ON OPTIMIZATION METHODS FOR SPEECH SIGNAL PROCESSING

HE Qi

MPhil.

The Hong Kong Polytechnic University

2019

The Hong Kong Polytechnic University
Faculty of Applied Science and Textiles
Department of Applied Mathematics

On Optimization Methods for Speech Signal Processing

HE Qi

A thesis submitted in partial fulfilment of the requirements for
the degree of Master of Philosophy
July 2019

Certificate of Originality

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

________________(signed)

He Qi (name of student)

Dedication

Dedicate to my parents, those beloved, and the friends who encouraged me to research.

Abstract

This thesis is concerned with optimization used in speech signal processing, including filters design problem, equalizer, and beamformer design problems. All these three problems play an essential role in signal speech processing.

Firstly, the minimax design of infinite impulse response(IIR) filter is considered. Partial fraction method is applied to decompose the transfer function into a sum of low order fractions; as a result, the minimax IIR filter design problem can be formulated with the stability condition as constraints. Then, among different possible decompositions of the transfer function, it can be proved that the second order decomposition can always achieve a good approximation to the optimal solution. Based on the second order decomposition formulation, several numerical experiments have been conducted, and better IIR filters can be designed using the proposed method compared with existing methods in the literature.

Secondly, a fixed equalizer aiming at noise reduction is designed by optimizing objective measures. There are many objective quality measures in speech quality testing, and one of the most popular ones is PESQ, and another latest measure is STOI. In this part, both of these two measures are considered as optimization criteria when designing weights of equalizers. To enhance robustly, there is a pre-training part of this method, and the average weights of the training section achieve as the coefficients of the fixed equalizer.

Thirdly, array gain optimization methods are considered in speech enhancement. Algorithms, LS, and SNR, are used to find optimal weights, and the improved performance using these two methods are presented. From the results, we could see that LS concentrate more on distortion control comparing with SNIR method, but SINR performs better in noise suppression than LS. Besides, in the experimental part, the results consist of two parts, in-

cluding the real data recorded by sensors and the created signals calculated by a transfer function.

Publications arising from the Thesis

Here lists the publication arising from this thesis:

- Qi, H., Feng, Z. G., Yiu, K. F. C., Nordholm, S. (2018). Optimal Design of IIR Filters via the Partial Fraction Decomposition Method. IEEE Transactions on Circuits and Systems II: Express Briefs.

Here lists the conference attending:

- the 7th International Conference on Optimization and Control with Applications (OCA2018).

Acknowledgements

The endeavor of research is a non-isolated process, and I am grateful to those who have supported me during my M.Phil study. I hereby acknowledge the assistance they provided.

First and foremost, I would like to offer my sincere appreciation to my supervisor, Prof. Yiu Ka-Fai, Cedric, for his guidance, encouragement, and continued patience during the whole procedure of my study. Without his unwavering support, I couldn't finish this work. Throughout the two years, I have learned a lot about this research field because of his patient guidance, insightful ideas, and invaluable discussions. Moreover, thank you taught me to be patient and confident.

My appreciation also goes to Prof. Sven Nordholm, Dr. Zhiguo Feng, Dr. Siow Yong Low, for their academically supports and guidance during my M.Phil study.

I am thankful for my parents and friends for their encouragement and comfort, accompanying me during my study period. Although some of them are not by my side, they do give me a lot of care, love, and attention. Thank you so much, and I will love you now and forever.

I give a special thank to fellow researchers at Hong Kong Polytechnic University for their support.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Speech Signal Processing

Speech signal processing is the study of speech signals and the processing methods of signals. It is also defined as the intersection of digital signal processing and natural language processing, as the speech signals are usually processed in a digital representation. This study contains the acquisition, manipulation, storage, transfer, and output of speech signals. Speech signal processing is the study of speech signals and their processing methods. It is also defined as the intersection of digital signal processing and natural language processing, as the speech signals are usually processed in a digital representation. The central processor of this study contains the acquisition, manipulation, storage, transfer, and output of speech signals. The main tasks of speech processing include speech recognition, speech synthesis, speaker recognition, speech coding, speech enhancement, etc.. This thesis is concerned with three essential problems in speech signal processing solved by optimization including IIR filter design problem, equalizer and beamformer design problems.

IIR filter design problem is quite crucial in Digital Signal Processing (DSP). Generally, two primary types of digital filters are wildly used, including Infinite-impulse-response(IIR) digital filter and finite impulse response(FIR). Because IIR filters can achieve a given filtering characteristic using less memory and calculations than a similar FIR filter, they are used in a wide range of applications where high selectivity and efficient processing of discrete signals are desirable. Hence, designing IIR filter coefficients became an essential problem.

Equalizers and beamformers are always used for speech enhancement. In the real environment, noise contamination is inevitable, which degrades the performances of the voice communication system. This results in the popularity of speech enhancement, which aims at improving the quality or the intelligibility of speech. It can be used in various domains such as hearing aids, assistive listening devices, speech recognition system, voice communication systems, etc.

1.2 Filter Design Problem

1.2.1 Background

The ability to communicate is the key to our society, and it signals that carry the information from one point to another that makes communications work. Signal processing is concerned with the representation, manipulation, and transformation of signals and the information they carry. [12] Digital signal processing, known as DSP, is a science of processing the information which is presented in digital form. As the development of this field, many efficient algorithms are used in various applications, such as the array processing, statistical spectral analysis, and image, multimedia processing. It also plays a vital role in speech signal processing, like the Equalizer Design Problem and Beamformer design problem, which would be discussed in the following subsection.

In signal processing, two types of filters are wildly used, Infinite Impulse Response Filters (IIR) and Finite Impulse Response Filters (FIR), respectively. FIR is the non-recursive filter, and the general transfer function of FIR filters is given as

$$H(z) = \frac{p(z)}{q(z)} = \sum_{k=0}^{n} p_n z^{-k} \tag{1.1}$$

while IIR is recursive filter which could be given as

$$H(z) = \frac{p(z)}{q(z)} = \frac{\sum_{k=0}^{n} p_n z^{-k}}{1 + \sum_{k=1}^{m} q_k z^{-k}} \tag{1.2}$$

FIR filters are wildly used in digital communication systems, noise control systems, etc. There are three advantages of FIR for the popularity of the FIR. Firstly, it is easy and efficient to adjust the filter coefficients. Second, the stability condition is easily guaranteed by ensuring the coefficients of filters are bounded. Third, FIR filters always have excellent performance in satisfying the design criteria. As for the IIR filters, they always performed better for a given filter order, which is the number of filter coefficients. However, there is a problem that instability may affect the filter general numerical sensitivity. Although there are some difficulties in designing, IIR filters have many applications where requires a recursive filter, such as in echo canceller.

1.2.2 FIR Design Methods

There are many methods of FIR design problems. Generally, there are four following basic steps to satisfy the desired specifications. Firstly, to approximate the ideal filter, we need to convert the desired specifications into a mathematical formulation according to the design technique. Second, solve the approximation problem to design the filter coefficients by minimizing or maximizing a performance measure. Third, choose a proper structure to realize filters and determine filter coefficients, including the input, output, intermediate variable word lengths. Finally, verify the performance of the filters by simulation and some other hardware for digital filters. [39]

One conventional method is applying the technique of windowing where the desired frequency response is translated into a Fourier series and truncated into the desired length. [24],[25],[4] Then, the output filter minimizes the least square between the filter response and the desired response. However, the maximum absolute value of the error of the filter designed by this method is quite large because of the Gibbs phenomenon. To get over this shortcoming, people seek to use smooth time-limited window by multiplying the coefficients of the Fourier series instead of just truncating the infinite Fourier series. Some windows, such as Hanning window,[4] Hamming window,[4] Kaiser window[25] and the Dolph-Chebyshev window [14] are popular in this method. Another wild used technique in designing FIR filter is a frequency sampling method, which was first raised by Jordan and Gold [11] and improved by Rabiner el al [37]. The general idea of this method is fixing most coefficients of Discrete Fourier Transform (DFT) and leaving those lying in the transition bands, as one filter can achieve the specified frequency response by fixing the passband and stopband. This problem can be considered as a linear programming problem with constraints.

In the following development, Herrmann [15] was originally proposed as an optimal method in FIR design. Based on Herrmann, Hofstetter et al. [16] developed an algorithm known as "reminiscent of the Remez exchange algorithm," which effectively solve the nonlinear equations. However, there is a drawback that it is not possible to specify a prior location of the stopband and passband cutoff frequencies. Then, a method [34] formulated the lowpass approximation of the desired response was proposed, but this method can be quite slow when computing the best Chebyshev approximation. Thus, Rabiner [38] proposed an alternative method for this complex calculating procedure.

1.2.3 IIR Design Methods

The conventional method to design an IIR filter is approximating the magnitude specification and use linearize the phase response by using allpass equalizers. Recently, many popular methods consider the simultaneous approximation of both phase and magnitude characteristic. There are some efficient and typical methods, such as Thiran's all-pole filter, allpass filters' applications, and Remez algorithm. Digital fractional delay filters provide a useful building block that can be used for fine-tuning the sampling instants, i.e., implement the required bandlimited interpolation.

Thiran [42] develops one traditional method; he derived the all-pole transfer function that approximates a maximally flat sense. Another method called the Remez exchange algorithm, also known as Paks-McClellan method, was proposed by Parks and McClellan. [19] This algorithm mainly contains two steps. Firstly, by solving a series of linear equations, determine candidate filter coefficients from alternation frequencies. Secondly, determine the candidate alternation frequencies according to the candidate filter coefficients from the first step. IIR design problem can also be considered as a nonlinear optimization problem. It proposed by A.G.Deczky [6], and he decomposed the transfer function in a cascade of second-order sections and then minimize the error function which involved a sum of error in the passband and stopband magnitude and group delay. In the second section, it proposed a new method to design IIR filters by an optimal algorithm.

