of the initial design is a critical step to find a satisfactory solution. Without any prior knowledge of optimal IIR filters, initial guesses can be chosen as optimal FIR filters as suggested in [20], or IIR filters designed by the LP method [32] as suggested in [18]. Some other methods utilize more complicated multistage initialization strategy [23].

In our designs, the initial IIR filters are obtained by solving the relaxed SOCP problem (4.9). For stability, the constraint (4.26) should be incorporated in (4.9). The initial denominator can be simply assumed as $q_0^{(-1)} = 1$ and $q_m^{(-1)} = 0$ for $m = 1, 2, ..., M$. In this situation, the stability constraint (4.26) is equivalent to the positive realness based

stability constraint (3.37) proposed in [32]. Although only the lower and upper bounds on the optimal value δ^* of the original nonconvex problem (4.8) can be obtained, we find that the corresponding filter coefficients $\boldsymbol{x}$ and the trigonometric coefficients $\boldsymbol{d}$ can always lead to satisfactory solutions for all the designs we have tried so far. Some other guesses can also be utilized as the initial points of the sequential procedure. But $\boldsymbol{q}^{(0)}$ and $\boldsymbol{d}^{(0)}$ should satisfy (4.9.b) and (4.9.d).

Finally, the major steps of the proposed sequential design method are summarized as follows:

Step 1. Given an ideal frequency response $D(\omega)$, filter orders N and M, and a weighting function $W(\omega)$, set $k = 0$ and solve the relaxed design problem (4.9) to obtain initial coefficients $\boldsymbol{x}^{(0)}$ and $\boldsymbol{d}^{(0)}$.

Step 2. Set $k = k+1$, and solve the SOCP problem (4.22) to obtain $\Delta\boldsymbol{x}^{(k)}$ and $\Delta\boldsymbol{d}^{(k)}$. Update coefficients $\boldsymbol{x}^{(k)}$ and $\boldsymbol{d}^{(k)}$ by (4.11) and (4.12).

Step 3. If the stopping criterion (4.18) is satisfied, terminate the sequential design procedure. Otherwise, go to Step 2 and continue.

Some remarks about the proposed design method are made below:

1. In practice, after the sequential procedure converges to the final solution, some local optimization methods can be further applied to refine the design results. In our post-processing, we keep the obtained denominator coefficients fixed, and then the numerator coefficients are updated by solving the following SOCP problem:

$$\min \quad \delta \tag{4.27}$$

$$\text{s.t.} \quad \|\boldsymbol{G}(\omega_i)\boldsymbol{p} - \boldsymbol{g}(\omega_i)\|_2 \le \delta, \qquad \omega_i \in \Omega_I, \quad i = 0,1,\dots,L \tag{4.27.a}$$

where

$$\boldsymbol{G}(\omega) = W(\omega)\begin{bmatrix} \mathrm{Re}\left\{\dfrac{\boldsymbol{\varphi}_N^T(e^{j\omega})}{Q(e^{j\omega})}\right\} \\ \mathrm{Im}\left\{\dfrac{\boldsymbol{\varphi}_N^T(e^{j\omega})}{Q(e^{j\omega})}\right\} \end{bmatrix} \tag{4.28}$$

$$\boldsymbol{g}(\omega) = W(\omega)\begin{bmatrix} \mathrm{Re}\{D(\omega)\} \\ \mathrm{Im}\{D(\omega)\} \end{bmatrix} \tag{4.29}$$

For a given $\boldsymbol{q}$, the numerator obtained by (4.27) is optimal.

2. According to the previous analysis, it is clear that parameter ζ used in (4.22) should be appropriately selected. Through a large number of simulations, it is found that generally ζ can be chosen within [10^{-6}, 10]. Simulation results also show that both $\left|\lambda(d_0^{(k)}, \boldsymbol{q}^{(k)})\right|$ and the minimax approximation error can be dramatically reduced at the first several iterations even though $\zeta \ll 1$. As k increases, the convergence speed gradually slows down. Moreover, as k is large enough, we cannot achieve much reduction on the approximation error at each iteration. This observation implies that in practice, we can also employ a variable ζ in (4.22) during the proposed sequential design procedure. At the beginning of the sequential procedure, ζ can be chosen as a small value, such that the feasible set defined by (4.21) can be as large as possible. As k increases, parameter ζ can be accordingly augmented, such that the convergence of the sequential design procedure can be accelerated. Example 4 will be presented in the next section to demonstrate the effectiveness of the usage of a variable ζ.

4.3 Simulations

In this section, four examples are presented to demonstrate the effectiveness of the proposed design method. We still use the *SeDuMi* [66] in MATLAB environment to solve the SOCP problems (4.9) and (4.22). Besides the peak and L_2 errors of magnitude (MAG) and group delay (GD) over Ω_I, we also employ the minimax error E_{MM} defined by (1.20) to evaluate the design results. Without explicit declaration, the weighting

function is always set to $W(\omega) = 1$ for $\forall\omega \in \Omega_I$ and $W(\omega) = 0$ otherwise. Similarly, the admissible maximum pole radius ρ is always set equal to 1, unless it is explicitly specified. Parameter K is always equal to 101. Let S be the set of equally-spaced grid points over $[0, \pi]$, *i.e.*, $S = \{\omega_j: \omega_j = \frac{\pi(j-1)}{K-1}, j = 1, 2, ..., K\}$. The hyperbolic constraints (4.9.c) and (4.22.e) are then imposed on a set of frequency points taken from S, that is, $\{\omega_i: \omega_i \in S \cap \Omega_I\}$. Generally speaking, a larger K can lead to a more accurate design. However, in practice, this effect is almost negligible when K is large enough, *e.g.*, 100 or more. With a larger K, the proposed sequential design procedure needs more computation time to find the final solution. Note that the total number of iterations is normally not changed. The extra computation time is expended to construct the extra constraints and solve the SOCP problem of a larger size at each iteration. Our simulation experience indicates that when K is between 100 and 500, the computation time is acceptable. In all the simulation examples, step size α and parameter ε used in (4.18) are set, respectively, to 0.5 and 10^{-5}. In our designs, parameter μ used in the stability constraint (4.26) is always chosen as 10^{-3}.

4.3.1 Example 1

The first example is to design a lowpass IIR filter whose specifications are the same as those adopted in [20]. The ideal frequency response is defined by

$$D(\omega) = \begin{cases} e^{-j12\omega} & 0 \le \omega \le 0.4\pi \\ 0 & 0.56\pi \le \omega < \pi \end{cases}$$

Filter orders are chosen as $N = 15$ and $M = 4$. In this design, parameter ζ used in (4.22) is set to 0.001. After 24 iterations, the sequential procedure converges to the final solution. The maximum pole radius of the obtained IIR filter is 0.8598. All the filter coefficients are listed in Table 4.1. The magnitude and group delay responses are shown as solid curves in Fig. 4.1. The magnitude of the weighted complex error $E(\omega)$ is plotted in Fig. 4.2. For comparison, we also employ the design method proposed by [19] to design an IIR filter under the same set of specifications. Instead of an Mth-order polynomial used in (1.4), the denominator utilized by [19] is expressed as a product of second-order factors

and a first-order factor if M is odd, *i.e.*, $Q(z) = (1 + b_0 z^{-1}) \prod_{i=1}^{I} (1 + b_{i,1} z^{-1} + b_{i,2} z^{-2})$, where $b_0 = 0$ if M is even, and $I = (M-1)/2$ if M is odd or $M/2$ if M is even. Then, the first-order Taylor series approximation is directly applied on the frequency response $H(e^{j\omega})$ with respect to the numerator coefficients p_n (n = 0, 1, …, N) and the factorized denominator coefficients b_0, $b_{i,1}$, and $b_{i,2}$ (i = 1, 2, …, I), and subsequently the design problem at each iteration can be formulated as an SOCP problem. The advantage of adopting the factorized denominator is that the stability constraint can be cast as a set of linear inequality constraints in terms of b_0, $b_{i,1}$, and $b_{i,2}$, which are independent of the frequency ω:

$$\begin{bmatrix} 1 \\ -1 \end{bmatrix} b_0 + \rho^2 \begin{bmatrix} 1 \\ 1 \end{bmatrix} \geq \mathbf{0}_{2\times 1} \tag{4.30}$$

$$\begin{bmatrix} 1 & 1 \\ -1 & 1 \\ 0 & -1 \end{bmatrix} \cdot \begin{bmatrix} b_{i,1} \\ b_{i,2} \end{bmatrix} + \rho^2 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \geq \mathbf{0}_{3\times 1}, \quad i = 1, 2, \ldots, I \tag{4.31}$$

These constraints are sufficient and near necessary conditions for stability. At the beginning of the SOCP design method [19], all poles are simply placed at the origin, *i.e.*, $b_0 = b_{i,1} = b_{i,2} = 0$ for i = 1, 2, …, I. The initial numerator is obtained by solving (4.27) with the initial denominator specified above. The maximum pole radius of the IIR filter designed by [19] is 0.8590. The magnitude and group delay responses of the corresponding IIR filter are also shown in Fig. 4.1 as dashed curves. All the error measurements of both designs are summarized in Table 4.2 for comparison. It can be observed that the proposed method can achieve slightly better performance in E_{MM} than the SOCP method [19].

Table 4.1 Filter Coefficients (p_0 to p_N and q_0 to q_M) of IIR Digital Filter Designed in Example 1

p_0~p_4	-2.7223e-003	-2.2388e-003	3.8713e-003	3.6209e-003	-6.7370e-003
p_5~p_9	-8.1179e-003	1.0796e-002	1.7887e-002	-1.7854e-002	-4.5345e-002
p_{10}~p_{14}	3.4012e-002	2.2196e-001	3.7787e-001	3.7109e-001	2.2094e-001
p_{15}	7.5230e-002				
q_0~q_4	1.0000e+000	-4.5908e-001	8.9299e-001	-2.5445e-001	8.1335e-002

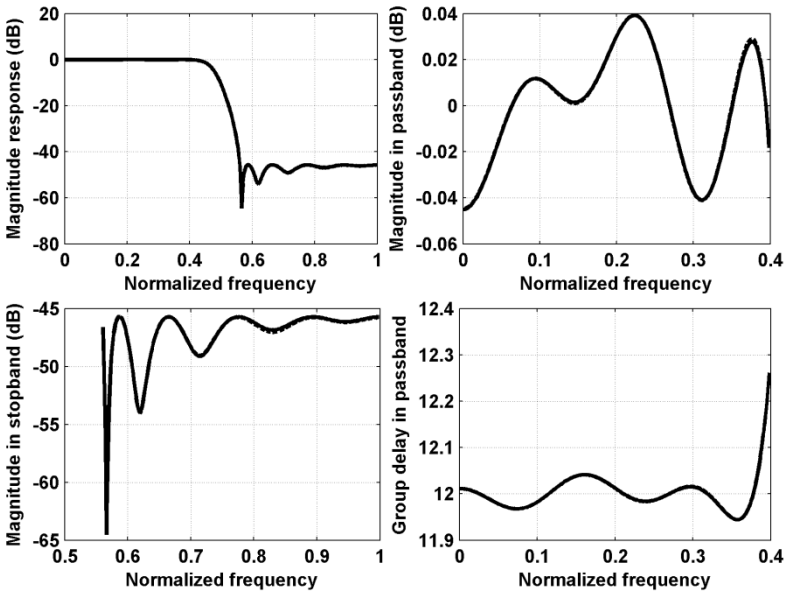

Fig. 4.1. Magnitude and group delay responses of IIR filters designed in Example 1. Solid curves: designed by the proposed method. Dashed curves: designed by the SOCP method [19].

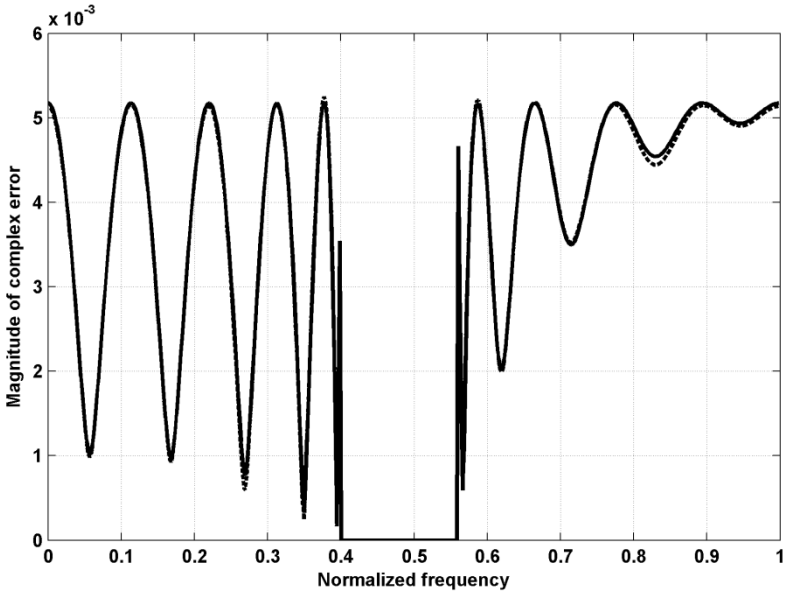

Fig. 4.2. Magnitude of weighted complex error of IIR filters designed in Example 1. Solid curves: designed by the proposed method. Dashed curves: designed by the SOCP method [19].

Table 4.2 Error Measurements of Design Results in Example 1

Method	Minimax Error E_{MM} (in dB)	Passband MAG (Peak/L_2 in dB)	Passband GD (Peak/ L_2)	Stopband MAG (Peak/L_2 in dB)
Proposed	-45.721	-45.722/ -55.167	2.814e-1/ 2.560e-2	-45.719/ -50.355
SOCP [19]	-45.615	-45.785/ -55.148	2.785e-1/ 2.538e-2	-45.657/ -50.396

In order to illustrate the effects of the regularization parameter ζ on the final design results, we repeat the experiment using ten different ζ's, which are taken within the interval [5×10^{-4}, 5×10^{-3}]. All the other design specifications are unchanged. Fig. 4.3 shows the variation of the minimax error E_{MM} versus the regularization parameter ζ. It can be noticed that the design performances can be improved by decreasing ζ. This coincides with our previous discussion. In all the designs, the sequential design procedure can converge to final solutions within at most 28 iterations. However, when ζ is too small (in this example, $\zeta \leq 10^{-4}$), the sequential design procedure converges in a very slow speed. Moreover, as ζ is sufficiently small (in this example, $\zeta \leq 10^{-3}$), it is difficult to further improve the design performance. Fig. 4.3 suggests us a way to find an appropriate regularization parameter ζ: First of all, we can choose a large value for ζ (*e.g.*, 1). Then, we gradually reduce the value of ζ until the improvement of design performances is negligible, or the sequential design procedure cannot converge within a prescribed maximum number of iterations (*e.g.*, 50). Except the variable ζ adopted in Example 4, the values of ζ used in all the other examples presented in this section are chosen in a similar way.

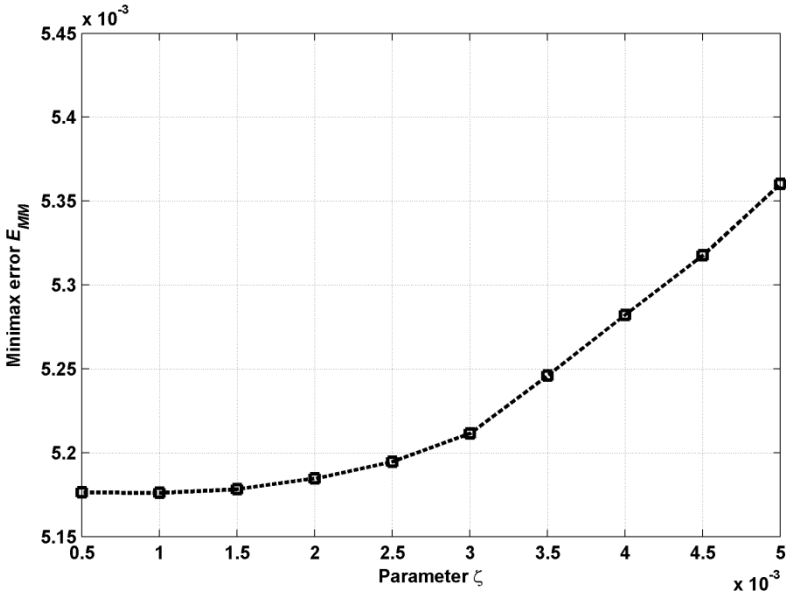

Fig. 4.3. Variation of minimax error E_{MM} versus parameter ζ.

4.3.2 Example 2

The second example is to design a highpass IIR filter [11], [28]. The filter orders are chosen as $N = M = 14$, and the ideal frequency response is defined by

$$D(\omega) = \begin{cases} e^{-j12\omega} & 0.525\pi \le \omega < \pi \\ 0 & 0 \le \omega \le 0.475\pi \end{cases}$$

Originally, the maximum pole radius is set as $\rho = 1$. However, the design results show that there is a magnitude overshoot within the transition band. Thereby, we reduce ρ from 1 to 0.96. Correspondingly, parameter ζ is set equal to 0.3. After 30 iterations, the sequential design procedure converges to the final solution. The maximum pole radius of the obtained IIR filter is 0.9559. All the filter coefficients are listed in Table 4.3. The magnitude and group delay responses are shown as solid curves in Fig. 4.4. The magnitude of the weighted complex error is plotted in Fig. 4.5. During the sequential design procedure, the optimal value $\delta_{rel}^{(k)}$ and the real minimax approximation error $\delta_{mm}^{(k)}$ are recorded at each iteration. The variation of discrepancy between $\delta_{mm}^{(k)}$ and $\delta_{rel}^{(k)}$, *i.e.*, $\delta_{mm}^{(k)} - \delta_{rel}^{(k)}$, versus the iteration index k is shown in Fig. 4.6. The changes of $\delta_{mm}^{(k)}$ and $\delta_{rel}^{(k)}$ at various iterations are also presented in Fig. 4.6. It can be observed that at the initial stage of the sequential procedure (in this example, $k \le 5$), $\delta_{mm}^{(k)}$ and $\delta_{mm}^{(k)} - \delta_{rel}^{(k)}$ decrease fast. Then, the sequential design procedure reaches a steady stage until the stopping condition is satisfied. In contrast, the optimal value $\delta_{rel}^{(k)}$ of the design problem (4.22) first increases, and then gradually decreases. Actually, in all the designs we have tried so far, $\delta_{mm}^{(k)} - \delta_{rel}^{(k)}$, $\delta_{mm}^{(k)}$ and $\delta_{rel}^{(k)}$ change at each iteration in a similar way.