1.3 Equalizer Design Problem

The equalizer design problem can be described as the following:

Suppose that F and F^* are known digital filters, a physical system is molded by F and the desired system is molded by F^*. The equalizer problem is constructing an equalizer E, which could approximate F^* by composing with F.

Generally, F can be obtained from prior knowledge according to the system characteristics or input, and output of sensors in the system and the desired filter can be known by certain performance criterions. This problem can be classified into two categories depending on whether the filters are linear. The linear filter situation is well studies. In the nonlinear system, Volterra filters are chosen, because they provide suitable mathematical models for various nonlinear systems.

Equalizer technique is an important technique in speech signal processing and has a wild application such as in speech enhancement and echo cancellation. Algorithms based on Volterra filter equalizer is prosed to solve the echo cancellation problem in [1] and [17]. These methods are typically minimizing the mean-square error of the equalization. However, this is a nonconvex problem. Thus it is difficult to grantee the convergence of the optimal solution. As for the speech enhancement problem, fixed or adaptive equalizers are considered. The input signal is divided into subbands which are individually and temporally weighted. According to whether the weighting of the subbands is changeable, they can be defined as fixed and adaptive equalizers. The fixed equalizer is easy to realize in practice, but it is not robust. The adaptive equalizer is active, only when the speech exists in a particular subband. This method is flexible and robust and efficient in various noise situations.

When designing the fixed or adaptive equalizer, filter-bank is always used. Suppose $s(k)$ is the input signal, it can be decomposed into many subbands signals by using bandpass filters

$$y_i(k) = \sum_{k=0}^{n-1} x(k-n) \cdot h_i(k), \; i = 0, 1, \cdots, N \tag{1.3}$$

where $h_i(k)$ is impulse response with length n, N is the length of bandpass filters and $y_i(k)$ is a block of data. For the aim of audio equalization, gain multipliers g_i are applied to every subband signals and the output singal is achieved as

$$y(k) = \sum_{i=0}^{N} g_(i) y_i(k) \tag{1.4}$$

In this situation, the value of gains is limited from zero to one and depend on the transient signal to noise ratios. As this process is in the time domain, there is no sample rate decimation and Discrete Fourier Transformation (DFT) in the signal path. The gains can be designed in the either from $y_i(k)$ or original signal $x(k)$. However, the shortcoming of this system is the complexity, because a great number of separate filters are necessary.

Also, some procedures in the frequency domain are used, and one conventional method is spectral subtraction to reduce noise. This method uses short-term spectral analysis and spectral-domain filtering. The main idea of spectral subtraction is that the DFT coefficients are multiplied by real values. The gains are designed according to a certain rule. There are various methods about the calculation of gains, for example in noise reduction, some ”spectral subtraction rules” are applied. [5] [33] Finally, the output spectrum is translated into the time domain by inverse discrete Fourier transformation.

1.4 Beamformer Design Problem

Beamforming design is a popular task in signal processing with wild applications in sonar, astronomy, acoustics, seismology, communications, radar, and medical imaging. It can be used from speech enhancement to source detection. This method improves a multichannel recording, often obtained from a microphone array with four to eight channels, by reconstructing the signals to converge better results. Generally, there is common ground in the filter technique and delay technique. Many well-known filter technique in signal processing is applied, such as Winer filters and Kalman filters.

There are three interfering paraments, reverberations, measurement noise, and additional signal in most methods. The reverberation is correlated with microphones and source signals. The noise can be considered as Gaussian Noise as it is supposed to be uncorrelated with the microphones. The additional sources are correlated between the microphones but independent with the source of interest. Therefore, these three kinds of disturbing signals have to be considered separately according to different situations. Besides, according to the aim of the beamformer, it can be divided into diverse acoustic signal processing techniques, signal direction-of-arrival estimation, noise reduction, echo cancellation and suppression, source separation and speech dereverberation, respectively.

In the task of sound source localization, as there is a difference of arrival times of the microphones, we could find a place most likely to be the source location by finding out a single value of angle most relative to the normal direction of the microphone array. In this process, the most important property is the phase of the signal, and it is different at each microphone. From the difference of phase among the sensors, the time delay of the signal can be calculated and the direction of arrival (DOA) can be obtained. Some traditional methods for this problem contain diagonal loading,[10] constrained minimum variance beamforming and some other combined approaches.

Source separation for acoustic signals consists of recovering the original sound source signals when a mixed signal with unknown coefficients is observed in a sensor. This technique has applications in speech recognition system and hands-free communication systems. A single channel and multi-channel inputs can achieve source separation. In single-channel input, various methods are proposed, such as tracking a formant structure [35], auditory scene analysis approach [45], the organization method for hierarchical perceptual in the auditory domain [26]. While in multichannel inputs of source separation, a method based on array speech signal processing is an effective technique. In this method, firstly the estimation of DOA of the signal is needed, and then according to using the directivity of the array, each of the source signals is separated. During this procedure, some conventional microphone arrays technique, such as the delay-and-sum and the adaptive beamformer, are wildly used in source separation.

In the situation where the location of the microphone and the user has a distance, the microphone may pick up the undesirable reflections because of the borders of an enclosure when in a car compartment or a room. As the reflections propagate in changeable paths due to many factors such as the environment changes, movements of the microphones, volume changes of the speakers and so on, the electro-acoustic circuit can be unstable and even produce highly undesirable howling. In many communication systems, echo cancellation is required as people may get annoyed when hearing the delayed voice of themselves, typically used in full-duplex communication environments. In some traditional methods, an adaptive filter is applied to cancel echo by estimating the echo according to a reference signal. This method always can cancel the echo, which is correlated to the reference signal. However, as for some nonlinear echo or high volume echo, this technique can not handle, and even in some stable condition, it may also cause distortion when speech signal at low signal-to-noise ratio. To further solve this problem, the microphone array is applied. This technique is efficient in canceling echo.

Signals recorded in indoor places, such as in a conference hall or a room, are always reverberated, as there are reflections on the room walls and other obstacles from the speech to the sensors. This reflection reduces the quality of the speech signal and intelligibility. Because of the existence of the reverberated speech, some subsequent processing of the speech signal might be rendered useless, like automatic speech recognition and speech coding. Although some dereverberation methods base on single microphone technique, the most successful one chooses multichannel measurements. On beamformer methods, the related transfer function of acoustic is estimated to reduce reverberation, when prior knowledge is known. This method is quite efficient and robust to small speaker movements. However, if the prior knowledge is not given, these methods can not perform well, and then some methods based on subspace and null subspace estimation are proposed.

Another big problem is denoising and noise reduction also known as speech enhancement. The aim of the speech enhancement is improving the quality of speech and reducing the level of noise in noisy environments. Because of the wild application of it, there are various kinds of methods when solving this problem, such as spectral subtraction methods, statistical spectral estimation, digital short-time Fourier analysis, harmonic model using comb filtering, Kalman filter, subspace methods, etc.. A simple method is using the Delay-and-Sum technique by estimating the uncorrelated noise obtained from different microphones, especially the DOA is known. By the knowledge DOA, the time delay can be calculated. If the time-shifted signals are averaged, the source signal can be enhanced while the noise can be reduced. In this method, the more microphones there are, the better the beamformer performs. [36] Another technique, known as 'adaptive arrays' also applied to solve the problem, which achieves a severe reduction of unwanted signal. Various approaches have been proposed in [8], [9] and optimal criterions, like minimizing the mean-square error and maximizing the SNR, are applied.

Chapter 2

Optimal Design of IIR Filters

2.1 Introduction

In general, two primary types of digital filters are wildly used in Digital Signal Processing, namely, Infinite-impulse-response(IIR) digital filter and finite impulse response(FIR). Because IIR filters can achieve a given filtering characteristic using less memory and calculations than a similar FIR filter, they are used in a wide range of applications where high selectivity and efficient processing of discrete signals are desirable[18]. Hence, designing IIR filter coefficients became an essential problem. If the parameters of the IIR such as magnitude and phase or group delay are known, an ideal frequency response by a stable IIR digital filter can be obtained, and the design problem of an IIR filter is to approximate the ideal digital filter which can be transformed into a nonconvex optimization problem under various optimization criteria. However, it is a more challenging task to design the coefficients of IIR filter comparing with FIR filter design [7] because of the following two reasons. Firstly, IIR is nonconvex, which result in the difficulty in attaining the globally optimal solution. Another challenge is the stability of IIR filters due to the existence of denominators in the transfer function. Nevertheless, many people have suggested different ways to solve the problem, but there are still some drawbacks in those current references.

To solve the nonconvex problem, many optimization methods have been applied. One of the main ideas is based on the Stieglitz-McBride (SM) scheme [23] to transform the nonconvex problems into convex problems under different criteria, such as the linear programming (LP)[44], quadratic programming (QP), [31, 43, 46] or second-order cone programming (SOCP) [21, 20]. Besides, Gauss-Newton (GN) method was also applied to solve the nonconvex problem in [44], where a nonconvex constrained least-squares problem was converted into convex QP problems by GN method, and a GN method is proposed under the weighted least squares criteria in [29]. In terms of the stability problem, many people proposed many methods to solve it, such as the iterative Lyapunov inequality constraints[30], a general positive realness constraint on denominator perturbations [31],[43],[44],[46]. However, there is a common drawback that all of these methods above are sufficient but not necessary, which may exclude some optimal solutions out. If we transform the transfer function into a form with factorized denominators by cascaded second-order factors, we can consider the necessary and sufficient stability condition as the stability-triangle-based stability conditions[2]. Then, the stability problem can be described as a set of linear inequality constraints on the denominator coefficients[32],[28].

This section presents a new algorithm because of the IIR filter design problem by designing the placement of poles. For the design problem, we choose a complex weighted error as an objective function and select the stability requirement as constraints. Then, according to different cases of poles, we decompose the transfer function into a sum of partial low order fractions. Then, for each decomposition, the parameters in the fractions together with the poles are the decisions vector to be optimized, where the stability constraints can be set up easily. Then, the optimal solution can be obtained by comparing all the optimal solutions of subproblems for every decomposition. Furthermore, we prove that the decomposition of quadratic denominators can always achieve the optimal solution. Then, we only consider the quadratic denominator decomposition method and this problem have been simplified. By decomposition, we translate the transfer function into a sum of low order factors and then apply the triangular stability condition. Comparing with the other existing methods, the method proposed can solve the stability problem and nonconvex problem well, which guarantee a better performance.

This section is organized as follows. Section 2 gives the main issue which has to be solved in this section. The first part of section 3 shows fraction decomposition and the second part transform the problem into several subproblems based on the decomposition, and the three-part simplify it into only one subproblem. In the fourth section, we present five examples with results of optimization for performance evaluation.

2.2 Problem Statement

The transfer function of an IIR digital filter in direct form is denoted by

$$H(z) = \frac{p(z)}{q(z)} = \frac{\sum_{k=0}^{n} p_n z^{-k}}{1 + \sum_{k=1}^{m} q_k z^{-k}} = \frac{\mathbf{p}^T \psi_n(z)}{\mathbf{q}^T \psi_m(z)} \tag{2.1}$$

where $\mathbf{p} = [p_0, p_1 \ldots, p_n]^T$, $\mathbf{q} = [q_0, q_1, q_2 \ldots, q_m]^T (q_0 = 1)$, $\psi_K(z) = [1, z^{-1}, z^{-2} \ldots, z^{-K}]$, and the superscript T denotes the transpose of a vector. We define $\mathbf{y} = [\mathbf{q}^T, \mathbf{p}^T]^T$ and all the coefficients of the filter are real valued. A weighted complex error is denoted by

$$E(\omega) = W(\omega)|D(\omega) - H(e^{j\omega})|^2, \forall \omega \in \Omega \tag{2.2}$$

where $D(\omega)$ is a prescribed ideal frequency response over $[0, \pi]$, $W(\omega)$ is a given nonnegative weighting function, and Ω is the union of frequency bands of interest within $[0, \pi]$. Our purpose is to find a vector $\mathbf{y}$ such that the maximum weighted error is minimized. The design problem of an IIR digital filter in the minimax sense can be strictly expressed as

$$\min_{\mathbf{y} \in S} \max_{\omega \in \Omega} E(\omega), \tag{2.3}$$

Where S is the stability domain of the filter.

In general, we can assume that $n \geq m$. Then by multiplying z^n in the denominator and numerator, we have

$$H(z) = z^{m-n} \bar{H}(z),$$

where $\bar{H}(z) = \frac{p^T \bar{\psi}_n(z)}{q^T \psi_m(z)}$, where $\bar{\psi}(k) = [z^k, \cdots, 1]$. Then, the problem (3) is equivalent to

$$\min_{\mathbf{y} \in S} \max_{\omega \in \Omega} E(\omega) = W(\omega)|\bar{H}(e^{j\omega}) - \bar{D}(\omega)|^2 \tag{2.4}$$

where $\bar{D}(\omega) = z^{n-m} \bar{D}(\omega)$. However, it would be challenging to find the stability domain if we directly optimize the coefficients of IIR filter, that is, for a vector $\mathbf{q}$, it is difficult or complicated to check whether $q(z)$ is stable.

2.3 Decomposition Method

For a vector $\mathbf{y}$, it is complicated to verify the stability of $H(z)$. Thus we decompose $H(z)$ into a sum of partial fractions with low order.

2.3.1 Fraction Decomposition

Decomposition of rational polynomial depends on the poles of the numerator. Then, for the polynomial where the coefficients are real numbers, we have the poles in the following situations.

(i) Single Real Pole

If there is only one single real pole z_0, then the decomposition must have a term $\frac{A}{z-z_0}$. If there are at least two single real poles z_1 and z_2, the decomposition have two terms $\frac{A_1}{z-z_1}$ and $\frac{A_2}{z-z_2}$. We can rewrite it as

$$\frac{cz+d}{z^2+az+b} = \frac{A_1}{z-z_1} + \frac{A_2}{z-z_2} \tag{2.5}$$

where $a = -z_1 - z_2$, $b = z_1 z_2$, $c = A_1 + A_2$, $d = -A_1 z_2 - A_2 z_1$.

(ii) Double Real Pole

If there is a double real pole z_0, then the decomposition must have the terms $\frac{A_1}{z-z_0} + \frac{B_0}{(z-z_0)^2}$, this can be rewritten as

$$\frac{cz+d}{z^2+az+b} = \frac{A_1}{z-z_0} + \frac{B_0}{(z-z_0)^2} \tag{2.6}$$

where $a = -2z_0$, $b = z_0^2$, $c = A_1$, $d = B_0 + A_1 z_0$.

(iii) Single Complex Pole

Note that the coefficients of $A(z)$ are real, the complex pole appear conjugately, which means the decomposition must contain the terms $\frac{A}{z-(\alpha+\beta j)}$ and $\frac{A}{z-(\alpha-\beta j)}$. By combining the conjugate poles, we can get the combined part:

$$\frac{cz+d}{z^2+az+b} = \frac{A}{z-(\alpha+\beta j)} + \frac{A}{z-(\alpha-\beta j)} \tag{2.7}$$

where $a=-2\alpha$, $b=\alpha^2+\beta^2$, $c=2A$, $d=-2A\alpha$.

(iv) Repeated Pole

If there exists repeated real pole z_0 of order $k(k \geq 3)$, then except the terms in (i) and (ii), there are serveral terms in the decomposition of filters

$$\frac{A_1}{(z-z_0)^k} + \cdots + \frac{A(n-2)}{(z-z_0)^3}$$

Similarly, if there exists repeated complex poles of order $k(k \geq 2)$, then there are several terms in the decomposition as follows

$$\frac{c_1 z + d_1}{(z^2+a_1 z+b_1)^n} + \cdots + \frac{c_n z + d_{n-1}}{(z^2+a_{n-1}z+b_{n-1})^2} \tag{2.8}$$

Hence, we can decompose $H(z)$ in the following possible cases:

(a) The multiplicities of all the poles are in the cases of (i), (ii), (iii).

In this case, the function $\bar{H}(z)$ can be decomposed into a uniform equation

$$\bar{H}(z) = \sum_{l=0}^{n-m} c_l z^l + \sum_{l=1}^{m/2} \frac{b_{l1}z+b_{l2}}{z^2+\alpha_l z+\beta_l} \tag{2.9}$$

when m is even, and

$$\bar{H}(z) = \frac{a_1}{z-d_1} + \sum_{l=1}^{(m-1)/2} \frac{b_{l1}z+b_{l2}}{z^2+\alpha_l z+\beta_l} + \sum_{l=0}^{n-m} c_l z^l \tag{2.10}$$

when m is odd.

(b) If there exists an k-repeated real pole d_0 of $\bar{H}(z)(k \geq 3)$. Besides the pole d_0, there are $m-k$ times of the other poles. Then, if the $m-k$ times of the poles are the case of (a) above, the decomposition is also the same. Hence, the decomposition is

$$\bar{H}(z) = \sum_{k=0}^{n-m} c_k z^k + \bar{H}_1(z) + r(z), \tag{2.11}$$

where

$$\bar{H}_1(z) = \sum_{l=1}^{m-k} \frac{b_{l1} z + b_{l2}}{z^2 + \alpha_l z + \beta_l} \tag{2.12}$$

when $m-k$ is even, and

$$\bar{H}_1(z) = \sum_{l=1}^{(m-k-1)/2} \frac{b_{l1} z + b_{l2}}{z^2 + \alpha_l z + \beta_l} + \frac{a_1}{z-d_1} \tag{2.13}$$

when $m-k$ is odd, and $r(z)$ is given by

$$r(z) = \frac{b_1 z + b_2}{z^2 - 2dz + d^2} + \sum_{l=3}^{k} \frac{b_k}{(z-d)^l}. \tag{2.14}$$

(c) If there are k-repeated complex conjugate poles z_0 and $\bar{z}_0$ of $A(z), (k \geq 2)$. Besides, the poles z_0 and $\bar{z}_0$, there are $m-2k$ times of the poles, which is similar to the situation (b). The decomposition is the same as (2.11), where $\bar{H}(z)$ is shown as (2.12) or (2.13) and $r(z)$ is given by

$$r(z) = \frac{b_{11} z + b_{12}}{(z^2 + 2\alpha z + \beta^2)^k} + \sum_{l=1}^{k} \frac{b_{2l-1} z + b_{2l}}{(z^2 + 2\alpha z + \beta)^l}. \tag{2.15}$$

In order to make the decomposition procedure more clear, we give the following two examples. For example, if $n = 15, m = 4$, where m is even, we consider the following decompositions:

$$\text{(1)}\ \bar{H}(z) = \sum_{i=0}^{11} c_i z_i + \sum_{i=1}^{2} \frac{b_{i1} z + b_{i2}}{z^2 + \alpha_i z + \beta_i}$$

$$\text{(2)}\ \bar{H}(z) = \sum_{i=0}^{11} c_i z_i + \frac{a_{11}}{z-d_1} + \frac{a_{21} z + a_{22}}{(z-d_2)^2} + \frac{a_{23}}{(z-d_2)^3}$$

$$\text{(3)}\ \bar{H}(z) = \sum_{i=0}^{11} c_i z_i + \frac{a_{11} z + a_{12}}{(z-d_1)^2} + \frac{a_{13}}{(z-d_1)^3} + \frac{a_{14}}{(z-d_1)^4}$$

$$(4)\ \bar{H}(z) = \sum_{i=0}^{11} c_i z_i + \frac{b_{11}z + b_{12}}{z^2 + \alpha_1 z + \beta_1} + \frac{b_{13}z + b_{14}}{(z^2 + \alpha_1 z + \beta_1)^2}$$