Table 4.3 Filter Coefficients (p_0 to p_N and q_0 to q_M) of IIR Digital Filter Designed in Example 2

$p_0 \sim p_4$	-9.1215e-003	1.7383e-002	5.5492e-003	-3.6005e-003	-9.0162e-003
$p_5 \sim p_9$	6.9193e-003	9.9406e-003	-1.5344e-002	-1.3480e-002	4.9121e-002
$p_{10} \sim p_{14}$	5.3284e-002	-1.8564e-001	3.5661e-001	-2.5639e-001	1.8654e-001
$q_0 \sim q_4$	1.0000e+000	7.6364e-001	1.225e+000	7.5411e-001	4.5514e-001
$q_5 \sim q_9$	3.1573e-001	2.5391e-001	1.4183e-001	-4.7181e-003	-8.9604e-002
$q_{10} \sim q_{14}$	-9.1586e-002	-7.0897e-002	-6.7160e-002	-5.2190e-002	-3.5435e-002

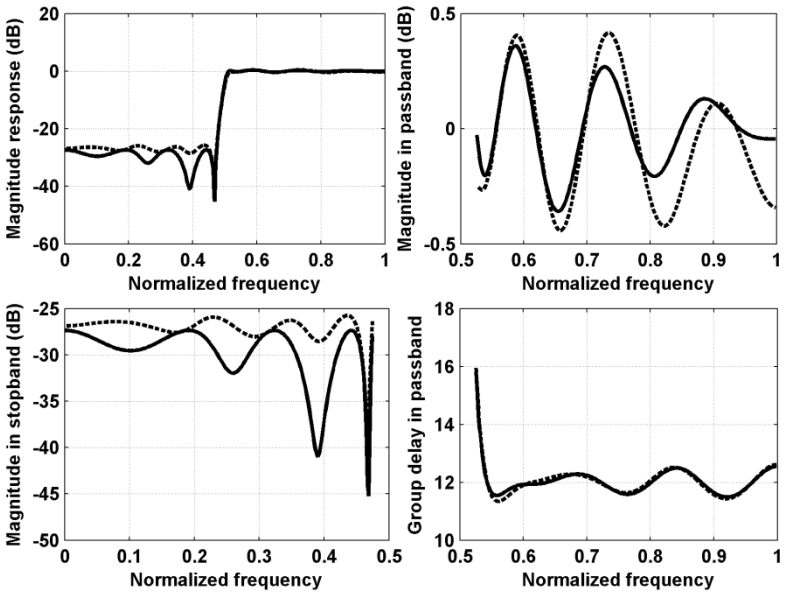

Fig. 4.4. Magnitude and group delay responses of IIR filters designed in Example 2. Solid curves: designed by the proposed method. Dashed curves: designed by the SM method [8].

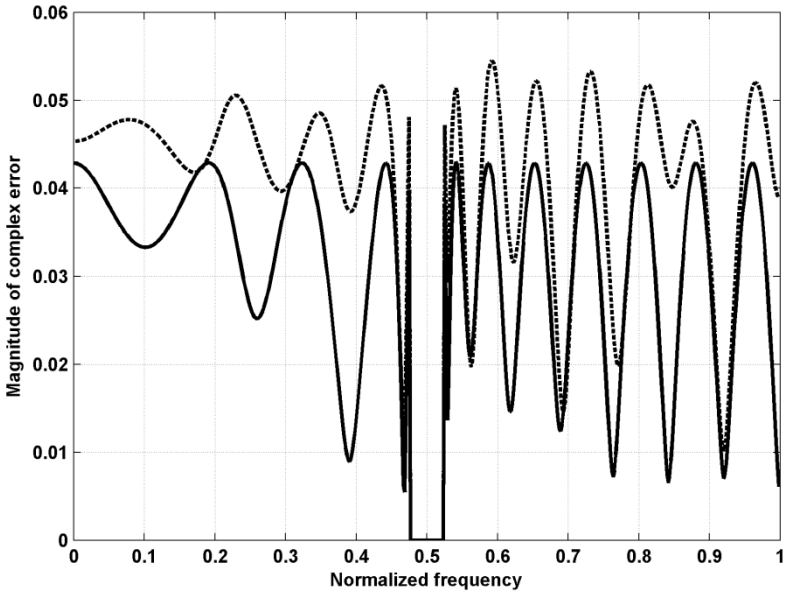

Fig. 4.5. Magnitude of weighted complex error of IIR filters designed in Example 2. Solid curves: designed by the proposed method. Dashed curves: designed by the SM method [8].

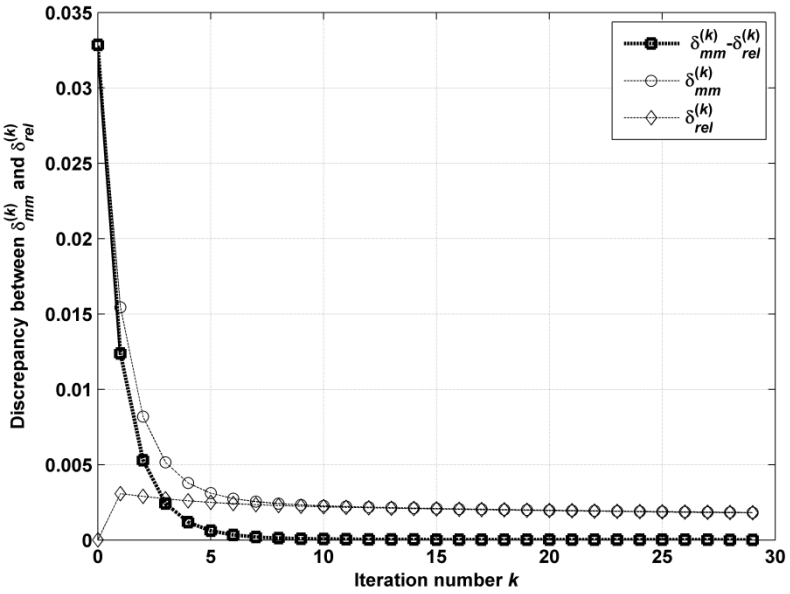

Fig. 4.6. Variation of discrepancy between $\delta_{mm}^{(k)}$ and $\delta_{rel}^{(k)}$ versus iteration number k.

For comparison, we also utilize the SM method [8] to design an IIR filter under the same set of specifications, except that the maximum pole radius is chosen as ρ = 1. The initial denominator for the SM method [8] is simply set as $\boldsymbol{q}^{(0)} = [1\ 0\ \ldots\ 0]^T$, and the initial numerator is chosen as the optimal FIR filter of order N = 14, which can be obtained by solving (4.27) with the initial denominator $\boldsymbol{q}^{(0)}$. The maximum pole radius of the obtained IIR digital filter is 0.9427. All the error measurements for both designs are listed in Table 4.4 for comparison. Apparently, the proposed method can achieve about 2dB reduction on the minimax approximation error E_{MM} than the SM method [8].

Table 4.4 Error Measurements of Design Results in Example 2

Method	Minimax Error E_{MM} (in dB)	Passband MAG (Peak/L_2 in dB)	Passband GD (Peak/ L_2)	Stopband MAG (Peak/L_2 in dB)
Proposed	-27.334	-27.456/ -37.093	3.823/ 3.076e-1	-27.359/ -32.341
SM [8]	-25.273	-26.081/ -33.597	4.011/ 3.399e-1	-25.740/ -30.240

4.3.3 Example 3

The third example is to design a two-band IIR digital filter [28] with the desired frequency response given by

$$D(\omega) = \begin{cases} e^{-j14.3\omega} & 0 \le \omega \le 0.46\pi \\ 0.5e^{-j20\omega} & 0.54\pi \le \omega < \pi \end{cases}$$

The maximum pole radius is set as 0.95. A group of IIR filters are designed by the proposed method, each of them totally having 31 filter coefficients, *i.e.*, N+M+1 = 31. The denominator order M changes from 0 to 15. Table 4.5 lists the minimax error of each design. The best design is attained when M = 6 and N = 24. The corresponding ζ used in this design is 0.005. The maximum pole radius of the obtained filter in the best design is 0.9486, and all the filter coefficients are given in Table 4.6. The magnitude and group delay responses over Ω_I are shown in Fig. 4.7. The magnitude of $E(\omega)$ is plotted in Fig. 4.8. For comparison, we also utilize the WISE method [28] to design an IIR filter. Since the WISE method is originally proposed for the WLS designs, the reweighting technique is used by [28] to achieve minimax designs. At each iteration, the original weighting function is successively multiplied by the envelope of $\left|E^{(k)}(\omega)\right|$, such that the minimax error can be accordingly reduced at the next iteration by solving the WLS design problem with the new weighting function. The minimax errors for IIR filters designed by the WISE method are also given in Table 4.5 for comparison. Obviously, the proposed sequential design method can achieve much better performances than the WISE method [28] in most of designs.

Table 4.5 Minimax Errors of Design Results in Example 3

M	E_{MM}		M	E_{MM}		M	E_{MM}	
	Proposed	WISE [28]		Proposed	WISE [28]		Proposed	WISE [28]
0	6.484e-2	2.605e-1	6	1.054e-2	1.180e-1	12	2.385e-2	1.611e-1
1	6.668e-2	2.625e-1	7	1.122e-2	1.182e-1	13	3.154e-1	5.744e-1
2	1.133e-2	1.133e-1	8	1.513e-2	1.452e-1	14	3.114e-1	5.590e-1
3	1.155e-2	1.141e-1	9	1.130e-2	1.191e-1	15	5.103e-1	7.232e-1
4	1.073e-2	1.211e-1	10	2.463e-2	1.859e-1			
5	1.077e-2	1.236e-1	11	2.964e-2	1.755e-1			

Table 4.6 Filter Coefficients (p_0 to p_N and q_0 to q_M) of IIR Digital filter (N = 24, M = 6) Designed in Example 3

$p_0 \sim p_4$	-2.6325e-003	1.2557e-002	3.2471e-003	4.5004e-003	-5.1975e-004
$p_5 \sim p_9$	7.3894e-003	4.2937e-003	-8.8385e-003	-6.6589e-003	1.5448e-002
$p_{10} \sim p_{14}$	1.2607e-002	-3.3186e-002	-3.3382e-002	1.3066e-001	4.3430e-001
$p_{15} \sim p_{19}$	6.8400e-001	7.0396e-001	4.9731e-001	1.9775e-001	-7.8242e-002
$p_{20} \sim p_{24}$	2.2166e-001	-3.0361e-001	2.5119e-001	-1.3502e-001	4.0037e-002
$q_0 \sim q_4$	1.0000e+000	1.2759e-001	1.1538e+000	1.1822e-001	2.0907e-001
$q_5 \sim q_6$	8.2517e-004	-1.7973e-002			

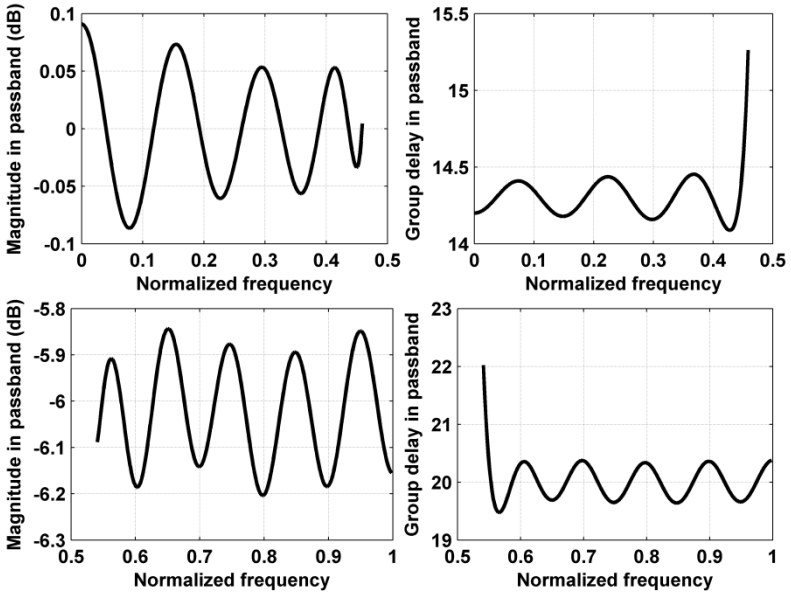

Fig. 4.7. Magnitude and group delay responses of IIR filter designed in Example 3.

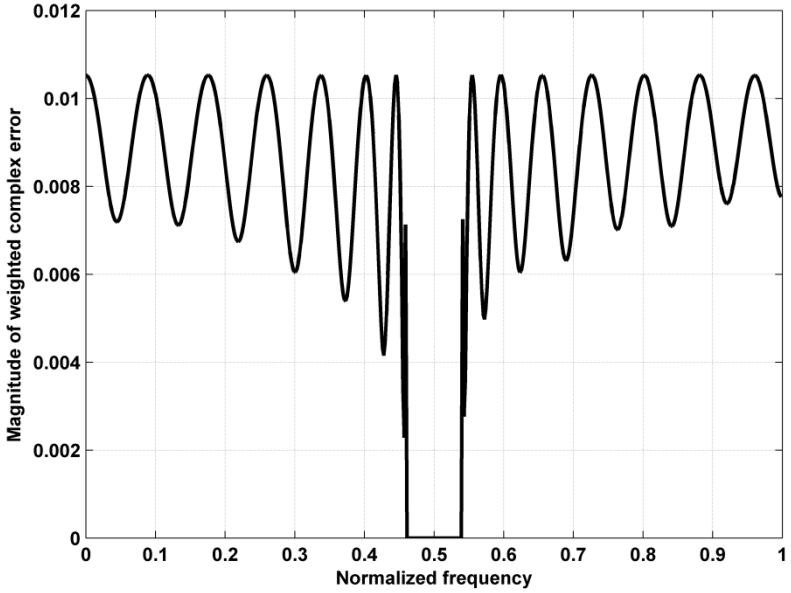

Fig. 4.8. Magnitude of weighted complex error of IIR filter designed in Example 3.

4.3.4 Example 4

The last example is to design a full-band differentiator [32]. The ideal frequency response is given by

$$D(\omega) = \frac{\omega}{\pi} e^{j[0.5\pi - (\tau_s + 0.5)\omega]}, \qquad 0 \le \omega \le \pi$$

where τ_s assumes an integer value. In the argument of $D(\omega)$ defined above, a half of sample delay is added to eliminate the discontinuity of the desired phase response [32]. Filter orders are chosen as $N = M = 17$. In this example, we adopt a variable regularization parameter $\zeta^{(k)}$ in (4.22). At the kth iteration, $\zeta^{(k)}$ is chosen as $\zeta^{(k)} = 0.01k$ for $k \ge 1$ and used in (4.22) to determine the search direction $\Delta \boldsymbol{x}^{(k)}$ and $\Delta \boldsymbol{d}^{(k)}$. Naturally, there are some other ways to select the variable $\zeta^{(k)}$ during the sequential procedure. We change the integer group delay τ_s from 8 to 17. The best design can be attained when τ_s is equal to 15. After 19 iterations, the sequential design procedure converges to the final solution. The maximum pole radius of the obtained IIR differentiator is 0.9635. All the numerator and denominator coefficients are listed in Table 4.7. The design characteristics and the approximation errors of magnitude and group delay responses are shown in Fig. 4.9. It can be seen that near the origin of the frequency axis, the group delay (or phase response) of the designed IIR differentiator has a large error. This is mainly because we use the absolute error in this design to construct the objective function. In practice, a better way to design differentiators is to adopt a relative or normalized error as the objective function [2]. Since the magnitude responses on the frequencies near the origin are almost equal to zero, the overall approximation errors on these frequencies are still quite small. This can be verified by the magnitude of the complex error, *i.e.*, $|E(\omega)|$, which is shown in Fig. 4.10. Therefore, when computing the error measurements of group delay listed in Table 4.8, the approximation errors of group delay within $[0, 0.01\pi]$ are neglected.

For comparison, we also design a group of IIR differentiators using the LP method [32] under the same set of specifications. The best design result can be attained when the integer group delay τ_s is equal to 14. The corresponding filter coefficients have

been reported in [32]. The maximum pole radius of this best IIR differentiator is 0.9821. All the error measurements are also summarized in Table 4.8. It can be observed that the LP method can achieve better group delay responses, whereas the proposed design method can obtain much better magnitude responses and much lower minimax approximation error.

Table 4.7 Filter Coefficients (p_0 to p_N and q_0 to q_M) of IIR Digital Differentiator (τ_s = 15) Designed in Example 4

$p_0 \sim p_4$	-3.0503e-003	-2.7413e-003	-3.8702e-004	-1.8694e-006	-2.0119e-004
$p_5 \sim p_9$	1.2524e-004	-2.3072e-004	3.4058e-004	-3.7793e-004	5.8881e-004
$p_{10} \sim p_{14}$	-9.4826e-004	1.7485e-003	-3.1625e-003	7.9066e-003	-2.8664e-002
$p_{15} \sim p_{17}$	3.5891e-001	1.8208e-002	-3.5417e-001		
$q_0 \sim q_4$	1.0000e+000	1.0531e+000	6.6768e-002	-1.0020e-002	3.7019e-003
$q_5 \sim q_9$	-1.7434e-003	7.5298e-004	-5.1406e-004	3.2175e-004	-3.5136e-004
$q_{10} \sim q_{14}$	2.4046e-005	-1.6413e-004	1.0602e-005	-1.7735e-004	-1.7353e-004
$q_{15} \sim q_{17}$	-6.6872e-004	-1.1545e-003	-8.5683e-004		

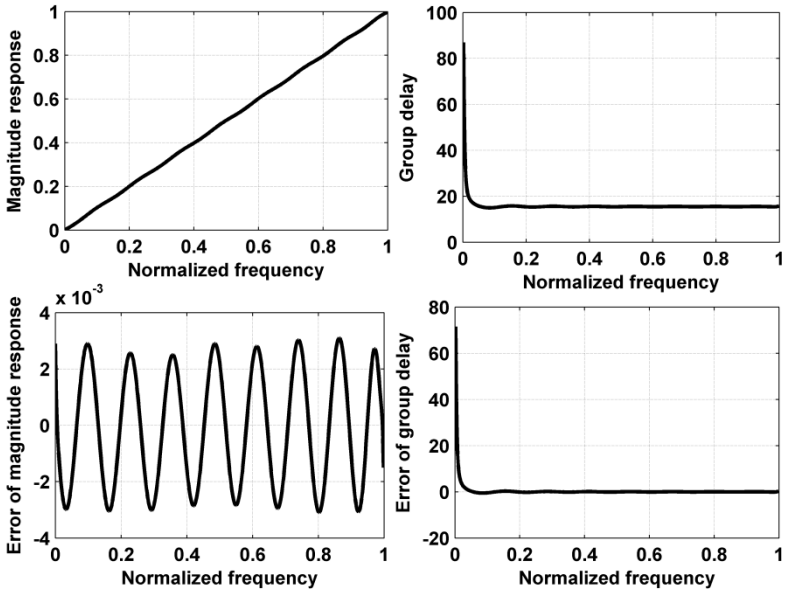

Fig. 4.9. Design characteristics and errors of IIR differentiator designed in Example 4.

Table 4.8 Error Measurements of Design Results in Example 4

Method	τ_s	Minimax Error E_{MM} (in dB)	MAG (Peak/L_2 in dB)	GD (Peak/ L_2)
Proposed	15	-50.102	-50.176/ -53.769	9.877/ 6.818e-1
LP [32]	14	-30.298	-30.636/ -51.783	3.585e-1/ 2.337e-2

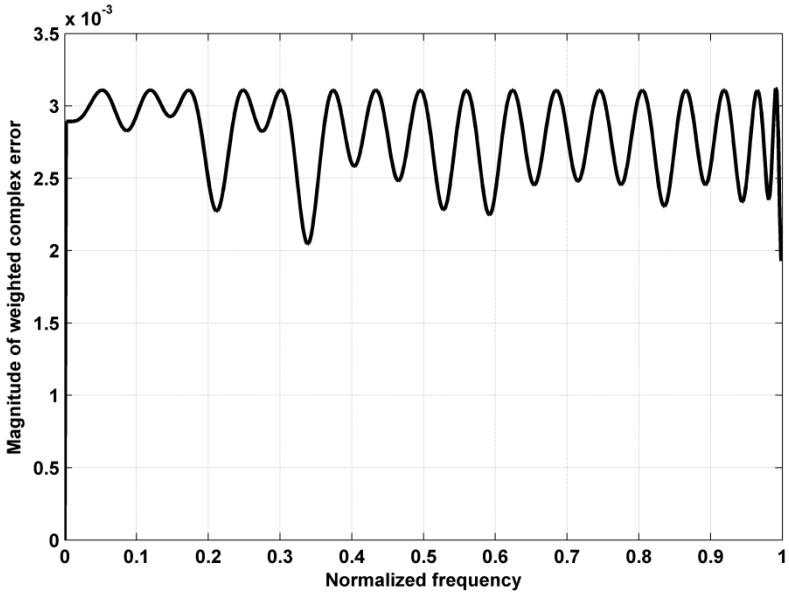

Fig. 4.10. Magnitude of weighted complex error of IIR differentiator designed in Example 4.

CHAPTER V
MINIMAX DESIGN OF IIR DIGITAL FILTERS USING SDP RELAXATION TECHNIQUE

Since IIR filter design problems are nonconvex, there are many local minima on error performance surfaces. From an initial point, using various local optimization methods, we can find a local optimum near the initial point. However, for nonconvex design problems, it is hard to guarantee that global solutions can be definitely obtained. On the other hand, even if a global solution were achieved, in practice it could be difficult or impossible to confirm that it was indeed the global solution. This difficulty, however, could be mitigated, to some extent, in the framework of convex optimization. In this chapter, a new design method will be proposed for minimax IIR filter designs. Using the SDP relaxation technique, the original design problem can be transformed to an SDP feasibility problem, which will be solved sequentially in a bisection search procedure. A sufficient condition for optimal designs can be derived from the proposed design method.

This chapter is organized as follows. In Section 5.1, a bisection search procedure is first introduced. Then, the SDP relaxation technique is applied to formulate a feasibility problem. A trace heuristic approximation method is presented later in Section 5.1 to achieve real minimax solutions. The stability of designed IIR filters can be ensured by a monitoring strategy, which is finally described in Section 5.1. Several numerical examples are presented in Section 5.2 to demonstrate the effectiveness of the proposed design method.

5.1 Minimax Design Method

5.1.1 Bisection Search Procedure

Instead of trying to find the minimum (squared) error limit δ^* by directly minimizing the error limit δ in (4.2), a bisection search procedure is employed in the

proposed design method. At each iteration, a fixed error limit $\delta^{(k)}$ is used to reformulate the constraint (4.2.b). The major steps of the bisection search procedure are shown below:

Step 1. Given a set of design specifications, set $k = 0$, and then estimate the initial upper bound $\delta_{+}^{(0)}$ and lower bound $\delta_{-}^{(0)}$ for the minimum error limit δ^{*}.

Step 2. Set $k = k+1$, and choose $\delta^{(k)} = \sqrt{\delta_{+}^{(k-1)} \cdot \delta_{-}^{(k-1)}}$, *i.e.*, the geometric mean of $\delta_{+}^{(k-1)}$ and $\delta_{-}^{(k-1)}$. Then, solve a feasibility problem, where the original constraint (4.2.b) is recast with the fixed error limit $\delta^{(k)}$. If a feasible solution is found, which indicates $\delta_{+}^{(k-1)} \geq \delta^{(k)} \geq \delta^{*} \geq \delta_{-}^{(k-1)}$, then choose the new upper and lower bounds as $\delta_{+}^{(k)} = \delta^{(k)}$ and $\delta_{-}^{(k)} = \delta_{-}^{(k-1)}$. On the contrary, if no feasible solution exists, which means $\delta_{+}^{(k-1)} \geq \delta^{*} > \delta^{(k)} \geq \delta_{-}^{(k-1)}$, then choose the new upper and lower bounds as $\delta_{-}^{(k)} = \delta^{(k)}$ and $\delta_{+}^{(k)} = \delta_{+}^{(k-1)}$. The formulation of the feasibility problem will be presented later.

Step 3. If a predetermined accuracy criterion of locating δ^{*} is satisfied, terminate the bisection search procedure. Otherwise, go to Step 2 and continue.

Several remarks on the bisection search procedure described above are made here:

1. This bisection search procedure is different from the usual bisection search procedure, where $\delta^{(k)}$ is chosen as the arithmetic mean of $\delta_{+}^{(k-1)}$ and $\delta_{-}^{(k-1)}$, *i.e.*, $\delta^{(k)} = 0.5\left[\delta_{+}^{(k-1)} + \delta_{-}^{(k-1)}\right]$. When δ^{*} is small, choosing $\delta^{(k)}$ as the geometric mean instead of the arithmetic mean can result in a smaller number of iterations required to achieve relative accuracy in locating δ^{*} [6]. Actually, the bisection search procedure presented above is performed with the arithmetic mean of $\log_{10}\delta_{+}^{(k)}$ and $\log_{10}\delta_{-}^{(k)}$.
2. The bisection search procedure will be terminated, if the following condition is satisfied:

$$\frac{\delta_+^{(k)} - \delta_-^{(k)}}{\delta_-^{(k)}} \leq \kappa_{\min} \tag{5.1}$$

where $\kappa_{\min} > 0$ is a prescribed small number. Let T_o denote the total number of iterations and it can be verified that

$$T_o \leq \left\lfloor \log_2 \left(\frac{\log_{10} \delta_+^{(0)} - \log_{10} \delta_-^{(0)}}{\log_{10}(1 + \kappa_{\min})} \right) \right\rfloor + 1 \tag{5.2}$$

For convenience of the latter discussion, we assume that $\kappa_{\min}$ can be chosen arbitrarily small so as to accurately locate the minimum error limit δ^*.

3. Normally, the initial upper and lower bounds of δ^* can be arbitrarily selected as long as the condition $0 < \delta_-^{(0)} \leq \delta^* \leq \delta_+^{(0)}$ is satisfied. However, it can be observed from (5.2) that T_o could be reduced if $\delta_-^{(0)}$ and $\delta_+^{(0)}$ are closer to each other. As an attempt to obtain a lower $\delta_+^{(0)}$, we first utilize the LP method [32] to design an IIR filter under the given specifications. Then, $\delta_+^{(0)}$ can be chosen as the squared error limit, *i.e.*, $\max_{\omega \in \Omega_I} |E(\omega)|^2$, of the obtained IIR filter. Some other design methods can also be deployed here to achieve smaller $\delta_+^{(0)}$ for δ^*. In order to obtain a reasonable $\delta_-^{(0)}$ for δ^*, we utilize the SDP relaxation technique to convert the nonconvex constraint (4.2.b) into a convex form. With the relaxed constraint, the design problem can be solved by directly minimizing δ. Since the feasible set defined by the relaxed constraint is larger than that of (4.2.b), we always have $\delta_{rel}^* \leq \delta^*$, where δ_{rel}^* denotes the optimal value of the relaxed design problem. Then, δ_{rel}^* can be chosen as the lower bound $\delta_-^{(0)}$. The formulation of such a relaxed design problem using the SDP relaxation technique is to be presented at the end of this section.

5.1.2 Formulation of Feasibility Problem Using SDP Relaxation Technique

In this section, we will construct a feasibility problem, in which the nonconvex constraint (4.2.b) is transformed to a convex form using the SDP relaxation technique.

This feasibility problem will be solved in Step 2 of the bisection search procedure described earlier. The feasibility problem will be first formulated without any stability constraint. The stability issue will be considered later in this section.

In practice, the constraint (4.2.b) can be imposed on a set of discrete frequency points, *i.e.*, $\omega_i \in \Omega_I$ for $i = 0, 1, ..., L$. At the kth iteration of the bisection search procedure, given the error limit $\delta^{(k)}$, the constraint (4.2.b) can be rewritten by

$$\begin{aligned} &\left|W(\omega_i)\left[D(\omega_i)Q\left(e^{j\omega_i}\right) - P\left(e^{j\omega_i}\right)\right]\right|^2 \\ &= W^2(\omega_i)[|D(\omega_i)|^2 + 2\mathrm{Re}\{D(\omega_i)\boldsymbol{c}^H(\omega_i)\}\overline{\boldsymbol{x}} + \overline{\boldsymbol{x}}^T\boldsymbol{A}(\omega_i)\overline{\boldsymbol{x}}] \\ &\le \delta^{(k)} \cdot \left|Q\left(e^{j\omega_i}\right)\right|^2 \\ &= \delta^{(k)} \cdot \left[1 + 2\mathrm{Re}\left\{\overline{\boldsymbol{\varphi}}_M^H\left(e^{j\omega_i}\right)\right\}\overline{\boldsymbol{q}} + \overline{\boldsymbol{q}}^T\boldsymbol{B}(\omega_i)\overline{\boldsymbol{q}}\right] \\ &\quad \omega_i \in \Omega_I, \quad i = 0, 1, \dots, L \end{aligned} \tag{5.3}$$

where

$$\overline{\boldsymbol{q}} = [q_1 \quad q_2 \quad \dots \quad q_M]^T \tag{5.4}$$

$$\overline{\boldsymbol{x}} = [\overline{\boldsymbol{q}}^T \quad \boldsymbol{p}^T]^T \tag{5.5}$$

$$\overline{\boldsymbol{\varphi}}_M(z) = [z^{-1} \quad z^{-2} \quad \dots \quad z^{-M}]^T \tag{5.6}$$

$$\boldsymbol{c}(\omega) = \begin{bmatrix} D(\omega)\overline{\boldsymbol{\varphi}}_M\left(e^{j\omega}\right) \\ -\boldsymbol{\varphi}_N\left(e^{j\omega}\right) \end{bmatrix} \tag{5.7}$$

$$\boldsymbol{A}(\omega) = \mathrm{Re}\{\boldsymbol{c}(\omega)\boldsymbol{c}^H(\omega)\} \tag{5.8}$$

$$\boldsymbol{B}(\omega) = \mathrm{Re}\left\{\overline{\boldsymbol{\varphi}}_M\left(e^{j\omega}\right)\overline{\boldsymbol{\varphi}}_M^H\left(e^{j\omega}\right)\right\} \tag{5.9}$$

It is noteworthy that now the constraint (5.3) is formulated in terms of $\overline{\boldsymbol{x}}$ instead of $\boldsymbol{x}$ in (4.2.b). Since the first denominator coefficient q_0 (namely, x_0) is always chosen equal to 1, the constraint formulated by (5.3) is still equivalent to (4.2.b). Although the terms on both sides of (5.3) are convex quadratic functions of $\overline{\boldsymbol{x}}$, it is difficult to directly transform

(5.3) into an equivalent convex constraint. Here, a symmetric matrix is introduced in order to further simplify (5.3)

$$\boldsymbol{X} = \overline{\boldsymbol{x}}\overline{\boldsymbol{x}}^T = \begin{bmatrix} \overline{\boldsymbol{q}}\overline{\boldsymbol{q}}^T & \overline{\boldsymbol{q}}\boldsymbol{p}^T \\ \boldsymbol{p}\overline{\boldsymbol{q}}^T & \boldsymbol{p}\boldsymbol{p}^T \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_q & \boldsymbol{X}_{q,p} \\ \boldsymbol{X}_{q,p}^T & \boldsymbol{X}_p \end{bmatrix} \tag{5.10}$$

Substituting $\boldsymbol{X}$ into (5.3) for the quadratic terms of $\overline{\boldsymbol{x}}$, we can rewrite the constraint (5.3) in a matrix form:

$$\begin{aligned} &|D(\omega_i)|^2 + 2\mathrm{Re}\{D(\omega_i)\boldsymbol{c}^H(\omega_i)\}\overline{\boldsymbol{x}} + \mathrm{Tr}\{\boldsymbol{X}\boldsymbol{A}(\omega_i)\} \\ &\le \frac{\delta^{(k)}}{W^2(\omega_i)} \cdot \left[1 + 2\mathrm{Re}\left\{\overline{\boldsymbol{\varphi}}_M^H\left(e^{j\omega_i}\right)\right\}\overline{\boldsymbol{q}} + \mathrm{Tr}\left\{\boldsymbol{X}_q\boldsymbol{B}(\omega_i)\right\}\right] \\ &\quad \omega_i \in \Omega_I, \quad i = 0, 1, \dots, L \end{aligned} \tag{5.11}$$

where Tr{·} denotes the trace of a matrix. By introducing $\boldsymbol{X}$, the original nonconvex constraint (4.2.b) is transformed into a linear inequality constraint in terms of $\overline{\boldsymbol{x}}$ and $\boldsymbol{X}$.

By combining (5.10) and (5.11), we can construct a feasibility problem as

$$\min \quad z \tag{5.12}$$

$$\begin{aligned} \text{s.t.} \quad &|D(\omega_i)|^2 + 2\mathrm{Re}\{D(\omega_i)\boldsymbol{c}^H(\omega_i)\}\overline{\boldsymbol{x}} + \mathrm{Tr}\{\boldsymbol{X}\boldsymbol{A}(\omega_i)\} \\ &\le \frac{\delta^{(k)}}{W^2(\omega_i)} \cdot \left[1 + 2\mathrm{Re}\left\{\overline{\boldsymbol{\varphi}}_M^H\left(e^{j\omega_i}\right)\right\}\overline{\boldsymbol{q}} + \mathrm{Tr}\left\{\boldsymbol{X}_q\boldsymbol{B}(\omega_i)\right\}\right] + z \\ &\quad \omega_i \in \Omega_I, \quad i = 0, 1, \dots, L \end{aligned} \tag{5.12.a}$$

$$\boldsymbol{X} = \overline{\boldsymbol{x}}\overline{\boldsymbol{x}}^T = \begin{bmatrix} \boldsymbol{X}_q & \boldsymbol{X}_{q,p} \\ \boldsymbol{X}_{q,p}^T & \boldsymbol{X}_p \end{bmatrix} \text{ where } \overline{\boldsymbol{x}} = \begin{bmatrix} \overline{\boldsymbol{q}} \\ \boldsymbol{p} \end{bmatrix} \tag{5.12.b}$$

An auxiliary variable z is introduced into (5.12). It can be verified that a feasible solution ($\overline{\boldsymbol{x}}$, $\boldsymbol{X}$) exists under the constraints (5.10) and (5.11) if and only if the minimum value of z obtained by solving (5.12) is less than or equal to 0. Then, the upper bound $\delta_+^{(k)}$ can be replaced by $\delta^{(k)}$, and taken into the next iteration of the bisection search procedure. On the contrary, if the minimum value of z is larger than 0, which means given $\delta^{(k)}$ the

constraints (5.10) and (5.11) cannot be simultaneously satisfied. Then, in Step 2 of the bisection search procedure, the lower bound $\delta_{-}^{(k)}$ will be replaced by $\delta^{(k)}$, and taken into the next iteration to determine $\delta^{(k+1)}$.

There is an obstacle to solve the feasibility problem (5.12). The matrix equality constraint (5.10) is nonconvex. In order to overcome this obstacle, we relax (5.10) as $\boldsymbol{X} \succeq \overline{\boldsymbol{x}}\overline{\boldsymbol{x}}^T$, which represents $\boldsymbol{X} - \overline{\boldsymbol{x}}\overline{\boldsymbol{x}}^T$ is a positive semi-definite (PSD) matrix. The relaxed constraint $\boldsymbol{X} \succeq \overline{\boldsymbol{x}}\overline{\boldsymbol{x}}^T$ is equivalent to [72]

$$\boldsymbol{Z} = \begin{bmatrix} 1 & \overline{\boldsymbol{x}}^T \\ \overline{\boldsymbol{x}} & \boldsymbol{X} \end{bmatrix} \succeq 0 \tag{5.13}$$

Then, the feasibility problem (5.12) can be recast as

$$\min \quad z \tag{5.14}$$

$$\text{s.t.} \quad |D(\omega_i)|^2 + 2\text{Re}\{D(\omega_i)\boldsymbol{c}^H(\omega_i)\}\overline{\boldsymbol{x}} + \text{Tr}\{\boldsymbol{X}\boldsymbol{A}(\omega_i)\}$$
$$\leq \frac{\delta^{(k)}}{W^2(\omega_i)} \cdot \left[1 + 2\text{Re}\left\{\overline{\boldsymbol{\varphi}}_M^H\left(e^{j\omega_i}\right)\right\}\overline{\boldsymbol{q}} + \text{Tr}\left\{\boldsymbol{X}_q\boldsymbol{B}(\omega_i)\right\}\right] + z \tag{5.14.a}$$
$$\omega_i \in \Omega_I, \quad i = 0, 1, \ldots, L$$

$$\boldsymbol{Z} = \begin{bmatrix} 1 & \overline{\boldsymbol{x}}^T \\ \overline{\boldsymbol{x}} & \boldsymbol{X} \end{bmatrix} \succeq 0 \text{ where } \boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_q & \boldsymbol{X}_{q,p} \\ \boldsymbol{X}_{q,p}^T & \boldsymbol{X}_p \end{bmatrix} \text{ and } \overline{\boldsymbol{x}} = \begin{bmatrix} \overline{\boldsymbol{q}} \\ \boldsymbol{p} \end{bmatrix} \tag{5.14.b}$$

Now (5.14.a) is a linear inequality constraint in terms the elements of $\boldsymbol{Z}$ and the auxiliary variable z. Compared with (5.12.b), the constraint (5.14.b) defines a larger feasible set. Thus, for a given $\delta^{(k)}$, if a feasible solution ($\overline{\boldsymbol{x}}$, $\boldsymbol{X}$) exists for (5.12), by taking ($\overline{\boldsymbol{x}}$, $\boldsymbol{X}$) into (5.14.a) and (5.14.b), it can be verified that the relaxed feasibility problem (5.14) also has a feasible solution $\boldsymbol{Z}$, and the corresponding minimum value of z is definitely less than or equal to 0. It should be mentioned that even if a feasible solution $\boldsymbol{Z}$ with $z \leq 0$ exists for (5.14), there is no guarantee that the original feasibility problem (5.12) also has a feasible solution ($\overline{\boldsymbol{x}}$, $\boldsymbol{X}$). On the contrary, if the minimum value of z for (5.14) is greater than 0, it implies that there is no feasible solution $\boldsymbol{Z}$ satisfying both the linear inequality constraint (5.11) and the relaxed LMI constraint (5.14.b). Accordingly, the original feasibility

problem (5.12) does not have a feasible solution ($\bar{\boldsymbol{x}}$, $\boldsymbol{X}$) for the given error limit $\delta^{(k)}$. However, even though there is no feasible solution existing for (5.12), the relaxed feasibility problem (5.14) may still have a feasible solution $\boldsymbol{Z}$ with $z \leq 0$. Here, it should be emphasized that if the rank of $\boldsymbol{Z}$ obtained by solving (5.14) is equal to 1, the relaxed constraint (5.14.b) is reduced to (5.12.b). Then, the feasibility problems (5.12) and (5.14) are equivalent to each other.