If $m = 5, n = 10$, where m is odd, we consider the following decompositions:

$$(1)\ \bar{H}(z) = \sum_{i=0}^{5} c_i z_i + \sum_{i=1}^{2} \frac{b_{i1}z + b_{i2}}{z^2 + \alpha_i z + \beta_i} + \frac{a_{11}}{z - d_1}$$

$$(2)\ \bar{H}(z) = \sum_{i=0}^{5} c_i z_i + \frac{b_{11}z + b_{12}}{z^2 + \alpha_1 z + \beta_1} + \frac{a_{11}z + a_{12}}{(z - d_1)^2} + \frac{a_{13}}{(z - d_1)^3}$$

$$(3)\ \bar{H}(z) = \sum_{i=0}^{5} c_i z_i + \frac{a_{11}}{z - d_1} + \frac{a_{21}z + a_{22}}{(z - d_2)^2} + \frac{a_{23}}{(z - d_2)^3} + \frac{a_{24}}{(z - d_2)^4}$$

$$(4)\ \bar{H}(z) = \sum_{i=0}^{5} c_i z_i + \frac{a_{11}z + a_{12}}{(z - d_1)^2} + \frac{a_{13}}{(z - d_1)^3} + \frac{a_{14}}{(z - d_1)^4} + \frac{a_{15}}{(z - d_1)^5}$$

$$(5)\ \bar{H}(z) = \sum_{i=0}^{5} c_i z_i + \frac{b_{11}z + b_{12}}{(z^2 + \alpha_1 z + \beta_1)^2} + \frac{b_{13}z + b_{14}}{(z^2 + \alpha_1 z + \beta_1)^2} + \frac{a_{11}}{z - d_1}$$

2.3.2 Problem Transformation

For each decomposition of $\bar{H}(z)$, the vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, $\mathbf{d}$, $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ can be treated as the decision vector. Denote $\mathbf{h}$ as these vectors the a subproblem can be formulated into

$$\min_{\mathbf{h}\in S} \max_{\omega\in\Omega} E(\omega) = W(\omega)|\bar{H}(e^{j\omega}) - \bar{D}(\omega)|^2, \tag{2.16}$$

Where S is the stability domain. Then, we can formulate all the subproblems for every decomposition of $\bar{H}(z)$. The number of decompositions is finite. Hence, we can solve all the subproblems and choose the best solutions. Thus, the problem (2.4) is equivalent to a series of subproblems.

The stability of $H(\omega)$ is a difficult topic in general. That is, it is hard to justify the stability of $H(\omega)$ from the coefficient vector $\mathbf{q}$. There are many papers which consider the stability of $H(\omega)$ the stability of $H(\omega)$ from $\mathbf{q}$. However, only sufficient condition can be derived. If we consider the decomposition of $\bar{H}(\omega)$, the stability becomes very easy. Note that the stability of $\bar{H}(\omega)$ only relates to the parameters α_i, β_i, d_i, we have the following theorem.

Theorem 2.1. *For each decomposition of $H(\omega)$, it is stable if and only if*

$$-1 < d_i < 1 \tag{2.17}$$

$$D \begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} + e > 0 \tag{2.18}$$

where,

$$\mathbf{D} = \begin{bmatrix} 1 & 1 \\ -1 & 1 \\ 0 & -1 \end{bmatrix}$$

$$\mathbf{e} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

Proof. As we all know, if one system is stable, all the poles of the system are inside a unit circle on a complex plane. (2.17) is the case of the single real pole and (2.18) is obviously for the quadratic case, which is known as stability triangle in [2]. □

It can be seen that the condition in Theorem 1 is sufficient and necessary, which is better than the other conditions in [21], [27], etc.. Based on Theorem 1, the IIR filter design problem for each decomposition is transformed into

$$\min_{\delta,\mathbf{h}} \quad \delta \tag{2.19a}$$

$$s.t \quad \left|\bar{H}(\omega) - \bar{D}(\omega)\right|^2 \le \delta \tag{2.19b}$$

$$-1 + \varepsilon \le d_i \le 1 - \varepsilon \tag{2.19c}$$

$$D \begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} + e \ge \varepsilon \tag{2.19d}$$

where ε is the error term and defined as $\varepsilon = e^{-6}$.

Problem (2.19) is a subproblem, which can be solved by optimization software. Then, the optimal parameters $c_i, \alpha_i, \beta_i, d_i, a_{ij}, b_{ij}$, can be obtained and the corresponding coefficients $\mathbf{p}, \mathbf{q}$ can be computed. After the optimal solutions of all the subproblems are obtained. We can compare the results and choose the optimal one.

2.3.3 Problem Simplification

As the order of numerator m increases, the number of the decompositions increases. Hence, it is required to simplify the problem. For this, we have the following lemmas:

Lemma 1: For any transfer function $Q(z) = \frac{A}{(z-a_0)^n}(n \geq 3)$, where $(|a_0| < 1)$ is a real number, then $\forall \varepsilon > 0$, there is a stable $P(z)$ satisfies $|Q(z) - P(z)| < \varepsilon$, where the multiplicity of $P(z)$ is 1.

Proof. Without loss of generality, we set $A = 1$. Note that the unit circle of the complex plane is compact, and $|a| \neq 1$, we can define $M = \min_{z \in \odot} |z - a_0|^n > 0$, where $\odot$ represents the unit circle of complex plane. Since $M > 0$, we set $\delta < M$ such that $M - \delta > 0$. Then, $\forall \varepsilon > 0$, we have

$$\begin{aligned} \left| \frac{1}{(z-a_0)^n} - \frac{1}{(z-a_0)^n - \delta} \right| &= \left| \frac{\delta}{(z-a_0)^n((z-a_0)^n - \delta)} \right| \\ &< \left| \frac{\delta}{M(M-\delta)} \right| < \varepsilon \end{aligned} \tag{2.20}$$

which implies that $\delta < M^2\varepsilon/(1 - M\varepsilon)$. Then, δ is given by

$$\delta < \min(M^2\varepsilon/(1 - M\varepsilon), M) \tag{2.21}$$

(2.20) is true. Hence, we can set $P(z) = \frac{1}{(z-a_0)^n - \delta}$ and have

$$|Q(z) - P(z)| < \varepsilon.$$

Set $(z - a_0)^n - \delta = 0$, we have $z = a_0 + \delta^{1/n} e^{j2k\pi/n}, k = 0, \cdots, n-1$. Note that in the cases k and $n - 1 - k$, the roots are $a_0 + \delta^{1/n} e^{j2\pi k/n}$ and $a_0 + \delta^{1/n} e^{-j2\pi k/n}$. Hence, the roots are conjugate. Note that δ can be set sufficient small such that the roots still satisfy $|z| < 1$, which means that $P(z)$ is stable. □

It can be seen that the multiplicities of poles of $P(z)$ are 1, which can be decomposed into the following form

$$P(z) = \sum_{r=1}^{n/2} \frac{b_{r1}z + b_{r2}}{z^2 + \alpha_r z + \beta_r} \tag{2.22}$$

when n is even and

$$P(z) = \sum_{r=1}^{(n-1)/2} \frac{b_{r1}z + b_{r2}}{z^2 + \alpha_r z + \beta_r} + \frac{a_1}{z - d_1} \tag{2.23}$$

when n is odd.

In order to illustrate Lemma 1, we give one example with $n = 5$, $a_0 = 0.3$ and then calculate the error between $P(z)$ and $Q(z)$ on frequency domain. We define $E =\| Q(z) - P(z) \|_2$ and Fig. 1 shows the error with different ε. From the Figure, we can easily find that when ε closes to zero, $P(z)$ approximates $Q(z)$.

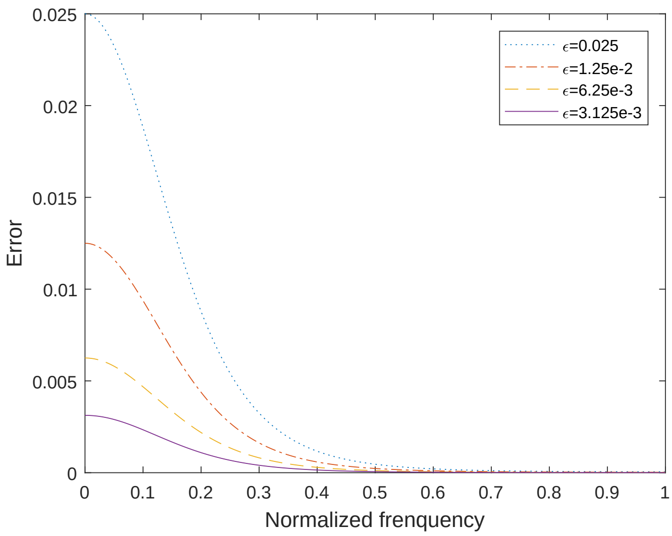


Figure 2.1: Error of different ε

Lemma 2: For any transfer function $Q(z) = \frac{bz+c}{(z^2+\alpha z+\beta)^n} (n \geq 2)$, where α, β satisfy the stability triangle in the next subsection, there is a P(z) satisfies $|Q(z) - P(z)| < \varepsilon$.

The proof of this lemma is similar to Lemma **??**, which is omitted. By Lemma 1 and Lemma 2, we have the theorem below.

Theorem: For any $\delta > 0$, and any decomposition of $H(\omega)$ which contains $n(n \geq 3)$ repeated real roots or $n(n \geq 2)$ complex roots, there is a decomposition of first kind which approached to $H(\omega)$.

It can be seen from Theorem 2 that we can always solve the first kind of decomposition to find the optimal solution and the other sort of decompositions are unnecessary, where $\bar{H}(z)$ can always be written as (2.9) or (2.10). Hence, we simplify the problem by only considering one subproblem.

2.4 Simulation

In this section, five examples are given to test the effectiveness of the proposed method, and we design the coefficients in the MATLAB environment. For convenience, we set the weighting function $W(\omega) = 1$.

2.4.1 Example 1

In this example, we choose a lowpass filter with a filter order as $M = 4$, $N = 15$, and the ideal transfer function is given by:

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

To demonstrate the Theorem 2, we optimize the coefficients vector for every possible decomposition of the transfer function and compare the results among the different subproblems. Note that m is equal to 4, there would be four possible decomposition which has been written in the last section. By optimizing the coefficients, we obtain four filters, and the positions of poles are shown in Fig.2. The magnitude, passband error and group delay of the four filters are depicted in Fig.3, and the error measurements are showed in Table 1, including Minimax Error(ME), Maximum Passband Error(P-MME), Maximum Stopband Error(S-MME) and Maximum Passband Group Delay(P-MGD). From the figure and table, we can easily find that the IIR designed using the first form performs best, which represents a pure quadratic decomposition. This is consistent with the theoretical result in Theorem 1.

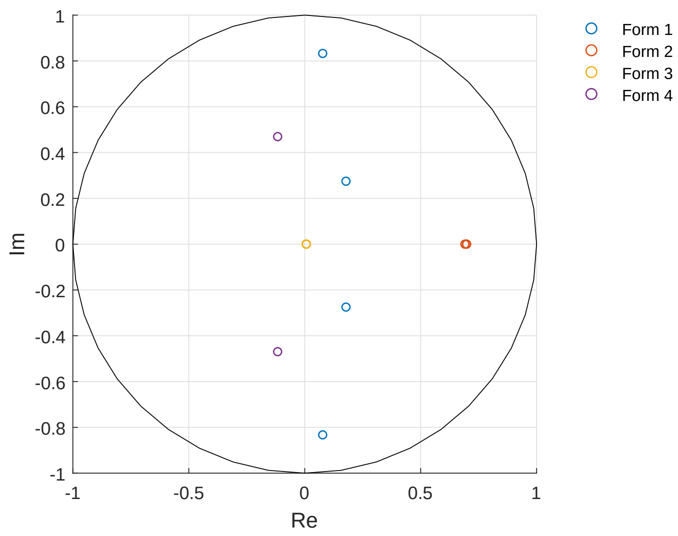


Figure 2.2: The placement of poles in Example 1

Table 2.1: Measurements of Design Result in Example 1

Filter	ME	P-MME	S-MME	P-MGD
Form 1	0.005526	0.005185	0.005526	0.3241
Form 2	0.11491	0.1041	0.1491	0.4242
Form 3	0.07004	0.07004	0.05672	1.295
Form 4	0.1604	0.04297	0.1604	2.476

2.4.2 Example 2

This specification of the example we chose is the same as the second example adopted in [21] and the second example in [28], which is a highpass filter with a filter order as $m = 14, n = 14$, for comparison. The ideal frequency response is defined as

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

According to the method we proposed, we decompose the transfer function into the form in (2.9). After optimizing, the seven pairs of conjugate poles we obtained are depicted

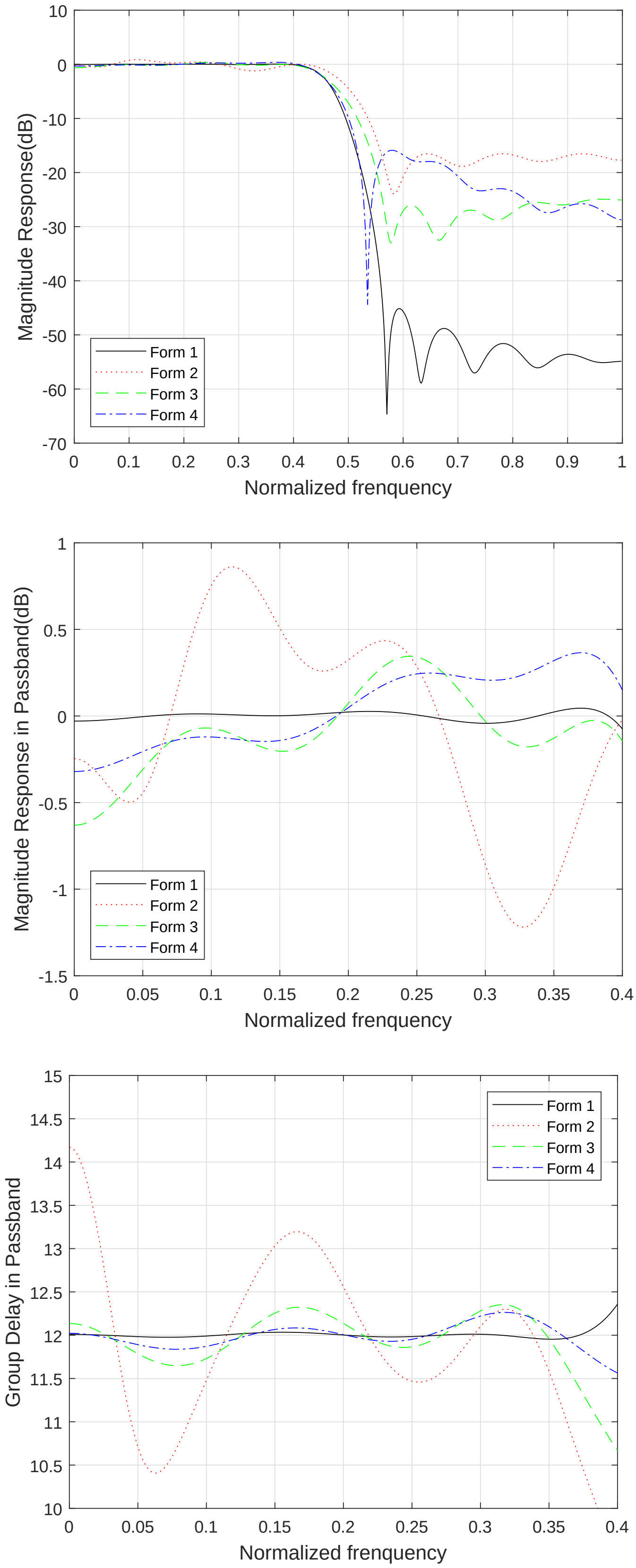


Figure 2.3: Magnitude, passband error, and group delay in Example 1

in Fig.4. Apparently, all the poles are located in the unit circle on the complex plane, thus the stability is guaranteed. The magnitude response is shown in Fig.5. The dashed curves represents the magnitude and group delay of the IIR obtained by using the method in [21], while the solid curves represents the IIR filter designed by our method. From the figure, we can see the solid one is slightly better than the dashed one. We also compare our method with the method in [28] and the measurements are shown in Table 2.

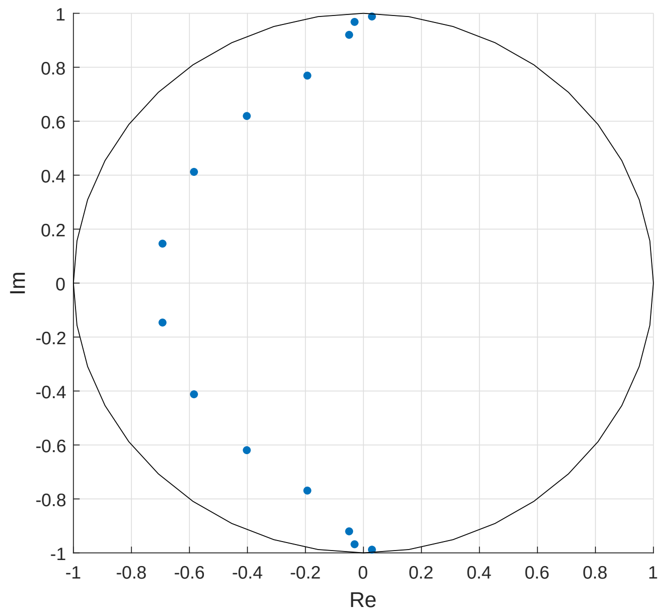


Figure 2.4: Placement of Poles in Example 2

Table 2.2: Measurements of Design Result in Example 2

Method	S-MME	P-MME	P-MGD
Method Proposed	0.01521	0.01221	0.4452
Method in [28]	0.01601	0.01598	0.4782
Method in [21]	0.0429	0.0424	3.823

2.4.3 Example 3-5

In this section, we compare our method with the method in [27], which is sequential constrained least-squares(SCLS) Steiglitz-McBride(SM), and we choose three examples as follows:

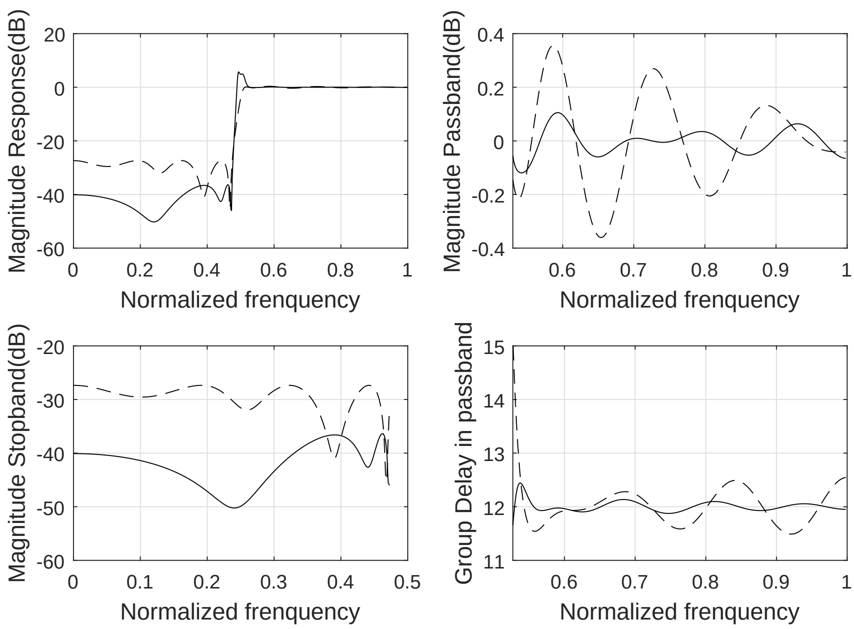


Figure 2.5: Magnitude and group delay of IIR designed in Example 2 and [21]

Table 2.3: Measurements of Design Result in Example 3-5

IIR	Method	P-MME	S-MME	P-MGD
Example 3	SCLS-SM	0.0147	0.0149	0.9257
	Proposed	0.01219	0.01099	0.05268
Example 4	SCLS-SM	0.00659	0.00664	1.789
	Proposed	0.006293	0.005958	0.922
Example 5	SCLS-SM	0.08125	0.08125	2.830
	Proposed	0.008522	0.008416	0.1047

Example 3: Low-pass filter with $m = 12, n = 6, \tau = 9$, a passband $[0, 0.5\pi]$, and a stopband $[0.6\pi, \pi]$.

Example 4: Low-pass filter with $m = n = 18, \tau = 15$, a passband $[0, 0.5\pi]$, and a stopband $[0.55\pi, \pi]$.

Example 5: High-pass filter with $m = n = 15, \tau = 14$, a passband $[0.75\pi, \pi]$, and a stopband $[0, 0.7\pi]$.

In these three examples, we decompose the transfer functions in Example 3 and Example 4 into form (2.9), while the transfer function in Example 5 into form (2.10). Table 3 shows some measurements of the filters. In the Fig.6, it represents the magnitude error(MAE) in passband and stopband in terms of the three examples above, respectively. The result shows that all the three filters performs better than those designed in [27].

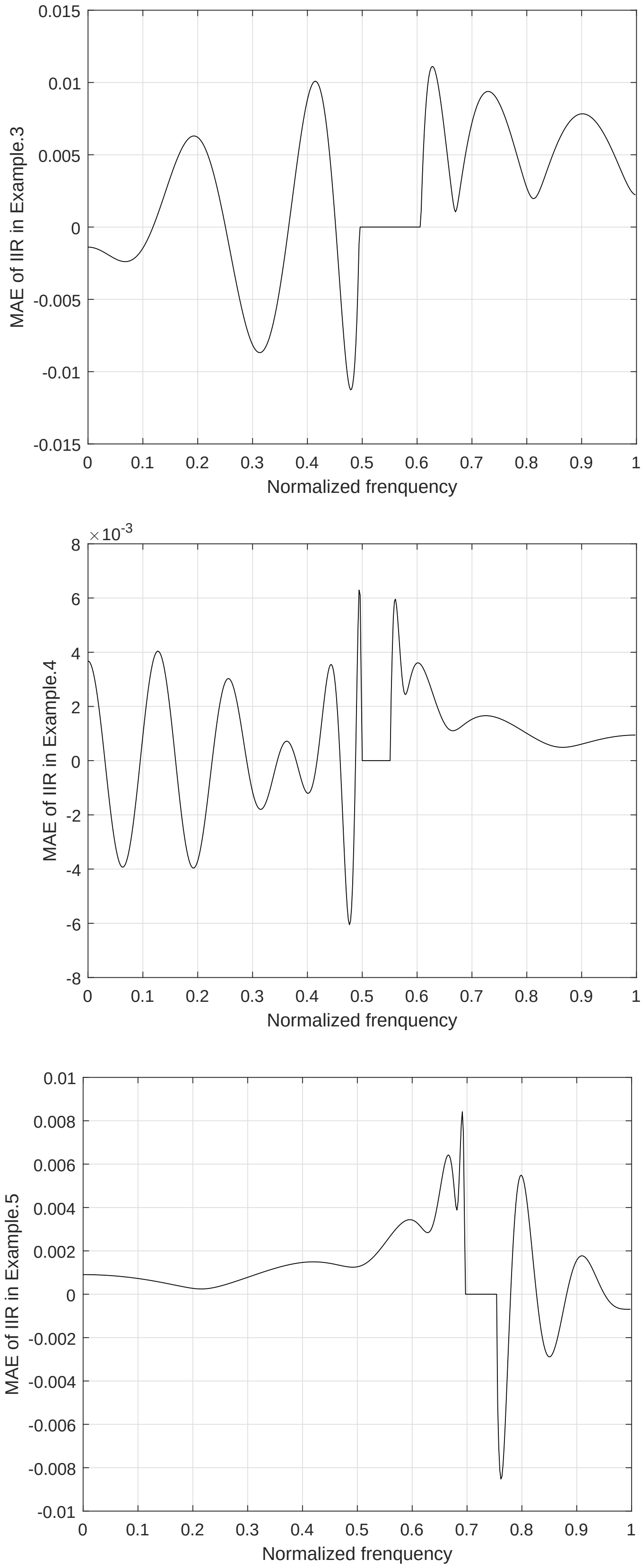


Figure 2.6: Magnitude, passband error, and group delay in example 3-5

2.5 Conclusion

In this section, a decomposition method for designing IIR digital filter by optimizing the poles is presented. We translate the direct form of an IIR into many parallel forms by fraction decomposition and simplify into only one form of a transfer function. Comparing with directly designing the parameters of the transfer function of an IIR filter, poles design gains advantages such as better stability conditions. According to the simulation result, we prove that this method performs well, especially in the situation of a high order system.

Chapter 3

Optimal Design of Fixed Equalizers

3.1 Introduction

Speech is the primary input for any voice communication system. However, in the real environment, noise contamination is inevitable, which degrades the performances of the voice communication system. This results in the popularity of speech enhancement, which aims at improving the quality or the intelligibility of speech. It can be used in various domains such as hearing aids, assistive listening devices, speech recognition system, voice communication systems, etc.

The problem of speech enhancement can be posed as an adaptive equalization design problem. In essence, the equalization process involves a known digital filter and desired system modeled by a know digital filter. The objective of the problem is constructing an equalizer such that the original system composed of the equalizer approximates the desired system.

There are some main algorithms when designing equalizers, such as spectral subtraction and Minimum Mean-Square Error (MMSE) method. One method used widely is spectral subtraction, which spectrally subtracts the noise estimate from the envelope of the noise spectrum. This method is non-linear and stand-alone noise suppression algorithms and used to reduce unwanted broadband noise acoustically added to a signal [22], [5]. However, for the Any spectral estimator erroneous estimate results in noise artifacts and this mismatch results in bad performances and then a Voice Activity Detector (VAD) is needed to detect speech.

In another kind of widely used systems, MMSE Short-Time Spectral Amplitude (STSA) estimator is applied. In the methods based on Wiener filters, a desired and input signals are given, and the Winer filter produces the minimum mean square error estimate of the desired signal [3], [13]. However, Wiener STSA estimator is derived from the optimal MMSE signal spectral estimator, which is not the optimal spectral amplitude estimators under the assumed statistical model and criterion. Then, more STSA estimators are proposed, [47] which based on modeling speech and noise spectral components as statistically independent Gaussian random variables. In constructing the enhanced signal, the MMSE STSA estimator is combined with the complex exponential of the noisy phase.

In the method proposed, we try to design a generalized equalizer based on the optimization of Perceptual Evaluation of Speech Quality (PESQ) and Short-Time Objective Intelligibility (STOI). We choose a certain number of signals as the training database added different types of noise with different SNR, and then calculate the gains of the equalizer by optimizing the two objective intelligibility measures. The average equalizer of the training database can be a more general one, which can be used to enhance speech polluted by similar kinds of noise. After this training procedure, there is no need to know or estimate the original signal. In addition, to reduce complexity, we transform signals into the frequency domain by Fast Fourier Transform (FFT) and divide signals into subbands. The average equalizer has proven to be effective and robust and eliminate the requirement od VAD.

3.2 Mould of Equalizer

Suppose that $s(k)$ is speech signal and $n(k)$ is noise signal, with the sum of $s(k)$ and $n(k)$, the noisy signal could denoted as

$$x(k) = s(k) + n(k) \tag{3.1}$$

Then, if taken Fourier Transform, $X(\omega)$ could be presented as:

$$X(\omega) = S(\omega) + N(\omega) \tag{3.2}$$

By applying Discrete Fourier Transformation (DFT), the noisy signal is translated from the time domain to the frequency domain, and then a frequency-domain filter could be applied. For the purpose of speech enhancement, a fixed equalizer of $x(k)$ is used by multiplying gain multipliers w_i to each of short-term spectral spectrum. The enhanced coefficient of output signal $y(k)$ in frequency domain could be presented as

$$Y(i,k) = w(i) \cdot X(i,k) \quad i = 0, 1, \cdots, N-1 \tag{3.3}$$

where $X(i,k)$ is the short spectrum of $x(k)$ at time k and N is the transform length of DFT. In this case, the value of the gains is limited to the range

$$0 \leq w(i) \leq 1$$

The fixed gain factor $w(i)$ are derived by solving some optimal problems which will be presented in detail in the next subsection. Then, the enhanced spectrum is converted back to the time domain by Inverse Discrete Fourier Transformation (IDFT) and the parameters of this procedure, including the window function, transform length and overlap, are the same as DFT. The output signal $y(k)$ is

$$y(k) = IDFT[Y(\omega)] \tag{3.4}$$

This is the whole procedure of the system, and the main problem we have to solve is designing the parameters of the equalizer.