Combined with the bisection search procedure described earlier, relaxed feasibility problems (5.14) with different $\delta^{(k)}$ are sequentially solved. Based on the analysis above, we arrive at the following sufficient condition for the optimal solution of the original design problem:

Proposition 1: Let $\widetilde{\boldsymbol{Z}}$ or, equivalently, $(\widetilde{\boldsymbol{x}},\ \widetilde{\boldsymbol{X}})$ be the final output of the bisection search procedure, in which the relaxed feasibility problem (5.14) is solved at each iteration. The corresponding final error limit is denoted by $\tilde{\delta}$. Then, $\tilde{\delta}$ is equal to δ^*, and $\boldsymbol{x}_{opt} = [1\ \widetilde{\boldsymbol{x}}^T]^T$ is the optimal solution of the minimax design problem (4.1), if the rank of $\widetilde{\boldsymbol{Z}}$ is equal to 1.

Proof: Suppose that the rank of $\widetilde{\boldsymbol{Z}}$ is equal to 1. Then, $\widetilde{\boldsymbol{x}}$ and $\widetilde{\boldsymbol{X}}$ satisfy the equality constraint (5.12.b), and $\widetilde{\boldsymbol{x}}$ is a minimax solution to the original design problem (4.1). On the other hand, from the discussion earlier, it follows that by successively solving the relaxed feasibility problem (5.14), we can find a lower bound of δ^*, *i.e.*, $\tilde{\delta} \leq \delta^*$. Suppose that $\tilde{\delta} < \delta^*$, which means that we could find another solution, which can achieve a lower minimum error limit than δ^*. However, it contradicts the assumption that δ^* is the minimum error limit of the original minimax design problem (4.1). Therefore, $\tilde{\delta}$ should be equal to δ^*. Accordingly, $\boldsymbol{x}_{opt} = [1\ \widetilde{\boldsymbol{x}}^T]^T$ is the optimal solution of (4.1).

□

This proposition implies that if we can find a rank-1 solution using the bisection search procedure, then it is the optimal solution of the original design problem indeed. Example 1 will be presented in Section 5.2 to demonstrate the capability of the proposed bisection search procedure to achieve optimal designs. However, rank-1 solutions cannot always be attained, especially when the denominator order M is large and/or the design

specifications are stringent. Furthermore, the stability issue has not been taken into account during the bisection search procedure. It is known that when $M > 2$, the stability domain cannot be strictly expressed as a convex set with respect to denominator coefficients q_m. On the other hand, when the obtained $\boldsymbol{Z}$ has a rank higher than 1, the corresponding solution $\boldsymbol{x} = [1\ \overline{\boldsymbol{x}}^T]^T$ is not a real minimax design. This problem will be addressed in the next section.

5.1.3 SDP Formulation Using Trace Heuristic Approximation

In order to obtain a rank-1 solution, we can constrain the rank of $\boldsymbol{Z}$ equal to 1 in the relaxed feasibility problem (5.14) during the bisection search procedure. However, in general, the rank constraint is nonconvex, and incorporating it could make the feasibility problem computationally intractable. Here, we employ a trace heuristic method [73] to approximate the design problem with the rank constraint. This approximation technique is based on the observation that the rank of the PSD matrix $\boldsymbol{Z}$, represented by $\operatorname{rank}\boldsymbol{Z}$, can be expressed by

$$\operatorname{rank}\boldsymbol{Z} = \sum_{i=1}^{N+M+2} I_0(\lambda_i(\boldsymbol{Z})) \tag{5.15}$$

where $\lambda_i(\boldsymbol{Z})$ (i = 1, 2, ..., N+M+2) denote the real eigenvalues of the symmetric matrix $\boldsymbol{Z}$. Without loss of generalization, we can assume that $\lambda_i(\boldsymbol{Z})$ are arranged in a non-ascending order, *i.e.*, $\lambda_1(\boldsymbol{Z}) \geq \lambda_2(\boldsymbol{Z}) \geq \cdots \geq \lambda_{N+M+2}(\boldsymbol{Z})$. In (5.15), $I_0(x)$ is an indicator function which is defined by

$$I_0(x) = \begin{cases} 1 & x > 0 \\ 0 & x \leq 0 \end{cases} \tag{5.16}$$

Then, we approximate the indicator function $I_0(x)$ by x in (5.15), and incorporate a regularization term $\mathrm{Tr}\{\boldsymbol{Z}\} = \sum_i \lambda_i(\boldsymbol{Z})$ into the objective function of (5.14). Since $\boldsymbol{Z}$ is PSD, $\mathrm{Tr}\{\boldsymbol{Z}\}$ equals $\|\boldsymbol{\Lambda}(\boldsymbol{Z})\|_1 = \sum_i |\lambda_i(\boldsymbol{Z})| = \sum_i \lambda_i(\boldsymbol{Z})$, where $\|\cdot\|_1$ denotes the l_1-norm of a vector and $\boldsymbol{\Lambda}(\boldsymbol{Z}) = [\lambda_1(\boldsymbol{Z})\ \lambda_2(\boldsymbol{Z})\ \ldots\ \lambda_{N+M+2}(\boldsymbol{Z})]^T$. If the regularization coefficient is sufficiently large, it is known that by minimizing $\|\boldsymbol{\Lambda}(\boldsymbol{Z})\|_1$ some components of $\boldsymbol{\Lambda}(\boldsymbol{Z})$

will be driven to zero, leading the final $\boldsymbol{\Lambda}(\boldsymbol{Z})$ to a sparse vector [51]. This means that minimizing $\mathrm{Tr}\{\boldsymbol{Z}\}$ renders many of the eigenvalues of $\boldsymbol{Z}$ as zeros, resulting in a low-rank matrix. When the rank of $\boldsymbol{Z}$ is close to 1, we have $\mathrm{Tr}\{\boldsymbol{Z}\} \approx \lambda_1(\boldsymbol{Z}) \approx 1+\|\overline{\boldsymbol{x}}\|_2^2$. Therefore, in practice, the regularization coefficient cannot be too large. Otherwise, $\mathrm{Tr}\{\boldsymbol{Z}\}$ could be over-attenuated and, accordingly, $\|\overline{\boldsymbol{x}}\|_2$ becomes too small to achieve a satisfactory design.

We modify the objective function of (5.14) as the weighted sum of $\mathrm{Tr}\{\boldsymbol{Z}\}$ and z. Then, the relaxed feasibility problem (5.14) is modified as

$$\min \quad \alpha \mathrm{Tr}\{\boldsymbol{Z}\} + (1-\alpha)z \tag{5.17}$$

$$\begin{aligned} \text{s.t.} \quad & |D(\omega_i)|^2 + 2\mathrm{Re}\{D(\omega_i)\boldsymbol{c}^H(\omega_i)\}\overline{\boldsymbol{x}} + \mathrm{Tr}\{\boldsymbol{X}\boldsymbol{A}(\omega_i)\} \\ & \le \frac{\delta^{(k)}}{W^2(\omega_i)} \cdot \left[1 + 2\mathrm{Re}\left\{\overline{\boldsymbol{\varphi}}_M^H\left(e^{j\omega_i}\right)\right\}\overline{\boldsymbol{q}} + \mathrm{Tr}\left\{\boldsymbol{X}_q \boldsymbol{B}(\omega_i)\right\}\right] + z \qquad (5.17.a) \\ & \omega_i \in \Omega_I, \qquad i = 0,1,\ldots,L \end{aligned}$$

$$\boldsymbol{Z} = \begin{bmatrix} 1 & \overline{\boldsymbol{x}}^T \\ \overline{\boldsymbol{x}} & \boldsymbol{X} \end{bmatrix} \succeq 0 \text{ where } \boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_q & \boldsymbol{X}_{q,p} \\ \boldsymbol{X}_{q,p}^T & \boldsymbol{X}_p \end{bmatrix} \text{ and } \overline{\boldsymbol{x}} = \begin{bmatrix} \overline{\boldsymbol{q}} \\ \boldsymbol{p} \end{bmatrix} \tag{5.17.b}$$

where $0 \le \alpha \le 1$. When $\alpha = 0$, the regularized feasibility problem (5.17) is reduced to (5.14). The regularization coefficient α should be chosen as small as possible so as to best approximate the relaxed feasibility problem (5.14) as well as avoid $\mathrm{Tr}\{\boldsymbol{Z}\}$ being over-attenuated. In order to determine an appropriate value for α, another bisection search procedure is introduced. Note that for the complete method, there are two nested bisection search procedures. They play different roles in the proposed design method. The outer bisection search procedure is used to locate the minimum error limit δ^*. Given a fixed error limit $\delta^{(k)}$ at the kth outer iteration, the inner bisection search procedure is invoked to find an appropriate α to make the rank of the obtained matrix $\boldsymbol{Z}$ close to 1. The inner bisection search procedure can also be used to restrict all the poles' positions for stability, which will be discussed in detail later. For clarity, in the following, we use l to represent the iteration index of the inner bisection search procedure, while k for the outer bisection search procedure. Accordingly, z, α, and $\boldsymbol{Z}$ in (5.17) are replaced by

$z^{(k,l)}$, $\alpha^{(k,l)}$, and $\boldsymbol{Z}^{(k,l)}$ at the iteration step (k, l), respectively. The major steps of the inner bisection search procedure are shown below:

Step 1. Given $\delta^{(k)}$, set $l = 0$, and then choose the initial upper bound $\alpha_{+}^{(k,0)}$ and lower bound $\alpha_{-}^{(k,0)}$, respectively.

Step 2. Set $l = l+1$, and choose $\alpha^{(k,l)} = \sqrt{\alpha_{+}^{(k,l-1)} \cdot \alpha_{-}^{(k,l-1)}}$. Using $\delta^{(k)}$ and $\alpha^{(k,l)}$, solve the regularized feasibility problem (5.17). If the ratio

$$\eta\left(\boldsymbol{Z}^{(k,l)}\right) = \frac{\lambda_2\left(\boldsymbol{Z}^{(k,l)}\right)}{\lambda_1\left(\boldsymbol{Z}^{(k,l)}\right)} < \varepsilon \tag{5.18}$$

set $\alpha_{+}^{(k,l)} = \alpha^{(k,l)}$ and $\alpha_{-}^{(k,l)} = \alpha_{-}^{(k,l-1)}$. Otherwise, choose $\alpha_{-}^{(k,l)} = \alpha^{(k,l)}$ and $\alpha_{+}^{(k,l)} = \alpha_{+}^{(k,l-1)}$.

Step 3. If the predetermined accuracy of locating the minimum value of α is satisfied, terminate the inner bisection search procedure. Otherwise, go to Step 2 and continue.

Some remarks regarding the inner bisection search procedure are made below:

1. In practice, we use $\eta\left(\boldsymbol{Z}^{(k,l)}\right) < \varepsilon$ to replace the condition that the rank of $\boldsymbol{Z}^{(k,l)}$ is equal to 1. Here, $\lambda_1\left(\boldsymbol{Z}^{(k,l)}\right)$ and $\lambda_2\left(\boldsymbol{Z}^{(k,l)}\right)$ denote the first and second largest eigenvalues of $\boldsymbol{Z}^{(k,l)}$, and parameter $\varepsilon > 0$ represents a pre-specified small positive value.
2. Before the inner bisection search procedure, the relaxed feasibility problem (5.14) should be solved first. Let $(z^{(k,0)}, \boldsymbol{Z}^{(k,0)})$ denote the result obtained from (5.14). If $z^{(k,0)} > 0$, which means there is no feasible solution for the relaxed feasibility problem (5.14), then new upper and lower bounds of δ^* are appropriately selected and the design program can directly go to Step 3 of the outer bisection search procedure. If $z^{(k,0)} \leq 0$ and $\eta\left(\boldsymbol{Z}^{(k,l)}\right) \geq \varepsilon$, the inner bisection search procedure will be triggered.

3. The inner bisection search procedure continues until the following condition is satisfied

$$\frac{\alpha_{+}^{(k,l)} - \alpha_{-}^{(k,l)}}{\alpha_{-}^{(k,l)}} \leq \kappa_{\min} \tag{5.19}$$

Like the outer bisection search procedure, the convergence of the inner bisection search procedure can be guaranteed. Let $T_i(k)$ (k = 1, 2, ..., T_o) represent the total number of the inner iterations at the kth outer iteration. Similar to (5.2), we have

$$T_i(k) \leq \left\lfloor \log_2 \left(\frac{\log_{10} \alpha_{+}^{(k,0)} - \log_{10} \alpha_{-}^{(k,0)}}{\log_{10}(1+\kappa_{\min})} \right) \right\rfloor + 1 \tag{5.20}$$

4. The initial upper and lower bounds of α can be arbitrarily selected as long as the condition $0 < \alpha_{-}^{(k,0)} \leq \alpha_{+}^{(k,0)} \leq 1$ is satisfied. In order to reduce the total number of the inner iterations, in our design the initial upper and lower bounds of α at the kth outer iteration are chosen as

$$\alpha_{+}^{(k,0)} = \gamma \alpha_{+}^{(k',T_i(k'))}, \qquad 1 \leq k' < k \leq T_o \tag{5.21}$$

$$\alpha_{-}^{(k,0)} = \gamma^{-1} \alpha_{-}^{(k',T_i(k'))}, \qquad 1 \leq k' < k \leq T_o \tag{5.22}$$

where $\gamma > 1$, and $\alpha_{+}^{(k',T_i(k'))}$ and $\alpha_{-}^{(k',T_i(k'))}$ denote the final upper and lower bounds of α determined by the inner bisection search procedure at the k'th outer iteration. Obviously, the search range of α can be extended by increasing γ. For the first time the inner bisection search procedure is invoked, the initial upper bound $\alpha_{+}^{(k_0,0)}$ and lower bound $\alpha_{-}^{(k_0,0)}$ ($k_0 \geq 1$) should be specified by designers. Since there is no prior information to determine them, normally we can choose $\alpha_{+}^{(k_0,0)}$ and $\alpha_{-}^{(k_0,0)}$ close to 1 and 0, respectively.

5. So far, it has not been strictly proved that there always exists some α for which the rank of $\boldsymbol{Z}$ is equal to 1. Nevertheless, in the extreme situation when α = 1, the constraint (5.17.a) can always be satisfied, because z can be arbitrarily selected

without any influence on the objective function of (5.17). Then, it can be deduced from (5.17) that the rank of $\boldsymbol{Z}$ should be equal to 1, and all eigenvalues are equal to 0 except $\lambda_1(\boldsymbol{Z}) = 1$. Thus, in practice, we can assume that when α is large enough, the rank of the final output $\boldsymbol{Z}$ is close to 1.

6. Since the regularization term $\mathrm{Tr}\{\boldsymbol{Z}\}$ is incorporated in the objective function of (5.17), even if the rank of the final output $\boldsymbol{Z}$ is equal to 1, it cannot be concluded that the optimal solution is attained. However, as the minimum value of α determined by the inner bisection search procedure is small enough, the regularized feasibility problem (5.17) can serve as a good approximation of the relaxed feasibility problem (5.14).

5.1.4 Stability Issue

So far, the proposed design method cannot definitely ensure the stability of designed IIR filters. Therefore, stability constraints need to be incorporated in the design procedure. Many stability constraints, such as the positive realness based stability constraint (3.37), the Rouché's theorem based stability constraint (3.41), and the generalized positive realness based stability constraint (4.26), can be readily used in the proposed design procedure.

In this dissertation, we adopt a monitoring strategy to make all poles lie inside the stability domain. The positive realness based stability condition [32] has been given in (3.37). This sufficient stability condition can be readily extended to the situation where all poles of the designed IIR filter are required to lie inside a circle of radius $\rho_{\max} \leq 1$ for robust stability:

$$\begin{aligned} &\mathrm{Re}\{Q(\rho_{\max} e^{j\omega})\} \\ &= 1 + \sum_{m=1}^{M} q_m \rho_{\max}^{-m} \cos m\omega \\ &\geq \nu, \qquad \forall \omega \in [0, \pi] \end{aligned} \tag{5.23}$$

From (5.23), we have

$$1+\sum_{m=1}^{M} q_m \rho_{\max}^{-m} \cos m\omega \geq 1-\sum_{m=1}^{M} |q_m| \rho_{\max}^{-m}$$
$$\geq 1-\sqrt{\sum_{m=1}^{M} |q_m|^2} \cdot \sqrt{\sum_{m=1}^{M} \rho_{\max}^{-2m}} \tag{5.24}$$

In (5.24), the second inequality is obtained by the Cauchy-Schwartz inequality. By combining (5.23) and (5.24), we can construct a stability condition as

$$\|\overline{\boldsymbol{q}}\|_2 = \sqrt{\sum_{m=1}^{M} q_m^2} \leq \frac{1-\nu}{\sqrt{\sum_{m=1}^{M} \rho_{\max}^{-2m}}} \tag{5.25}$$

It can be observed from (5.25) that if ν is fixed, we can force the poles to move towards the origin (*i.e.*, $\rho_{\max} \to 0$) by suppressing $\|\overline{\boldsymbol{q}}\|_2$. When all poles lie on the origin (*i.e.*, $\rho_{\max} = 0$), we have $\|\overline{\boldsymbol{q}}\|_2 = 0$ and the designed IIR digital filter essentially degenerates to an FIR digital filter. However, the stability condition (5.25) is too restrictive to be directly applied in practical designs. Instead of employing a fixed upper bound for $\|\overline{\boldsymbol{q}}\|_2$, we can gradually reduce $\|\overline{\boldsymbol{q}}\|_2^2$ during the design procedure. Note that when $\operatorname{rank} \boldsymbol{Z} = 1$, the relaxed LMI constraint (5.14.b) is reduced to (5.12.b), and then we have $\operatorname{Tr}\{\boldsymbol{X}_q\} = \|\overline{\boldsymbol{q}}\|_2^2$. Therefore, we can attenuate $\|\overline{\boldsymbol{q}}\|_2^2$ by reducing $\operatorname{Tr}\{\boldsymbol{X}_q\}$, which can be accomplished by augmenting the regularization coefficient α in the objective function of (5.17). Since a large α may result in an over-attenuated $\operatorname{Tr}\{\boldsymbol{Z}\}$, which degrades the performance of obtained IIR filters, the value of α should be carefully selected. Here, we also resort to the inner bisection search procedure. In Step 2 of the inner bisection search procedure described earlier, after solving the regularized feasibility problem (5.17), besides the ratio $\eta\left(\boldsymbol{Z}^{(k,l)}\right)$, we also need to check the maximum pole radius of the obtained IIR filter, which is represented by $\rho\left(\boldsymbol{q}^{(k,l)}\right)$ where $\boldsymbol{q}^{(k,l)} = \begin{bmatrix} 1 & \left(\overline{\boldsymbol{q}}^{(k,l)}\right)^T \end{bmatrix}^T$. If $\eta\left(\boldsymbol{Z}^{(k,l)}\right) \leq \varepsilon$ and $\rho\left(\boldsymbol{q}^{(k,l)}\right) \leq \rho_{\max}$, choose $\alpha_+^{(k,l)} = \alpha^{(k,l)}$ and $\alpha_-^{(k,l)} = \alpha_-^{(k,l-1)}$ such that at the next iteration $\alpha^{(k,l+1)}$ will be augmented. Otherwise, set $\alpha_-^{(k,l)} = \alpha^{(k,l)}$ and $\alpha_+^{(k,l)} = \alpha_+^{(k,l-1)}$ such that

at the next iteration $\alpha^{(k,l+1)}$ will be reduced. Similarly, at each outer iteration, we need to check $z^{(k,0)}$, $\eta\left(\boldsymbol{Z}^{(k,0)}\right)$, and $\rho\left(\boldsymbol{q}^{(k,0)}\right)$ after solving (5.14) in order to determine whether or not the inner bisection search procedure needs to be invoked.

In practice, some other constraints can be imposed on $\overline{\boldsymbol{x}}$ and $\boldsymbol{X}$ to refine the formulation of the feasibility problems (5.14) and (5.17), such that the relaxed feasibility problem can approach the original design problem as well as possible or the obtained IIR filters can satisfy some specific requirements. In our designs, the following linear inequality constraints in terms of the denominator coefficients $\overline{\boldsymbol{q}}$ and the diagonal elements $\left[\boldsymbol{X}_q\right]_{(m,m)}$ of $\boldsymbol{X}_q$ are also incorporated:

$$|q_m| \le C(M,m)\rho_{\max}^m, \qquad m = 1,2,\dots,M \tag{5.26}$$

$$\left[\boldsymbol{X}_q\right]_{(m,m)} \le [C(M,m)\rho_{\max}^m]^2, \qquad m = 1,2,\dots,M \tag{5.27}$$

where $C(M,m) = M!/[m!\,(M-m)!]$. It can be verified that (5.26) and (5.27) are necessary conditions for the stability of designed IIR filters.

The flowchart of the complete design method is shown in Fig. 5.1. The dashed box indicates the inner bisection search procedure described in Section 5.1.3. It can be seen from Fig. 5.1 that the major computation is expended to solve the SDP feasibility problem (5.17).

5.1.5 Initial Lower Bound Estimation Using SDP Relaxation

The last issue we need to address is how to estimate the initial lower bound $\underline{\delta}^{(0)}$ of the minimum error limit δ^* for the outer bisection search procedure. Obviously, the initial design (4.9) used by the sequential SOCP design method presented in Chapter IV can be directly applied here. In this section, we shall make use of the SDP relaxation technique described in Section 5.1.2 to reformulate an SDP design problem. By solving this relaxed design problem, we can also obtain an initial lower bound $\underline{\delta}^{(0)}$. It will be shown that this SDP design problem is related to the SOCP design problem (4.9).

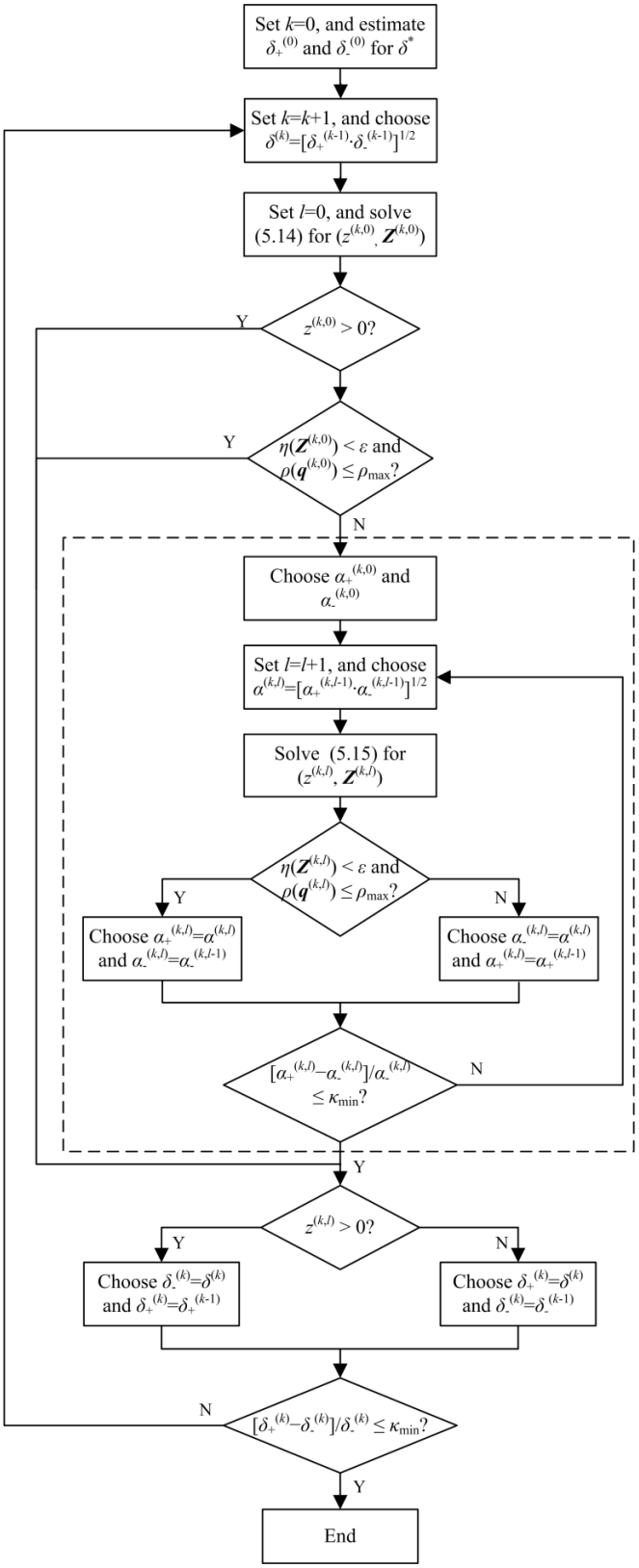

Fig. 5.1 Flowchart of the complete design method.

The SDP relaxation technique described in Section 5.1.2 can be applied only on the right-hand side of (5.3), and then we can obtain the following SDP design problem

$$\min \quad \delta \tag{5.28}$$

$$\text{s.t.} \quad |D(\omega_i)|^2 + 2\text{Re}\{D(\omega_i)\boldsymbol{c}^H(\omega_i)\}\overline{\boldsymbol{x}} + \overline{\boldsymbol{x}}^T\boldsymbol{A}(\omega_i)\overline{\boldsymbol{x}} \leq \frac{\delta}{W^2(\omega_i)} \cdot \left[1 + 2\text{Re}\left\{\overline{\boldsymbol{\varphi}}_M^H\left(e^{j\omega_i}\right)\right\}\overline{\boldsymbol{q}} + \text{Tr}\{\boldsymbol{X}_q\boldsymbol{B}(\omega_i)\}\right] \tag{5.28.a}$$

where $\overline{\boldsymbol{x}} = [\overline{\boldsymbol{q}}^T \quad \boldsymbol{p}^T]^T$, and $\omega_i \in \Omega_I,\ i = 0,1,\dots,L$

$$\boldsymbol{Y} = \begin{bmatrix} 1 & \overline{\boldsymbol{q}}^T \\ \overline{\boldsymbol{q}} & \boldsymbol{X}_q \end{bmatrix} \succeq 0 \tag{5.28.b}$$

In (5.28), the decision variables are δ, $\boldsymbol{p}$, and $\boldsymbol{Y}$. Unlike the linear inequality constraint (5.11) which is expressed in terms of $\overline{\boldsymbol{x}}$ and $\boldsymbol{X}$, now (5.28.a) is a hyperbolic constraint, which can be recast as an LMI constraint [74]. Compared with (4.2.b), the constraints (5.28.a) and (5.28.b) define a larger feasible set. Therefore, a lower bound on the optimal value of the original design problem (4.2) can be obtained by solving (5.28). The major difference between (5.28) and (4.9) is that the trigonometric function $\boldsymbol{d}^T\boldsymbol{s}(\omega)$ in (4.9) has been replaced by a linear function of the denominator coefficients $\overline{\boldsymbol{q}}$ and the elements of $\boldsymbol{X}_q$. When $\text{rank}\,\boldsymbol{Y} = 1$, we have $\boldsymbol{X}_q = \overline{\boldsymbol{q}}\overline{\boldsymbol{q}}^T$ and, accordingly,

$$\text{Tr}\{\boldsymbol{Y}\} = 1 + \text{Tr}\{\boldsymbol{X}_q\} = \sum_{m=0}^{M} q_m^2 \tag{5.29}$$

$$\text{Tr}_m\{\boldsymbol{Y}\} = \sum_{i=0}^{M-m} q_i q_{i+m}, \quad m = 1,2,\dots,M \tag{5.30}$$

where $\text{Tr}_m\{\boldsymbol{Y}\}$ denotes the sum along the mth diagonal of $\boldsymbol{Y}$. Comparing (5.29) and (5.30) with (4.6), we can find that if (4.5) is satisfied, $\text{Tr}\{\boldsymbol{Y}\} = d_0$ and $\text{Tr}_m\{\boldsymbol{Y}\} = d_m$ where d_m is defined by (4.6). In both initial designs, *i.e.*, (4.9) and (5.28), the convex relaxation techniques have been employed to transform the original nonconvex constraint (4.2.b) into convex forms. Specifically, the equality constraint (4.7) is relaxed to (4.9.d), while

the constraint $\boldsymbol{X}_q = \overline{\boldsymbol{q}}\overline{\boldsymbol{q}}^T$ is relaxed to (5.28.b). In practice, these two initial designs can generate similar lower bounds of the minimum error limit δ^*.

It should be mentioned that although the relaxed constraints of (5.14) and (5.28) are both obtained by applying the SDP relaxation technique on the original nonconvex constraint (4.2.b), they are used in different situations and cannot be replaced by each other. In (5.14) the error limit δ must be fixed. Otherwise, (5.14.a) is cannot be directly incorporated in the convex feasibility problem (5.14). However, in (5.28), the objective function is chosen as δ subject to a set of relaxed constraints. Therefore, the relaxed constraints of (5.14) cannot be applied to find an initial lower bound $\delta_-^{(0)}$ in (5.28). Given $\delta^{(k)}$, the relaxed constraints (5.28.a) and (5.28.b) could be applied to construct the following feasibility problem, which is similar to (5.14)

$$\min \quad z \tag{5.31}$$

$$\begin{aligned} \text{s.t.} \quad & |D(\omega_i)|^2 + 2\text{Re}\{D(\omega_i)\boldsymbol{c}^H(\omega_i)\}\overline{\boldsymbol{x}} + \overline{\boldsymbol{x}}^T\boldsymbol{A}(\omega_i)\overline{\boldsymbol{x}} \\ & \leq \frac{\delta^{(k)}}{W^2(\omega_i)} \cdot \left[1 + 2\text{Re}\left\{\overline{\boldsymbol{\varphi}}_M^H\left(e^{j\omega_i}\right)\right\}\overline{\boldsymbol{q}} + \text{Tr}\left\{\boldsymbol{X}_q\boldsymbol{B}(\omega_i)\right\}\right] + z \\ & \text{where } \overline{\boldsymbol{x}} = [\overline{\boldsymbol{q}}^T \quad \boldsymbol{p}^T]^T, \text{ and } \omega_i \in \Omega_I,\ i = 0,1,\dots,L \end{aligned} \tag{5.31.a}$$

$$\boldsymbol{Y} = \begin{bmatrix} 1 & \overline{\boldsymbol{q}}^T \\ \overline{\boldsymbol{q}} & \boldsymbol{X}_q \end{bmatrix} \succeq 0 \tag{5.31.b}$$

However, this formulation will lead to problematical solutions. Assume that by solving (5.31) with a given $\delta^{(k)}$, a set of z, $\overline{\boldsymbol{x}}$, and $\boldsymbol{X}_q$ (or $\boldsymbol{Y}$) have been obtained. Since $\boldsymbol{X}_q$ is PSD, we can construct another PSD matrix $b\boldsymbol{X}_q$ for any $b > 1$, which satisfies $\text{Tr}\{b\boldsymbol{X}_q\boldsymbol{B}(\omega)\} > \text{Tr}\{\boldsymbol{X}_q\boldsymbol{B}(\omega)\} > 0$ and $b\boldsymbol{X}_q \succeq \overline{\boldsymbol{q}}\overline{\boldsymbol{q}}^T$. Then, by taking $b\boldsymbol{X}_q$ into (5.31.a) and (5.31.b), it can be verified that the scaled matrix $b\boldsymbol{X}_q$ can also satisfy these two constraints with the obtained z and $\overline{\boldsymbol{x}}$. Thereby, in (5.31.a) the value of z can be slightly reduced without changing the inequality sign of (5.31.a). This implies that by sufficiently scaling $\boldsymbol{X}_q$, we can always make $z \leq 0$. Under this circumstance, $\delta_+^{(k)}$ will be chosen as $\delta^{(k)}$ and eventually reduced to the initial lower bound $\delta_-^{(0)}$. Obviously, the desired

minimum error limit and the corresponding filter coefficients cannot be obtained by applying (5.31) in the outer bisection search procedure to locate δ^*.

5.2 Simulations

In this section, four examples are presented to demonstrate the effectiveness of the proposed design method. Theoretically speaking, in order to approach a rank-1 solution, the value of parameter ε should be chosen as small as possible. In practice, however, this parameter cannot be too small, otherwise Tr$\{\boldsymbol{Z}\}$ could be over-attenuated. In all the examples presented in this section, parameter ε is chosen as 5×10^{-2}, which is also suitable for most of designs we have tried so far. The value of parameter $\kappa_{\min}$ can be arbitrarily selected. In general, a smaller $\kappa_{\min}$ leads to a more accurate design, but the total number of iterations will accordingly be increased. In our designs, $\kappa_{\min}$ is set equal to 10^{-3}. The initial upper and lower bounds of parameter α used by the outer iterations in which the inner bisection search procedure is invoked for the first time can be arbitrarily selected, provided they are sufficiently close to 1 and 0, respectively. In our designs, they are chosen, respectively, as 10^{-2} and 10^{-12}. At the succeeding iterations, we choose $\gamma = 5$ in (5.21) and (5.22) to determine the upper and lower bounds $\alpha_+^{(k,l)}$ and $\alpha_-^{(k,l)}$ of the regularization coefficient. Parameter γ can take some larger value to extend the search range of α. However, according to (5.20)-(5.22), the inner bisection search procedure needs more iterations to find an appropriate α. Linear inequality constraints (5.14.a) and (5.17.a) are both imposed on a set of discrete frequency points taken from 101 equally-spaced grid points over the whole frequency band. If the weighting function $W(\omega)$ is not explicitly defined in the specifications, it is always set equal to 1 over Ω_I, and 0 otherwise. Similarly, without any explicit declaration, the admissible maximum pole radius is always chosen as $\rho_{\max} = 1$. Besides the peak and L_2 errors of the magnitude (MAG) and group delay (GD) responses over Ω_I, we also adopt the weighted minimax error E_{MM} defined by (1.20) to evaluate the performance of the designed filters. In our designs, all the SDP problems are solved by *SeDuMi* [66] in MATLAB environment.

5.2.1 Example 1

The first example is to design a lowpass digital filter with the following ideal frequency response

$$D(\omega) = \begin{cases} e^{-j12\omega} & 0 \le \omega \le 0.4\pi \\ 0 & 0.56\pi \le \omega < \pi \end{cases}$$

The numerator and denominator orders are chosen, respectively, as $N = 15$ and $M = 4$. The design specifications are exactly the same as those adopted by the first example of [20]. Using the proposed method, we design an IIR digital filter. All the filter coefficients are summarized in Table 5.1. The maximum pole radius of the obtained IIR filter is 0.8589. The magnitude and group delay responses are shown as solid curves in Fig. 5.2. The magnitude of the weighted complex approximation error, *i.e.*, $|E(\omega)|$, is plotted in Fig. 5.3. Simulation result reveals that in this design $T_o = 13$ and $T_i(k) = 0$ for $k = 1, 2, ..., 13$, which implies that the inner bisection search procedure is actually not invoked. By analyzing the final output $\boldsymbol{Z}$, we find that except the largest eigenvalue $\lambda_1(\boldsymbol{Z})$ (= 2.4617), all the other eigenvalues of $\boldsymbol{Z}$ are negligible ($\le 8.9715\times10^{-7}$). Then, by ignoring $\lambda_i(\boldsymbol{Z})$ ($i = 2, 3, ..., N+M+2$), the obtained $\boldsymbol{Z}$ can be approximately regarded as a rank-1 matrix. In view of the Proposition 1 described in Section 5.1.2, it can be concluded that the final solution is very close to the optimal solution of the original design problem. Note that based on the Proposition 1, we can detect the optimality of the obtained IIR filter. However, there is no guarantee that it is the unique optimal solution. In this example the denominator order M is not too high and the design specifications are not stringent. Hence, the optimal design can be obtained by only successively solving the relaxed feasibility problem (5.14). In general, however, the inner bisection search procedure has to be used to attain rank-1 solutions. The same set of specifications have been used by Example 1 in Chapter IV. By comparing the error measurements listed, respectively, in Table 4.2 and Table 5.2, we can find that the IIR filter designed by the sequential SOCP method proposed in Chapter IV is also very close to the optimal design, although its optimality cannot be verified therein.