3.3 Equalizer Design

To measure the performance of the equalizer, we introduce some indicators. There are many objective measures in the literature for evaluating the performance of the equalizer outputs. In this paper, we mainly choose two indicators, namely PESQ and STOI.

Perceptual evaluation of speech quality (PESQ) is the result of the integration of the perceptual analysis measurement system (PAMS) and, an enhanced version of the Perceptual Speech Quality Measure(PSQM), PSQM99. The model begins by level, aligning both signals to a standard listening level. [40] ITU-T recommends it for speech quality assessment for 3.2 kHz handset telephony and narrow-band speech codes. It is obtained by a linear combination of the average disturbance value D_{ind} and the average asymmetrical disturbance value A_{ind}

$$P = 4.5 - 0.1D_{ind} - 0.0309A_{ind} \tag{3.5}$$

A Short-Time Objective Intelligibility measure(STOI) is an objective machine-driven intelligibility measure, which based on shorter time segments shows a high correlation with the intelligibility of noisy and time-frequency weighted noisy speech. STOI is a function of a time-frequency-dependent intermediate intelligibility measure, which compares the temporal envelopes of the clean and degraded speech in short-time regions using a correlation coefficient. Then, the intermediate intelligibility is defined as the sample correlation coefficient between the two vectors. Finally, the average of the intermediate intelligibility measure is calculated. This is different from other measures, which typically consider the complete signal at once, or use a short analysis length(20-30ms). Experiments show STOI has a high correlation with speech intelligibility for different listening tests.[41]

To design the coefficients of the equalizer, arranging the multipliers $w(i)$ as a vector

$$\mathbf{w} = (w(0), w(1), \cdots, w(N-1))^T \tag{3.6}$$

where N is the number of DFT. Assume a noise equalizer parametrized by the vector $\mathbf{w}$ is employed, the output signal in the time domain after the equalizer denoted by $y(k)$. According the $y(k)$ and $s(k)$, the score of PESQ and STOI could be calculated, defined as $P(\mathbf{w})$ and $S(\mathbf{w})$. Then, the equalizer design problem can be written as two maximization problems, denoted by

$$\max \ P(\mathbf{w}) \tag{3.7a}$$

$$s.t \quad 0 \leq w(i) \leq 1 \quad \forall i \tag{3.7b}$$

and

$$\max \ S(\mathbf{w}) \tag{3.8a}$$

$$s.t \quad 0 \leq w(i) \leq 1 \quad \forall i \tag{3.8b}$$

As each objective measure has its unique property in evaluating the speech quality, we attempt to form a combination of the two objective measures which would adjust the speech quality in different aspects. Let the optimal problem be given by

$$\max \; A * S(\mathbf{w}) + (1 - A) * P(\mathbf{w})/4.5 \tag{3.9a}$$

$$s.t \quad 0 \leq w(i) \leq 1 \quad \forall i \tag{3.9b}$$

where A is a real number from 0 to 1.

This is a constrained nonlinear optimization problem, and there are many optimal methods could be applied in solving the problems. Sequential quadratic programming methods represent the state of the art in nonlinear programming methods. Based on the works of Biggs, Han and Powell; an SQP method mimics Newtons method for constraint optimization. It is an iterative method of starting from some initial point and converging to a constrained local minimum (Edwin and Kumaanan, 2001). At each iteration, one solves a quadratic program (QP) that models the original nonlinear constrained problem at the current point. The solution to the QP is used as a search direction to find an improving point, which is the next iteration. The algorithm could be described as follows:

(1) Choose an initial guess $\mathbf{w}^{(0)}$ that satisfies the constraints and set $k = 0$.

(2) Compute $P(\mathbf{w})$ and $S(\mathbf{w})$.

(3) Compute the optimum update $\delta^{(k)}$ by solving QP problem.

(4) Set $\mathbf{w}^{(k+1)} = \mathbf{w}^{(k)} + \alpha\delta^{(k)}$, with $\alpha \in (0, 1)$.

(5) If $\| \delta^{(k)} \|_2 / \| \mathbf{w}^{(k+1)} \|_2 < c$, stop. Otherwise, set $k = k + 1$ and then go to (2).

Set $\alpha = 0.5$ and $c = 10^{-3}$.

3.4 Experimental Result

In the experimental part, the optimal problem was solved in MATLAB environment, and nonlinear programming function "Fmincon" in the optimization toolbox was used in this paper. This experiment can be divided into three parts according to different indicators. Every part contains three groups according to the types of variables, including the different SNR, number of FFT, and the types of noise.

First of all, we discuss how the quality of the enhanced signals changes as the size of the training data increases. In this part, the Number of FFT is 1024, the SNR is 0dB, and the noise is the babble noise. The noisy and enhanced average STOI and PESQ value and Average Percentage Improvements(API) are presented in Table 3.1, in which S(ori) and P(ori) means the STOI and PESQ value of noisy signal and S(pro) and P(pro) stand for the STOI and PESQ value of the enhanced signals and P(SPI) and S(SPI) mean the average percentage improvement of PESQ and STOI. The average percentage of improvement is calculated by:

$$A = \frac{PESQ(STOI)\ after\ enhanced - PESQ(STOI)\ before\ enhanced}{PESQ(STOI)\ before\ enhanced} \times 100$$

From the table, we could see that as the size of the training sample increases, the performances of the enhanced signals improve little. In the following part, we choose 50 training samples and 100 test samples.

Table 3.1: Results Based on Different Size of Training Samples

Size of Training	S(ori)	S(pro)	S(API)	P(ori)	P(pro)	P(API)
50	0.6706	0.6904	3.0233	1.6909	1.8274	8.1722
100	0.6706	0.6905	3.0196	1.6909	1.8101	7.1229
150	0.6706	0.6905	3.0279	1.6909	1.8113	7.1971
200	0.6706	0.6906	3.0425	1.6909	1.8141	7.3646
250	0.6706	0.6910	3.1031	1.6909	1.8184	7.6228
300	0.6706	0.6911	3.1256	1.6909	1.8266	8.119

3.4.1 Experiment based on STOI

Different SNR

In this subsection, we add the noise with different signal-to-noise (SNR) conditions, 0dB, 5dB, 10dB, 15dB respectively. To see the impact of SNR, we choose the same babble noise and the same number of FFT which is 512. In addition, we make the noisy signals under different SNR conditions pass different equalizers trained under different SNR. All the results are shown in Table 3.2

Table 3.2: Results Based on Different SNR

SNR(Test Sample)	S(ori)	P(pro)	Equalizer	S(pro)	S(API)	P(ori)	P(API)
0dB	0.6706	1.6909	0dB	0.6874	2.5567	1.8301	8.3124
			5dB	0.6877	2.5860	1.8313	8.4058
			10dB	0.6847	2.1211	1.7984	6.4272
			15dB	0.6808	1.5245	1.7561	3.8847
5dB	0.7772	2.0554	0dB	0.7879	1.4322	2.1727	5.7430
			5dB	0.7925	2.0021	2.1748	5.8510
			10dB	0.7914	1.8510	2.1459	4.4339
			15dB	0.7887	1.4952	2.1095	2.6466
10dB	0.8620	2.4141	0dB	0.8636	0.2299	2.5111	4.0334
			5dB	0.8724	1.2374	2.5152	4.2074
			10dB	0.8738	1.3992	2.4906	3.1807
			15dB	0.8731	1.3013	2.4594	1.8840
15dB	0.9220	2.7618	0dB	0.9137	-0.8632	2.8417	2.9071
			5dB	0.9249	0.3414	2.8469	3.0946
			10dB	0.9286	0.7385	2.8266	2.3544
			15dB	0.9298	0.8695	2.7999	1.3841

Different number of FFT

In this part, we use a different number of FFT when optimizing signals, and we add the babble noise with a 0dB noise. The equalizers obtained from the training samples are used to process test samples. The results are shown in Table 3.3.

Table 3.3: Results Based on Different No. of FFT

No.FFT	S(ori)	S(pro)	S(API)	P(ori)	P(pro)	P(API)
64	0.6706	0.6775	1.0315	1.6909	1.8504	9.5578
128	0.6706	0.6800	1.4286	1.6909	1.8623	10.2637
256	0.6706	0.6855	2.2789	1.6909	1.8923	12.0634
512	0.6706	0.6874	2.5567	1.6909	1.8301	8.3124
1024	0.6706	0.6904	3.0233	1.6909	1.8274	8.1722
2048	0.6706	0.6873	2.5334	1.6909	1.7572	3.9508

Different Noise

In this part, we choose three types of noise, including babble, white and subway noise, to compare the average values of parameters, and the results are shown in Table 3.4. All three types of noise are set as 0dB SNR, and the number of FFT is 512.