Table 5.1 Filter Coefficients (p_0 to p_N and q_0 to q_M) of IIR Digital Filter Designed in Example 1

p_0~p_4	-2.7732e-003	-2.2843e-003	3.9183e-003	3.6388e-003	-6.7658e-003
p_5~p_9	-8.1916e-003	1.0842e-002	1.7997e-002	-1.7897e-002	-4.5489e-002
p_{10}~p_{14}	3.3981e-002	2.2224e-001	3.7818e-001	3.7119e-001	2.2084e-001
p_{15}	7.5101e-002				
q_0~q_4	1.0000e+000	-4.5733e-001	8.9053e-001	-2.5287e-001	8.0733e-002

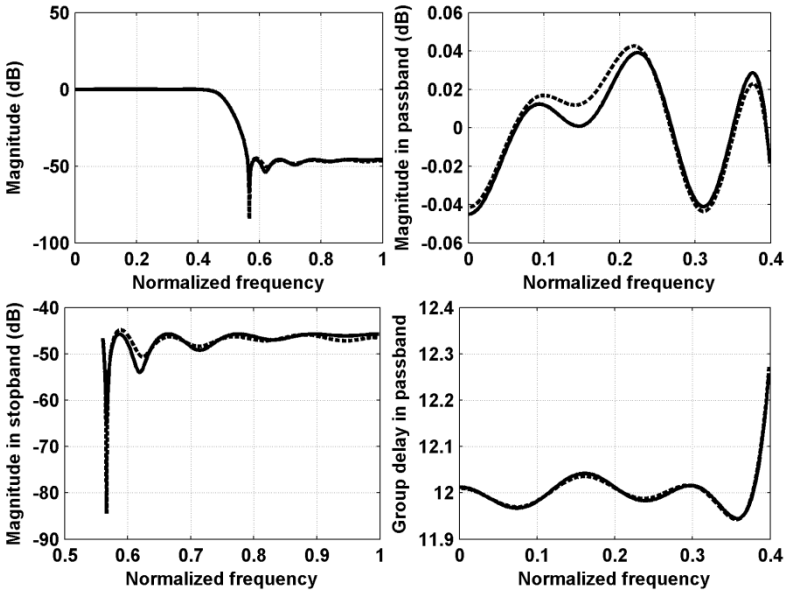

Fig. 5.2 Magnitude and group delay responses of IIR filters designed in Example 1. Solid curves: designed by the proposed method. Dashed curves: designed by the SM method [8].

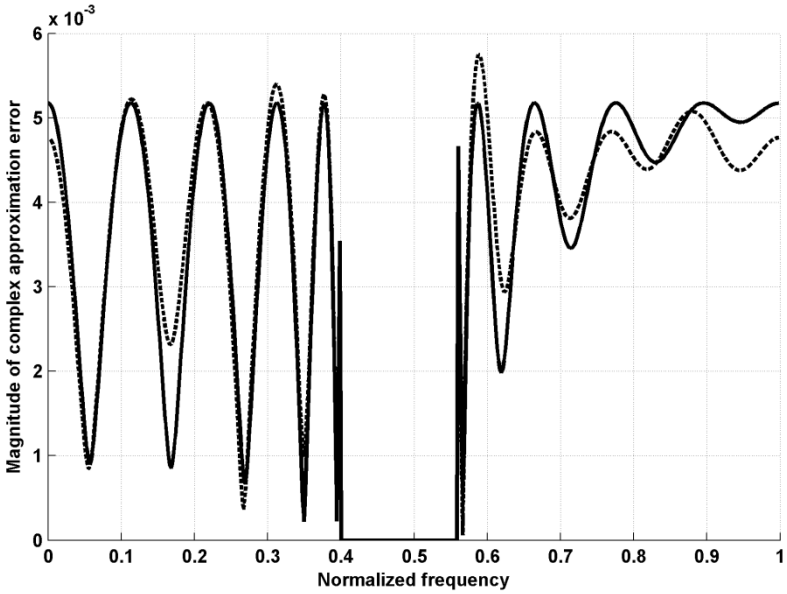

Fig. 5.3 Magnitude of complex approximation error $|E(\omega)|$ in Example 1. Solid curves: designed by the proposed method; Dashed curves: designed by the SM method [8].

Table 5.2 Error Measurements of Design Results in Example 1

Method	Minimax Error E_{MM} (in dB)	Passband MAG (Peak/L_2 in dB)	Passband GD (Peak/ L_2)	Stopband MAG (Peak/L_2 in dB)
Proposed	-45.721	-45.721/ -55.162	2.773e-1/ 2.537e-2	-45.720/ -50.378
SM [8]	-44.810	-45.998/ -54.561	2.933e-1/ 2.604e-2	-44.807/ -50.543

For comparison, we also utilize the SM method [8] to design an IIR filter under the same set of specifications. The initial point is chosen as the optimal FIR design with the filter order equal to N. The design result shows that the SM method can achieve a stable IIR filter even without the positive realness based stability constraint (3.37). The maximum pole radius of the obtained IIR filter is 0.8622. The magnitude and group delay responses are also plotted in Fig. 5.2 as dashed curves. The magnitude of the corresponding complex approximation error is also shown as dashed curves in Fig. 5.3. All the error measurements are summarized in Table 5.2. It can be observed that the proposed method can achieve slightly better performance except in peak error of the passband magnitude and L_2 error of the stopband magnitude than those obtained by the SM method [8].

5.2.2 Example 2

The second example, which is taken from [25], is to design another lowpass filter. The ideal frequency response is defined by

$$D(\omega) = \begin{cases} e^{-j5\omega} & 0 \le \omega \le 0.2\pi \\ 0 & 0.4\pi \le \omega < \pi \end{cases}$$

Numerator and denominator orders are set equal to $N = M = 4$. After 14 outer iterations, *i.e.*, $T_o = 14$, the outer bisection search procedure converges to the final solution. Only at the second outer iteration, the inner bisection search procedure is invoked, and $T_i(2) = 15$. The minimum value of α determined by the inner bisection search procedure is 2.3714×10^{-6}. The maximum pole radius of the obtained IIR filter is 0.8975. The first and second largest eigenvalues of the final output $\boldsymbol{Z}$ of the proposed design method are 19.6301 and 2.1717×10^{-5}. Both numerator and denominator coefficients of the obtained

IIR filter are summarized in Table 5.3. The magnitude and group delay responses are plotted as solid curves in Fig. 5.4. The magnitude of the complex approximation error $E(\omega)$ is shown in Fig. 5.5.

For comparison, we also design an IIR digital filter using the design method [25] under the same set of specifications. This IIR filter design method is based on the formulation of a generalized eigenvalue problem by using the Remez exchange algorithm. Numerator and denominator coefficients of the corresponding IIR filter have been given in [25]. The maximum pole radius of the obtained IIR filter is 0.8771. The magnitude and group delay responses and the magnitude of complex approximation error are also shown as dashed curves in Fig. 5.4 and Fig. 5.5, respectively. All the error measurements are summarized in Table 5.4. It is obvious that the proposed method can achieve better performance than the design method [25].

Table 5.3 Filter Coefficients (p_0 to p_N and q_0 to q_M) of IIR Digital Filter Designed in Example 2

$p_0 \sim p_4$	-2.3339e-002	4.1194e-002	1.1390e-002	1.1163e-002	4.4441e-002
$q_0 \sim q_4$	1.0000e+000	-2.5935e+000	2.9782e+000	-1.6947e+000	3.9670e-001

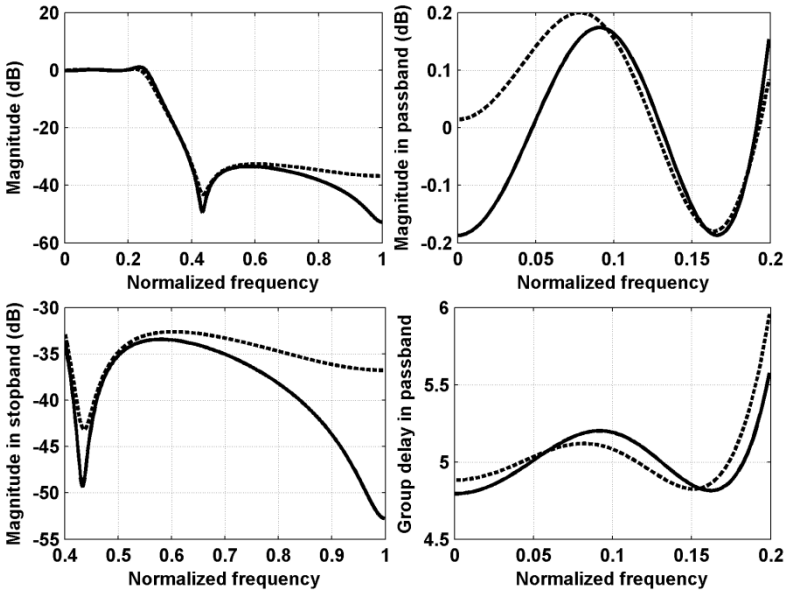

Fig. 5.4 Magnitude and group delay responses of IIR filters designed in Example 2. Solid curves: designed by the proposed method. Dashed curves: designed by the Remez multiple exchange method [25].

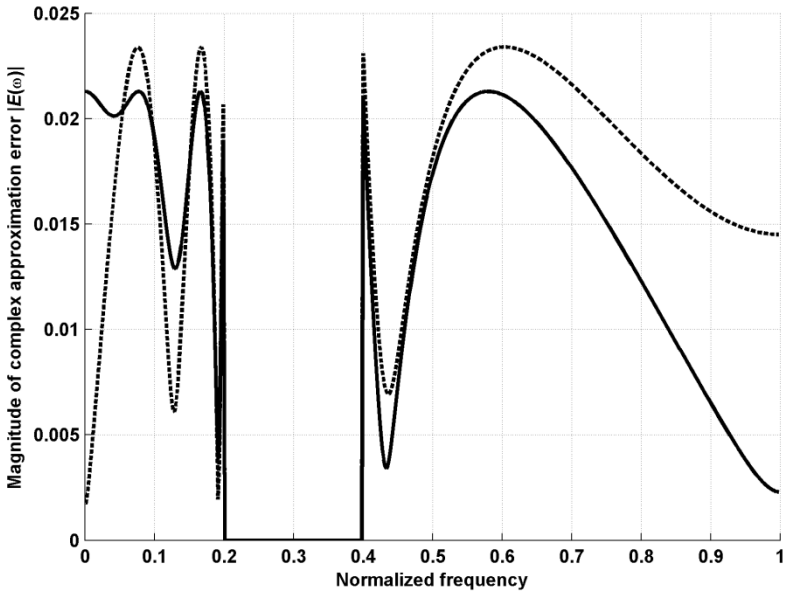

Fig. 5.5 Magnitude of complex approximation error $|E(\omega)|$ in Example 2. Solid curves: designed by the proposed method; Dashed curves: designed by the Remez multiple exchange method [25].

Table 5.4 Error Measurements of Design Results in Example 2

Method	Minimax Error E_{MM} (in dB)	Passband MAG (Peak/L_2 in dB)	Passband GD (Peak/ L_2)	Stopband MAG (Peak/L_2 in dB)
Proposed	-33.437	-33.437/ -43.697	5.766e-1/ 7.114e-2	-33.437/ -38.931
Remez [25]	-32.613	-32.669/ -43.598	9.573e-1/ 8.654e-2	-32.617/ -36.826

5.2.3 Example 3

The third example is to design two full-band digital differentiators [18] with the ideal frequency response

$$D(\omega) = \frac{\omega}{\pi} e^{j[0.5\pi - (\tau_s + 0.5)\omega]}, \qquad 0 \le \omega < \pi$$

where τ_s is an integer delay. The first differentiator is of order 8, *i.e.*, $N = M = 8$. And the filter order in the second design is set to 5. In both designs, τ_s is chosen as 3. Therefore, the ideal group delay is equal to 3.5 over the whole frequency band. As proposed in [18], the weighting functions in both designs are chosen as

$$W(\omega) = \begin{cases} \pi/\omega & 0.1\pi < \omega < \pi \\ 10 & 0 \le \omega \le 0.1\pi \end{cases}$$

In [18], an IIR differentiator of order 8 is first designed by the modified Ellacott-Williams (EW) algorithm, which utilizes the first-order Taylor series to simplify the denominator design at each iteration, while the optimal numerator for a given denominator can be obtained by solving (4.27). However, the obtained differentiator of order 8 is a degenerate filter. There are three pairs of poles and zeros which nearly cancel each other. After removing these poles and zeros, the remaining poles and zeros are then used to construct an IIR differentiator of order 5, from which a new IIR differentiator of order 5 with the same ideal group delay is redesigned by the modified EW algorithm. The poles and zeros of these two differentiators are given in [18]. In both designs of [18], the admissible maximum pole radius is specified as 0.98. The maximum pole radii of the designed differentiators of order 8 and order 5 are 0.6829 and 0.4400, respectively.

For comparison, we choose the admissible maximum pole radii as 0.7 and 0.5 in our designs. In the design of differentiator of order 8, after 14 outer iterations, *i.e.*, T_o = 14, the design procedure converges to the final solution. At each outer iteration, the inner bisection search procedure is invoked, and simulation result shows that $T_i(1)$ = 15 and $T_i(k)$ = 12 for k = 2, 3, ..., 14. The minimum value of α determined by each inner bisection search procedure is within the range of [7.2448×10^{-7}, 6.7989×10^{-5}]. The largest eigenvalue of the final output $\boldsymbol{Z}$ is 1.3300, and all the other eigenvalues are less than 9.0490×10^{-8}. Filter coefficients of the designed IIR differentiator of order 8 are listed in Table 5.5. In the design of differentiator of order 5, T_o = 13, and $T_i(1) = T_i(2)$ = 0, $T_i(3)$ = 15, $T_i(k)$ = 12 for k = 4, 5, ..., 13. The minimum value of α determined by each inner bisection search procedure is within the range of [7.2385×10^{-6}, 3.5142×10^{-5}]. The final output $\boldsymbol{Z}$ has eigenvalues $\lambda_1(\boldsymbol{Z})$ = 1.3651 and $\lambda_i(\boldsymbol{Z}) \le 6.6688\times10^{-7}$ (i = 2, 3, ..., 12). Filter coefficients of the obtained differentiator of order 5 are also given in Table 5.5. The design characteristics and errors of these two IIR differentiators are shown in Fig. 5.6 and Fig. 5.7, respectively. As in Example 4 of Section 4.3, the approximation errors of group delay response within the frequency band $[0.05\pi, \pi]$ are ignored when evaluating the peak and L_2 errors of group delay. The magnitudes of $E(\omega)$ of IIR differentiators are

both shown in Fig. 5.8, where solid and dashed curves correspond, respectively, to the IIR differentiators of order 8 and order 5. All the error measurements are summarized in Table 5.6.

Table 5.5 Filter Coefficients (p_0 to p_N and q_0 to q_M) of IIR Digital Differentiators Designed in Example 3

Order 8	$p_0 \sim p_4$	-1.0371e-002	1.9258e-002	-4.2066e-002	3.9520e-001	-3.2480e-001
	$p_5 \sim p_8$	-6.0965e-002	5.1963e-002	-6.7226e-002	3.6988e-002	
	$q_0 \sim q_4$	1.0000e+000	1.8608e-001	-6.2769e-002	7.4670e-002	-9.7975e-002
	$q_5 \sim q_8$	-8.4768e-003	1.8686e-004	-2.3250e-004	-2.0158e-003	
Order 5	$p_0 \sim p_4$	-1.0459e-002	1.7395e-002	-3.9043e-002	3.8891e-001	-2.5654e-001
	p_5	-1.0236e-001				
	$q_0 \sim q_4$	1.0000e+000	3.6826e-001	4.5442e-003	4.3005e-003	-7.1699e-004
	q_5	7.3662e-003				

Table 5.6 Error Measurements of Design Results in Example 3

Method	Order	Minimax Error E_{MM} (in dB)	MAG (Peak/L_2 in dB)	GD within $[0.05\pi, \pi]$ (Peak/ L_2)
Proposed	8	-34.656	-35.122/ -43.737	3.197e-1/ 6.447e-2
	5	-33.032	-33.418/ -43.294	2.434e-1/ 6.143e-2
Modified EW [18]	8	-30.918	-32.776/ -41.718	3.580e-1/ 7.582e-2
	5	-27.883	-28.122/ -41.666	3.265e-1/ 7.859e-2

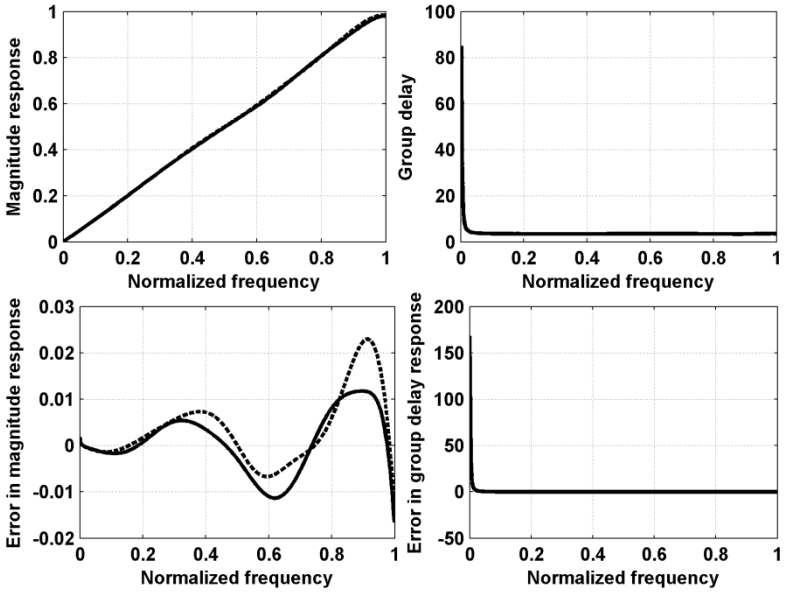

Fig. 5.6 Design characteristics and errors of the differentiator of order 8 in Example 3. Solid curves: designed by the proposed method. Dashed curves: designed by the modified EW method [18].

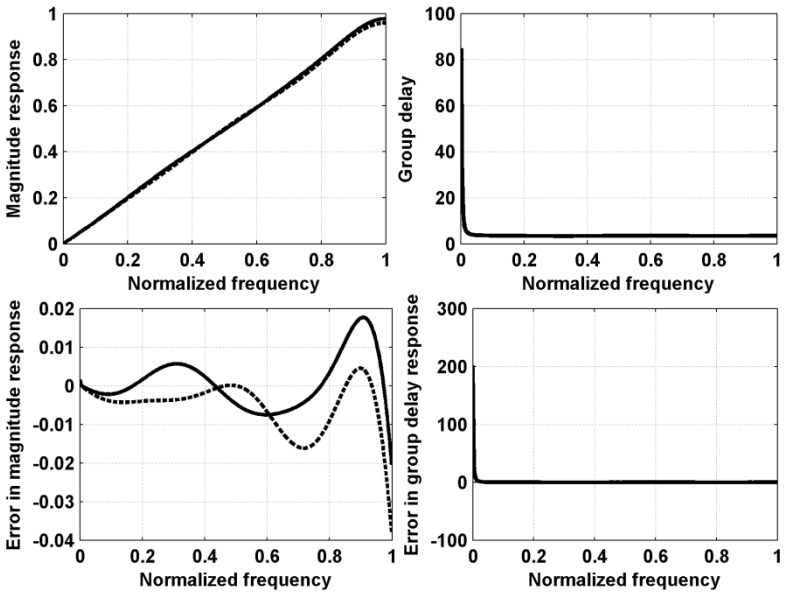

Fig. 5.7 Design characteristics and errors of IIR differentiator of order 5 designed in Example 3. Solid curves: designed by the proposed method. Dashed curves: designed by the modified EW method [18].

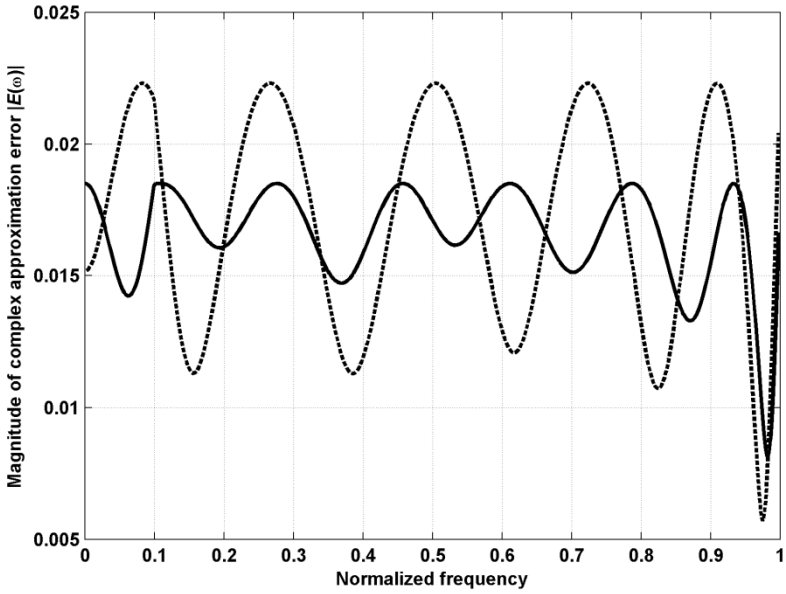

Fig. 5.8 Magnitudes of complex approximation error $|E(\omega)|$ of IIR differentiators designed in Example 3. Solid curves: differentiator of order 8; Dashed curves: differentiator of order 5.

5.2.4 Example 4

The last example is to design a halfband highpass filter [11], [28]. The desired frequency response is given by

$$D(\omega) = \begin{cases} e^{-j12\omega} & 0.525\pi \le \omega < \pi \\ 0 & 0 \le \omega \le 0.475\pi \end{cases}$$

Numerator and denominator orders are chosen as $M = N = 14$. First of all, we directly utilize the proposed method to design an IIR filter with $\rho_{\max} = 0.98$. The final solution is obtained after 14 outer iterations. The total number of inner iterations at each outer iteration is $T_i(1) = 0$, $T_i(2) = 15$, and $T_i(k) = 12$ for $k = 3, 4, ..., 14$. The regularization coefficients determined by these inner bisection search procedures are within the range of [1.1814×10^{-6}, 3.6685×10^{-6}]. The largest eigenvalue of the final output $\boldsymbol{Z}$ is equal to 2.5978, whereas $\lambda_i(\boldsymbol{Z}) \le 7.2489\times10^{-5}$ for $i = 2, 3, ..., 30$. The maximum pole radius of the designed filter is 0.9800. All the filter coefficients are given in Table 5.7. The magnitude and group delay responses, and the magnitude of $|E(\omega)|$ are shown as dash-dotted curves in Fig. 5.9 and Fig. 5.10, respectively. The corresponding error measurements (referred as Proposed-1) are given in Table 5.8. For comparison, the SM method [8] is employed to design an IIR digital filter under the same specifications. The design procedure starts from an optimal FIR filter design. The maximum pole radius of the obtained IIR filter is 0.9346. The corresponding magnitude of $E(\omega)$ is also shown as dashed curves in Fig. 5.10. Obviously, the proposed method can achieve much better performance.

Table 5.7 Filter Coefficients (p_0 to p_N and q_0 to q_M) of IIR Digital Filters Designed in Example 4

Proposed-1	$p_0 \sim p_4$	-8.9283e-003	1.5280e-002	6.9703e-003	-1.9689e-004	-7.7944e-003
	$p_5 \sim p_9$	6.5802e-003	9.8544e-003	-1.7955e-002	-1.9061e-002	4.5495e-002
	$p_{10} \sim p_{14}$	4.2842e-002	-2.2074e-001	3.3228e-001	-2.8527e-001	1.6577e-001
	$q_0 \sim q_4$	1.0000e+000	5.8712e-001	6.9620e-001	-9.5168e-002	-4.0565e-001
	$q_5 \sim q_9$	-1.6947e-001	2.1597e-001	2.6214e-001	-2.6027e-002	-2.4264e-001
	$q_{10} \sim q_{14}$	-1.4439e-001	1.0613e-001	2.3135e-001	1.6991e-001	5.8671e-002
Proposed-2	$p_0 \sim p_4$	-5.9409e-003	1.3554e-002	7.6070e-003	-4.7667e-003	-1.9294e-002
	$p_5 \sim p_9$	-5.1246e-003	1.2685e-002	-1.6718e-003	-1.5201e-002	2.8330e-002
	$p_{10} \sim p_{14}$	4.2385e-002	-1.9479e-001	2.7198e-001	-2.0772e-001	1.3183e-001
	$q_0 \sim q_4$	1.0000e+000	9.7306e-001	1.1889e+000	3.3032e-001	-4.1280e-001
	$q_5 \sim q_9$	-5.5491e-001	-1.1038e-001	3.5085e-001	3.6545e-001	1.2937e-002
	$q_{10} \sim q_{14}$	-3.0728e-001	-3.5532e-001	-2.1124e-001	-6.7584e-002	-5.0548e-003

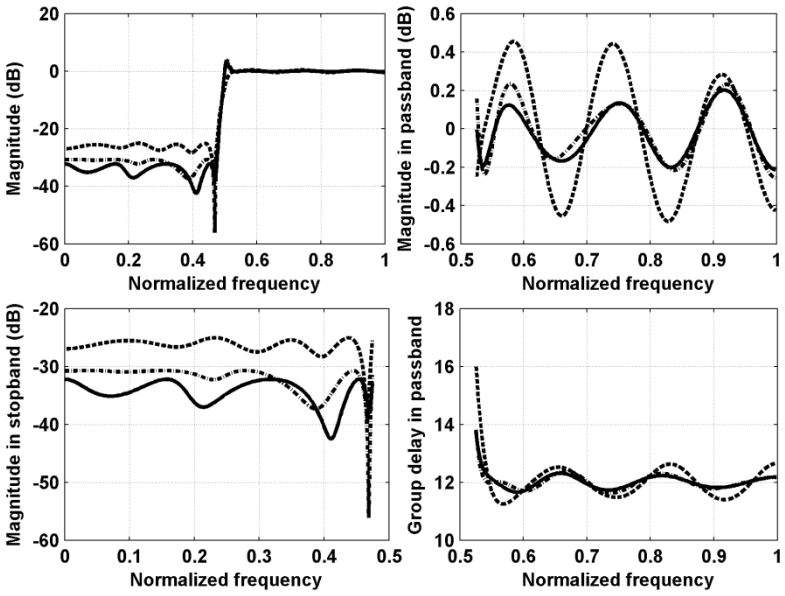

Fig. 5.9 Magnitude and group delay responses of IIR filters designed in Example 4. Solid curves: designed by the proposed method ($\rho_{\max} = 1$) followed by rescaling $\boldsymbol{q}$ through (5.32) and solving (4.27). Dash-dotted curves: designed by the proposed method ($\rho_{\max} = 0.98$). Dash curves: designed by the SM method [8].

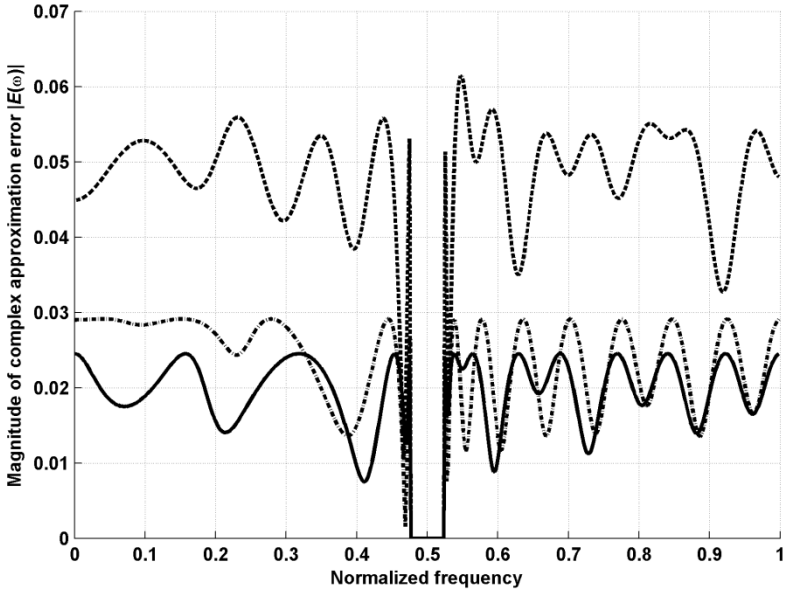

Fig. 5.10 Magnitude of complex approximation error $|E(\omega)|$ in Example 4. Solid curves: designed by the proposed method ($\rho_{\max} = 1$) followed by rescaling $\boldsymbol{q}$ through (5.32) and solving (4.27). Dash-dotted curves: designed by the proposed method ($\rho_{\max} = 0.98$). Dash curves: designed by the SM method [8].

Table 5.8 Error Measurements of Design Results in Example 4

Method	Minimax Error E_{MM} (in dB)	Passband MAG (Peak/L_2 in dB)	Passband GD (Peak/ L_2)	Stopband MAG (Peak/L_2 in dB)
Proposed-1	-30.714	-30.720/ -38.988	1.814/ 1.689e-1	-30.714/ -34.970
Proposed-2	-32.212	-32.218/ -40.058	1.716/ 1.649e-1	-32.211/ -37.270
SM [8]	-24.231	-25.333/ -32.497	4.062/ 4.030e-1	-25.051/ -29.538

In Section 5.1.3, we mentioned that the regularization parameter α should be appropriately selected in order to avoid Tr{$\boldsymbol{Z}$} and, accordingly, $\|\bar{\boldsymbol{x}}\|_2$ being over-attenuated. In order to demonstrate the effects of over-attenuation on the design performances, we redesign an IIR filter using the proposed method under the same set of specifications except the admissible maximum pole radius $\rho_{\max} = 1$. In so doing, the final solution can be obtained after 14 outer iterations. Simulation results show that $T_i(2) = 15$, $T_i(k) = 12$ for $k = 3, 7, 9$, and $T_i(k) = 0$ for $k = 1, 4, 5, 6, 8, 10, \ldots, 14$. The regularization coefficients determined by the inner iterations are within the range of $[5.5412\times10^{-8}, 1.0228\times10^{-6}]$. The largest eigenvalue of the obtained $\boldsymbol{Z}$ is 5.1767, and other eigenvalues $\lambda_i(\boldsymbol{Z})$ ($i = 2, 3, ..., 30$) are less than 0.0888. In order to make all poles lie inside the circle of the radius $\rho_{\max} = 0.98$, we can simply rescale the denominator coefficients of the obtained IIR filter (with $\rho_{\max} = 1$) as

$$\hat{q}_m = q_m \left[\frac{\rho_{\max}}{\rho(\boldsymbol{q})}\right]^m, \quad m = 1, 2, \ldots, M \tag{5.32}$$

where $\hat{q}_m$ denotes the rescaled denominator coefficients. Given $\hat{\boldsymbol{q}} = [1\ \hat{q}_1\ \ldots\ \hat{q}_M]^T$, the optimal numerator coefficients $\hat{\boldsymbol{p}} = [\hat{p}_0\ \hat{p}_1\ \ldots\ \hat{p}_N]^T$ can be determined by solving the SOCP problem (4.27). Filter coefficients of the obtained IIR filter are listed in Table 5.7. The design results and the magnitude of $E(\omega)$ are plotted as solid curves in Fig. 5.9 and Fig. 5.10, respectively. The corresponding error measurements (referred as Proposed-2) are summarized in Table 5.8. Although by using (5.32) and (4.27) the obtained IIR filter is not guaranteed to be optimal, it can be observed from Table 5.8 that the current design can achieve better performance than the one directly obtained by the proposed method with $\rho_{\max} = 0.98$. We also find that the regularization coefficient α determined in the previous design (Proposed-1) is larger than the one determined in the current design

(Proposed-2). Consequently, in the previous design, the obtained $\boldsymbol{Z}$ is much closer to a rank-1 solution, which can be verified by the ratio $\eta(\boldsymbol{Z})$ of both designs: $\eta(\boldsymbol{Z})$ = 2.7904×10^{-5} for Proposed-1 and $\eta(\boldsymbol{Z})$ = 1.7154×10^{-2} for Proposed-2. We can further compare these two designs by examining the l_2-norms of obtained filter coefficients, *i.e.*, $\|\boldsymbol{x}\|_2 = \|\boldsymbol{p}\|_2+\|\boldsymbol{q}\|_2$. From Table 5.7, we can obtain $\|\boldsymbol{x}\|_2$ = 1.6118 in the first design and $\|\boldsymbol{x}\|_2$ = 2.1594 in the second design. Obviously, compared with the design result obtained by the Proposed-2 method, the l_2-norm of filter coefficients obtained by the Proposed-1 method has been over-attenuated. This is the major reason for the better performance of the Proposed-2 method in this example.

It should be emphasized that such over-attenuation does not always appear when $\rho_{\max} < 1$. For Examples 2 and 3 presented before, and many other designs with a similar level of filter requirements, the Proposed-1 method is able to arrive at a satisfactory design and no further improvement can possibly be achieved by the Proposed-2 method. The Proposed-2 method is also not necessary in those designs with much less stringent filter requirements such as Example 1, since the inner bisection search procedure is not even invoked.

CHAPTER VI

CONCLUSIONS AND FUTURE STUDY

6.1 Conclusions

In this dissertation, we have mainly studied three IIR filter design methods. Given a complex-valued desired frequency response $D(\omega)$, our design objective is to find an IIR digital filter with the transfer function $H(z)$ defined by (1.4), which can best approximate $D(\omega)$ under the WLS or minimax criterion. Due to the existence of the denominator $Q(z)$ whose roots can be anywhere in the z plane, IIR filter design problems primarily face two difficulties: 1) The design problems are essentially nonconvex. Hence, there may be many local optima existing on error performance surfaces. 2) When $M > 2$, the stability domain is also nonconvex. In this dissertation, we have proposed three IIR filter design methods under the framework of convex optimization. The most important advantage of using convex optimization to solve design problems is that if a design problem can be strictly formulated as an equivalent convex optimization problem, its globally optimal solution can be efficiently and reliably obtained. For nonconvex IIR filter design problems, approximation and convex relaxation techniques have to be employed to transform original design problems into convex forms.

In Chapter III, a sufficient and necessary stability condition has been presented for WLS IIR filter designs. A sequential design procedure is developed, in which the original design problem is transformed to an SOCP optimization problem using the SM scheme. The stability condition given by (3.29) is derived from the argument principle of convex analysis. However, in practice we cannot directly utilize this stability condition since it is also in a nonconvex form. As an attempt to tackle this difficulty, we first adopt an approximation technique similar to the SM scheme to transform the stability condition (3.29) into a quadratic inequality constraint, and then combine this approximate stability constraint with the sequential design procedure. It has been shown that if this sequential procedure is convergent and the regularization parameter α is appropriately selected, the

argument principle based stability constraint can finally guarantee the stability of designed IIR filters.

In Chapter IV, a sequential design method has been developed in the minimax sense. It has been demonstrated in (4.2) that the nonconvexity of the original minimax design problem is reflected by the constraint (4.2.b). By introducing a new polynomial $R(z) = Q(z)Q(z^{-1})$ and then replacing $\left|Q(e^{j\omega})\right|^2$ on the right-hand side of (4.2.b) by $R(e^{j\omega})$, we can transform (4.2.b) into a hyperbolic constraint. However, in order to maintain the equivalence between $R(z)$ and $Q(z)Q(z^{-1})$, we need to incorporate a nonconvex constraint $R(e^{j\omega}) = \left|Q(e^{j\omega})\right|^2$ for $\forall\omega \in [0, \pi]$ into (4.2). An SOCP design problem can be obtained by relaxing this quadratic equality constraint as $\left|Q(e^{j\omega})\right|^2 \leq R(e^{j\omega})$. By solving this relaxed design problem, we can achieve the lower and upper bounds of the optimal value of the original design problem (4.2). In practice, a real minimax solution can be attained by gradually reducing the discrepancy between $\left|Q(e^{j\omega})\right|^2$ and $R(e^{j\omega})$ over the whole frequency band $[0, \pi]$. We can achieve this goal through a sequential procedure developed in Section 4.1.3. The convergence of this sequential procedure is definitely ensured. In order to increase the convergence speed, a regularization term can be incorporated in the objective function of the design problem. The generalized positive realness based stability constraints (4.26) are used to ensure the stability of designed IIR filters.

Another minimax design method has been presented in Chapter V. A bisection search procedure is introduced to locate the minimum error limit. A feasibility problem with a fixed error limit is solved at each iteration of this bisection search procedure. In order to construct the feasibility problem, a symmetric matrix $\boldsymbol{X} = \overline{\boldsymbol{x}}\overline{\boldsymbol{x}}^T$ is introduced. By reformulating (4.2.b) in terms of $\overline{\boldsymbol{x}}$ and $\boldsymbol{X}$, we can transform the constraint (4.2.b) to a linear inequality constraint (5.11). The equality constraint $\boldsymbol{X} = \overline{\boldsymbol{x}}\overline{\boldsymbol{x}}^T$ can be further relaxed to $\boldsymbol{X} \succeq \overline{\boldsymbol{x}}\overline{\boldsymbol{x}}^T$ or, equivalently, $\boldsymbol{Z} \succeq 0$ where $\boldsymbol{Z}$ is defined by (5.13), such that the feasibility problem is in a convex form. It has been proved in Section 5.1.2 that if the final solution $(\overline{\boldsymbol{x}}, \boldsymbol{X})$ of the bisection search procedure satisfies $\operatorname{rank}\boldsymbol{Z} = 1$, the globally optimal design is attained. This condition can be used to detect the optimality of IIR filters designed by

the proposed method. In practice, however, we cannot always obtain rank-1 solutions. Therefore, the constraint $\text{rank}\,\boldsymbol{Z} = 1$ has to be incorporated. Unfortunately, this rank constraint is still nonconvex. As an attempt to tackle this difficulty, the regularization term $\text{Tr}\{\boldsymbol{Z}\}$ is introduced into the objective function of the SDP feasibility problem so as to drive many eigenvalues of $\boldsymbol{Z}$ to zeros. Another bisection search procedure needs to be deployed within the outer bisection search procedure to determine an appropriate regularization parameter. The stability of designed IIR filters can also be assured by the inner bisection search procedure.

The effectiveness of all the proposed design methods described in this dissertation has been validated by various simulation examples. The design performances have also been compared with some prevalent design methods. It has been demonstrated that the proposed design methods can achieve satisfactory designs in the WLS and minimax senses, respectively.

6.2 Further Study

All the design methods proposed in this dissertation are primarily devoted to tackle the nonconvexity and stability issues of IIR filter design problems. So far, the prevalent way to accomplish this purpose is to employ some approximation techniques to transform the original design problems to some simpler forms. For example, local approximation techniques, such as first-order Taylor series, can be used to achieve convex formulations of these design problems. In this dissertation, we prefer the convex relaxation techniques to local approximation techniques, since some more important information about optimal solutions can be simultaneously obtained. However, the remaining difficulty is that generally these relaxation techniques can only lead to approximate solutions rather than optimal designs. Thus, we still need to resort to some other approximation techniques to refine the design results. Apparently, if the relaxed design problems can be better defined, we can gain more information about optimal designs. Correspondingly, it is more possible to achieve optimal designs through the subsequent local search procedures. Following this idea, some more relationships between the original and relaxed design problems can be exploited to refine the convex formulations of the relaxed design

problems. Moreover, some special characteristics of the original design problems in time and/or frequency domains can also be used to screen out unqualified solutions from the enlarged feasible sets of the relaxed design problems.

Stability is another important issue which needs to be addressed in IIR filter design methods. In this dissertation, a sufficient and necessary condition, *i.e.*, (3.29), for the stability of designed IIR filters has been presented. The major difficulty of using this stability condition in practical designs is its nonconvexity, which is mainly incurred by the dependence of $\boldsymbol{G}(r, \boldsymbol{q})$ on denominator coefficients $\boldsymbol{q}$ and the infiniteness of $\boldsymbol{G}(r, \boldsymbol{q})$. In Chapter III, we adopted an approximation technique similar to the SM scheme to tackle these difficulties. The major concern about the approximate stability condition is that by introducing the approximation technique, at each iteration the approximate stability condition may be neither sufficient nor necessary. Although the stability of designed IIR filters can still be assured if the sequential design method is convergent and the regularization parameter is appropriately selected, generally speaking, sufficient conditions are more desirable in practical designs, since stable IIR filters can always be obtained by sufficient conditions even in nonsequential design methods. Such sufficient conditions should satisfy the following properties:

1. Such stability conditions can be readily incorporated into a variety of optimization-based design methods. In general, sufficient stability conditions in convex forms are most suitable for this purpose.

2. The feasible set defined by such stability conditions should be large enough. In other words, these sufficient conditions can approximate the sufficient and necessary condition (3.29) as well as possible.

Although some sufficient stability conditions, which satisfy the first requirement, have been developed so far, the stability domains defined by these conditions are much smaller than the real stability domains. Two illustrative examples have been given by Figs. 1 and 2 in [23], where the stability domains defined by the positive realness based stability condition (3.37), the Rouché's theorem based stability condition (3.41), and the generalized positive realness based stability condition (4.24) are compared with the real stability domains. It can be found that the feasible sets defined by these sufficient

conditions are much smaller than the real stability domains. Thereby, the optimal designs could be excluded from the feasible sets of the design problems, especially when they are close to the boundary of the real stability domains. Since the stability condition (3.29) is both sufficient and necessary, the real stability domains can be strictly defined by (3.29). Thus, we can exploit appropriate approximation and convex relaxation techniques to derive sufficient stability conditions.

REFERENCES

[1] T. W. Parks and C. S. Burrus, *Digital Filter Design*. New York: John Wiley & Sons, 1987.

[2] A. Antoniou, *Digital Filters: Analysis, Design, and Applications*, 2nd Ed. New York: McGraw-Hill, 2000.

[3] T. W. Parks and J. H. McClellan, "Chebyshev approximation for nonrecursive digital filters with linear phase," *IEEE Trans. Circuit Theory*, vol. CT-19, pp. 189-194, Mar. 1972.

[4] L. R. Rabiner, "Linear programming design of finite impulse response (FIR) digital filters," *IEEE Trans. Audio Electroacoust.*, vol. AU-20, pp. 280-288, 1972.

[5] K. M. Tsui, C. Chan, and K. S. Yeung, "Design of FIR digital filters with prescribed flatness and peak error constraints using second-order cone programming," *IEEE Trans. Circuits Syst. II*, vol. 52, pp. 601-605, 2005.

[6] L. R. Rabiner, N. Y. Graham, and H. D. Helms, "Linear programming design of IIR digital filters with arbitrary magnitude function," *IEEE Trans. Acoust. Speech, Signal Process.*, vol. ASSP-22, no. 2, pp. 117-123, Apr. 1974.

[7] C. Tseng, "Design of stable IIR digital filter based on least *p*-power error criterion," *IEEE Tans. Circuits Syst. I*, vol. 51, pp. 1879-1888, 2004.

[8] C. Tseng and S. Lee, "Minimax design of stable IIR digital filter with prescribed magnitude and phase responses," *IEEE Tans. Circuits Syst. I*, vol. 49, pp. 547-551, Apr. 2002.

[9] R. A. Vargas and C. S. Burrus, "On the design of L_p IIR filters with arbitrary frequency response," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Process.*, Salt Lake City, USA, vol. 6, May 2001, pp. 3829-3832.

[10] W.-S. Lu, "Design of stable IIR digital filters with equiripple passbands and peak-constrained least-squares stopbands," *IEEE Trans. Circuits Syst. II*, vol. 46, pp. 1421-1426, Nov. 1999.

[11] W.-S. Lu, S. Pei, and C. Tseng, "A weighted least-squares method for the design of stable 1-D and 2-D IIR digital filters," *IEEE Trans. Signal Process.*, vol. 46, pp. 1-10, Jan. 1998.

[12] W.-S. Lu, "Design of stable minimax IIR digital filters using semidefinite programming," in *Proc. IEEE Int. Symp. Circuits and Systems*, Geneva, Switzerland, vol. 1, May 2000, pp. 355-358.

[13] A. Jiang and H. K. Kwan, "IIR digital filter design with novel stability criterion based on argument principle," *IEEE Trans. Circuits and Systems I*, vol. 56, pp. 583-593, Mar. 2009.

[14] A. K. Shaw, "Optimal design of digital IIR filters by model-fitting frequency response data," *IEEE Trans. Circuits Syst. II*, vol. 42, pp. 702-710, Nov. 1995.

[15] A. K. Shaw, "Optimal Identification of Discrete-Time Systems from Impulse Response Data," *IEEE Trans. Signal Process.*, vol. 42, pp. 113-120, Jan. 1994.

[16] Y. C. Lim, J.-H. Lee, C. K. Chen, and R.-H. Yang, "A weighted least squares algorithm for quasi-equiripple FIR and IIR digital filter design," *IEEE Trans. Signal Process.*, vol. 40, pp. 551-558, Mar. 1992.

[17] A. Alkhairy, "An efficient method for IIR filter design," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Process.*, Adelaide, Australia, vol. 3, Apr. 1994, pp. 569-571.

[18] X. Chen and T. W. Parks, "Design of IIR filters in the complex domain," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, pp. 910-920, Jun. 1990.

[19] W.-S. Lu and T. Hinamoto, "Optimal design of IIR digital filters with robust stability using conic-quadratic-programming updates," *IEEE Trans. Signal Process.*, vol. 51, pp. 1581-1592, Jun. 2003.

[20] M. C. Lang, "Least-squares design of IIR filters with prescribed magnitude and phase responses and a pole radius constraint," *IEEE Trans. Signal Process.*, vol. 48, pp. 3109-3121, Nov. 2000.

[21] W.-S. Lu, "An argument-principle based stability criterion and application to the design of IIR digital filters," in *Proc. IEEE Int. Symp. Circuits and Systems*, Island of Kos, Greece, May 2006, pp. 4431-4434.

[22] W.-S. Lu, Y. Cui, and R. L. Kirlin, "Least *p*th optimization for the design of 1-D filters with arbitrary amplitude and phase responses," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Process.*, Minneapolis, USA, vol. 3, Apr. 1993, pp. 61-64.

[23] B. Dumitrescu and R. Niemistö, "Multistage IIR filter design using convex stability domains defined by positive realness," *IEEE Trans. Signal Process.*, vol. 52, pp. 962-974, Apr. 2004.

[24] R. Niemistö and B. Dumitrescu, "Simplified procedures for quasi-equiripple IIR filter design," *IEEE Signal Process. Letters*, vol. 11, pp. 308-311, Mar. 2004.

[25] X. Zhang, K. Suzuki, and T. Yoshikawa, "Complex Chebyshev approximation for IIR digital filters based on eigenvalue problem," *IEEE Trans. Circuits Syst. II*, vol. 47, pp. 1429-1436, Dec. 2000.

[26] A. G. Deczky, "Equiripple and minimax (Chebyshev) approximations for recursive digital filters," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-22, pp. 98-111, Apr. 1974.

[27] M. Ikehara, H. Tanaka, and H. Kuroda, "Design of IIR digital filters using all-pass networks," *IEEE Trans. Circuits Syst. II*, vol. 41, pp. 231-235, Mar. 1994.

[28] Tarczynski, G. D. Cain, E. Hermanowicz, and M. Rojewski, "A WISE method for designing IIR filters," *IEEE Trans. Signal Process.*, vol. 49, pp. 1421-1432, Jul. 2001.

[29] A. Jiang and H. K. Kwan, "Unconstrained IIR filter design method using argument principle based stability criterion," in *Proc. IEEE Int. Asia Pacific Conf. Circuits and Systems*, Macao, China, Dec. 2008, pp. 866-869.

[30] A. G. Deczky, "Synthesis of recursive digital filters using the minimum *p*-error criterion," *IEEE Trans. Audio Electroacoust.*, vol. AU-20, pp. 257-263, Oct. 1972.

[31] G. Cortelazzo and M. R. Lightner, "Simultaneous design in both magnitude and group-delay of IIR and FIR filters based on multiple criterion optimization," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-32, pp. 949-967, Oct. 1984.

[32] T. Chottera and G. A. Jullien, "A linear programming approach to recursive filter design with linear phase," *IEEE Trans. Circuits Syst.*, vol. CAS-29, pp. 139-149, Mar. 1982.

[33] H. K. Kwan and A. Jiang, "Recent advances in FIR approximation by IIR digital filters," ICCCAS 2006, vol. 1, pp. 185-190, Jun. 2006.

[34] N. Wong and C.-U. Lei, "IIR approximation of FIR filters via discrete-time vector fitting," *IEEE Trans. Signal Process.*, vol. 56, pp. 1296-1302, Mar. 2008.

[35] S. C. Chan, K. M. Tsui, and K. W. Tse, "Design of constrained causal stable IIR filters using a new second-order-cone-programming-based model-reduction technique," *IEEE Trans. Circuits Syst. II*, vol. 54, pp. 107-111, Feb. 2007.

[36] C. Xiao, J. C. Olivier, and P. Agathoklis, "Design of linear phase IIR filters via weighted least-squares approximation," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Process.*, Salt Lake City, USA, vol. 6, May 2001, pp. 3817-3820.

[37] H. Brandenstein and R. Unbehauen, "Weighted least-squares approximation of FIR by IIR digital filters," *IEEE Trans. Signal Process.*, vol. 49, pp. 558-568, Mar. 2001.

[38] L. Li, L. Xie, W.-Y. Yan, Y. C. Soh, "Design of low-order linear-phase IIR filters via orthogonal projection," *IEEE Trans. Signal Process.*, vol. 47, pp. 448-457, Feb. 1999.

[39] H. Brandenstein and R. Unbehauen, "Least-squares approximation of FIR by IIR digital filters," *IEEE Trans. Signal Process.*, vol. 46, pp. 21-30, Jan. 1998.

[40] S. Holford and P. Agathoklis, "The use of model reduction techniques for designing IIR filters with linear phase in the passband," *IEEE Trans. Signal Process.*, vol. 44, pp. 2396-2404, Oct. 1996.

[41] V. Sreeram and P. Agathoklis, "Design of linear-phase IIR filters via impulse-response Gramians," *IEEE Trans. Signal Process.*, vol. 40, pp. 389-394, Feb. 1992.

[42] B. Beliczynski, I. Kale, and G. D. Cain, "Approximation of FIR by IIR digital filters: an algorithm based on balanced model reduction," *IEEE Trans. Signal Process.*, vol. 40, pp. 532-542, Mar. 1992.

[43] A. Betser and E. Zeheb, "Reduced order IIR approximation to FIR filters," *IEEE Trans. Signal Process.*, vol. 39, pp. 2540-2544, Nov. 1991.

[44] R. Fletcher, *Practical Methods of Optimization*, 2nd Ed. Chichester, UK: John Wiley & Sons, 1987.

[45] S. G. Nash and A. Sofer, *Linear and Nonlinear Programming*. New York: McGraw-Hill, 1996.

[46] A. Antoniou and W.-S. Lu, *Practical Optimization: Algorithms and Engineering Applications*. New York: Springer, 2007.

[47] K. P. Chong and S. H. Żak, *Introduction to Optimization*. New York: John Wiley & Sons, 2008.

[48] D. G. Luenberger and Y. Yu, *Linear and Nonlinear Programming*, 3rd Ed. New York: Springer, 2008.

[49] K. E. Steiglitz and L. E. McBride, "A technique for the identification of linear systems," *IEEE Trans. Automat. Contr.*, vol. AC-10, pp. 461-464, Oct. 1965.

[50] R. Fletcher and M. J. D. Powell, "A rapidly convergent descent method for minimization," *Computer Journal*, vol. 6, pp. 163-168, 1963.

[51] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[52] J. M. Borwein and A. S. Lewis, *Convex Analysis and Nonlinear Optimization*. New York: Springer, 2000.

[53] H. Hindi, "A tutorial on convex optimization," in *Proc. American Control Conf.*, Boston, USA, vol. 4, Jul. 2004, pp. 3252-3265.

[54] H. Hindi, "A tutorial on convex optimization II: duality and interior-point methods," in *Proc. American Control Conf.*, Minneapolis, USA, Jun. 2006, pp. 686-696.

[55] Z.-Q. Luo and W. Yu, "An introduction to convex optimization for communication and signal processing," *IEEE Journ. Sel. Areas Commun.*, vol. 24, pp. 1426-1438, Aug. 2006.

[56] S.-P. Wu, S. Boyd, and L. Vandenberghe, "FIR filter design via semidefinite programming and spectral factorization," in *Proc. 35th Conf. Decision and Control*, Kobe, Japan, vol. 1, Dec. 1996, pp. 271-276.

[57] W.-S. Lu, "Design of nonlinear-phase FIR digital filters: A semidefinite programming approach," in *Proc. IEEE Int. Symp. Circuits and Systems*, Orlando, USA, vol. 3, May 1999, pp. 263-266.

[58] S. C. Chan and K. M. Tsui, "On the design of real and complex FIR filters with flatness and peak error constraints using semidefinite programming," in *Proc. IEEE Int. Symp. Circuits and Systems*, Vancouver, Canada, vol. 3, May 2004, pp. 125-128.

[59] J. O. Coleman and D. P. Scholnik, "Design of nonlinear phase FIR filters with second-order cone programming," in *Proc. Midwest Symp. Circuits and Systems,* Las Cruces, USA, vol. 1, Aug. 1999, pp. 409-412.

[60] W.-S. Lu, "A unified approach for the design of 2-D digital filters via semidefinite programming," *IEEE Trans. Circuits Syst. I*, vol. 49, pp. 814-826, Jun. 2002.

[61] H. D. Tuan, T. T. Son, H. Tuy, and T. Nguyen, "New linear-programming-based filter design," *IEEE Trans. Circuits Syst. II*, vol. 52, pp. 276-281, May 2005.

[62] S. C. Chan, H. H. Chen, and K. S. Pun, "The design of digital all-pass filters using second-order cone programming (SOCP)," *IEEE Trans. Circuits Syst. II*, vol. 52, pp. 66-70, Feb. 2005.

[63] C.-C. Tseng, "Design of IIR digital all-pass filters using least *p*th phase error criterion," *IEEE Trans. Circuits Syst. II*, vol. 50, pp. 653-656, Sep. 2003.

[64] N. Levinson and R. M. Redheffer, *Complex Variables*. San Francisco: Holden Day, 1970.

[65] C. R. Wylier and L. C. Barrett, *Advanced Engineering Mathematics*, 6th Ed. New York: McGraw-Hill, 1995.

[66] J. F. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optim. Math. Soft.*, vol. 11-12, pp. 625-653, 1999.

[67] K. Preuss, "On the design of FIR filters by complex Chebyshev approximation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, pp. 702-712, May 1989.

[68] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret, "Applications of second-order cone programming," *Linear Algebra and its Applications*, vol. 284, pp. 193-228, 1998.

[69] B. Dumitrescu, *Positive Trigonometric Polynomials and Signal Processing Applications*. New York: Springer, 2007.

[70] Y. Nesterov and A. Nemirovskii, *Interior Point Polynomial Methods in Convex Programming*. Philadelphia, Philadelphia: SIAM, 1994, vol. 13, Studies in Applied Mathematics.

[71] A. S. Nemirovski and M. J. Todd, "Interior-point methods for optimization," *Acta Numerica*, vol. 17, pp. 191-234, May 2008.

[72] S. Boyd, L. E. Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequality in System and Control Theory*. Philadelphia: SIAM, .1994, vol. 15, Studies in Applied Mathematics.

[73] M. Fazel, H. Hindi, and S. Boyd, "Rank minimization and applications in system theory," in *Proc. American Control Conf.*, Boston, USA, vol. 4, Jul. 2004, pp. 3273-3278.

[74] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM Review*, vol. 38, pp. 49-95, Mar. 1996.

VITA AUCTORIS

Aimin Jiang was born in Taixing, China in 1979. Now, he is a Ph.D. candidate at the Department of Electrical and Computer Engineering, University of Windsor, Ontario, Canada. He has received his M.E. and B.E. both from the Department of Electronic and Information Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu, China, in 2004 and 2001, respectively.

In 1999, Aimin Jiang received the first prize with the other two teammates in China Undergraduate Mathematical Contest in Modeling. From 1997 to 2001, he received undergraduate student scholarship.

Owing to his outstanding performance during the undergraduate study, Aimin Jiang was recommended for admission to the graduate study. He was working on the imaging algorithms for spotlight synthetic aperture radar (SAR).

After receiving his M.E., he joined Fortemedia (Nanjing) Ltd., Nanjing, China as a senior software engineer. He developed pre-processor, assembler, and linker system for the software development platform of digital signal processor. He also developed C libraries of acoustic signal processing algorithms for ease of the research and software development.

From 2006 to 2009, Aimin Jiang received the Doctoral Tuition Scholarship of University of Windsor. His research interests include optimization and its applications to digital signal processing and communications.