Table 3.4: Quality Improvement Based on Different Kinds of Noise

Noise(Test Sample)	S(ori)	P(ori)	Equalizer	S(pro)	S(API)	P(pro)	P(API)
babble	0.6706	1.6909	babble	0.6874	2.5567	1.8301	8.3124
			subway	0.6726	0.3056	1.8774	11.1679
			white	0.6778	1.0787	1.8492	9.4623
subway	0.6471	1.5121	babble	0.6425	-0.6371	1.5580	3.0850
			subway	0.6691	3.4854	1.6972	12.6750
			white	0.6562	1.4560	1.6648	10.4414
white	0.6577	1.4201	babble	0.6516	-0.8284	1.4582	2.6895
			subway	0.6589	0.2718	1.5931	12.7030
			white	0.6698	1.9144	1.5762	11.5461

3.4.2 Experiments based on PESQ

Different SNR

The same as the experiment of STOI, we add the babble noise with different signal-to-noise (SNR) conditions, 0dB, 5dB, 10dB, 15dB respectively and set the same number of FFT which is 512. In addition, we also use the equalizers trained by different SNR to test the signals. The average stoi ans pesq improvements are given in Table 3.5

Different number of FFT

In this part, we compare the average value among different FFT number. We add a babble noise with a 0dB SNR and choose the different number of FFT as 64, 128, 256, 512, 1024, 2048. The result is shown in the Table 3.6

Different Noise

In this part, we compare the average value among three different kinds of noise, babble, subway and white noise with 0dB SNR and 512 FFT number. Quality improvements are displayed in Table 3.7

Table 3.5: Quality Improvement Based on Different SNR

SNR(Test Sample)	S(ori)	P(pro)	Equalizer	S(pro)	S(API)	P(ori)	P(API)
0dB	0.6706	1.6909	0dB	0.6559	-2.1648	2.4274	44.0943
			5dB	0.6561	-2.1193	2.4224	43.7984
			10dB	0.6588	-1.7078	2.3961	42.2353
			15dB	0.6622	-1.2065	2.3011	36.5442
5dB	0.7772	2.0554	0dB	0.7480	-3.7152	2.7041	31.7734
			5dB	0.7487	-3.6298	2.7010	31.6223
			10dB	0.7522	-3.1750	2.6763	30.4185
			15dB	0.7562	-2.6639	2.5906	26.2177
10dB	0.8620	2.4141	0dB	0.8179	-5.0845	3.0021	24.4545
			5dB	0.8193	-4.9185	3.0010	24.4055
			10dB	0.8236	-4.4293	2.9786	23.4748
			15dB	0.8280	-3.9149	2.8975	20.0952
15dB	0.9220	2.7618	0dB	0.8653	-6.1276	3.3024	19.6445
			5dB	0.8675	-5.8882	3.3053	19.7481
			10dB	0.8727	-5.3371	3.2886	19.1330
			15dB	0.8777	-4.7940	3.2137	16.4097

Table 3.6: Average Quality Improvement Based on Different No. of FFT

No.FFT	S(ori)	S(pro)	S(API)	P(ori)	P(pro)	P(SPI)
64	0.6706	0.6512	-2.8830	1.6909	2.5592	51.9842
128	0.6706	0.6478	-3.3930	1.6909	2.5174	49.4813
256	0.6706	0.6543	-2.4094	1.6909	2.4799	47.2406
512	0.6706	0.6559	-2.1648	1.6909	2.4274	44.0943
1024	0.6706	0.6567	-2.0522	1.6909	2.3505	39.4980
2048	0.6706	0.6699	-0.0848	1.6909	2.0309	20.3526

Table 3.7: Quality Improvements Based on Different Kinds of Noise

Noise(Test Sample)	S(ori)	P(ori)	Equalizer	S(pro)	S(API)	P(pro)	P(SPI)
babble	0.6706	1.6909	babble	0.6559	-2.1648	2.4274	44.0943
			subway	0.6399	-4.5785	2.3603	40.0694
			white	0.6534	-2.5506	2.3437	39.1051
subway	0.6471	1.5121	babble	0.6113	-5.5190	2.2055	47.1796
			subway	0.6125	-5.3434	2.1647	44.3789
			white	0.6198	-4.2304	2.2470	50.1001
white	0.6577	1.4201	babble	0.6117	-6.9396	2.1339	51.9364
			subway	0.6100	-7.2301	2.0982	49.3750
			white	0.6246	-5.0020	2.2002	56.9087

3.4.3 Experiments based on PESQ and STOI

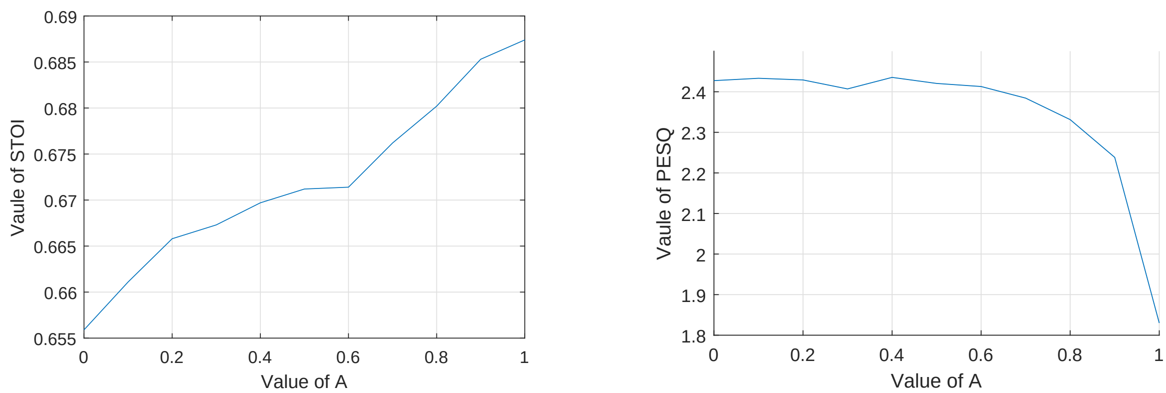
In this section, we consider both PESQ and STOI when optimizing. We set the objection function as the following form:

$$max \quad A * S + (1 - A) * P/4.5$$

where S represents the value of STOI, P means the value of PESQ, and we give a set of values of A. The same as the experiments before, we choose 50 male and 50 female samples as training database and 100 male and 100 female samples as a test database. The quality improvements are shown in Table 3.8 and Fig.1.

Table 3.8: Quality Improvements of Different Value of A

Value of A	S(ori)	S(pro)	S(API)	P(ori)	P(pro)	P(SPI)
0		0.6559	-2.1648		2.4274	44.0943
0.1		0.6611	-1.3807		2.4332	44.4350
0.2		0.6658	-0.6673		2.4291	44.1867
0.3		0.6673	-0.4393		2.4070	42.8724
0.4	0.6706	0.6697	-0.0712	1.6909	2.4354	44.5811
0.5		0.6712	0.1533		2.4205	43.6943
0.6		0.6714	0.1830		2.4129	43.2475
0.7		0.6762	0.9081		2.3843	41.5414
0.8		0.6802	1.5011		2.3313	38.3736
0.9		0.6853	2.2535		2.2383	32.8038
1		0.6874	2.5567		1.8301	8.3124



(a) STOI Improvements (b) PESQ Improvements

Figure 3.1: Quality Improvements

Chapter 4

Optimal Design of Beamformer

4.1 Introduction

In this section, optimal methods for beamformer design for hands-free devices are discussed. The use of hands-free devices increases because of its conveniences. As we all know, if there is a near field noise with low SNR, it could have a severe effect on the performances of hands-free devices. However, the beamformer technique works well in speech enhancement, which can solve this problem efficiently. One common approach is setting a pre-process array. Many methods are applied in microphone array technique, such as sidelobe cancellers, constrained beamforming, and gain optimization methods. All of these methods consider the design problem in optimization.

In this part, we discuss gain optimization methods with analytic expressions, which are commonly used for designing filter weights, including least-squares and the signal to noise ratio technique. Least-square method (LS) concentrate more on distortion control compared with the signal to noise interference ratio (SNIR) method. However, SINR performs better in noise suppression than LS. In the experimental part, we could see the value of relative measures of these two approaches.

4.2 Method and Performance Measure

4.2.1 Least Square Method

Suppose that there is a microphone array with N elements situated in a fixed position and the signal received by it consists of two parts including source signal, $s[n]$, and the noise signal, $v[n]$. Besides, we consider the source is a wideband source located in the nearfield of a microphone array. Then, the received signal can be represented by

$$x_i[n] = s_i[n] + v_i[n], \quad i = 1, 2, \cdots, N \tag{4.1}$$

where $x_i(k)$ is the ith microphone observation of the signals.

In order to enhance the source signal and reduce the noise signal, digital linear filters are used at each microphone signal. The output of the beamformer is given by

$$y[n] = \sum_{i=1}^{N} \sum_{j=0}^{L-1} \omega_i[j] x_i[n-j] \tag{4.2}$$

where the order of filers is $L - 1$. The filter taps for the beamformer is arranged as a vector:

$$\omega = (\omega_1^T, \omega_2^T, \cdots, \omega_N^T) \tag{4.3}$$

where ω_i^T is the weight for channel number i.

By using some formulation, we could design the filter, which makes the output of the beamformer resembles the source signal, while the noise components are canceled or attenuated. This objective can be implemented in various methods depending on the objective functions we chose. If we use the least square formulation to minimize the difference between the output signal $y[n]$ and the reference source signal $s_r[n]$, the beamformer design problem can be considered as

$$\omega_{LS}^{opt} = arg\ min\{\sum_{n=0}^{N-1} [|y[n] - s_r[n]|^2]\} \tag{4.4}$$

where the reference calibration sequences $s_r[n]$ is not directly available, but it could be gained in a quiet environment and this signal will represent the temporal and spatial information about the source. In addition, the real data $x[n]$ is independent of the $s_r[n]$ when N is large and the problem could be solved as

$$\omega_{LS}^{opt} = arg\ min \sum_{n=1}^{N-1} [|\omega^H s[n] - s_r[n]|^2 + |\omega^H x[n]|^2] \tag{4.5}$$

where the superscript ^{H} denotes Hermitian transpose. If calculating the sum, it can be changed into

$$\omega_{LS}^{opt} = arg\ min\{w^H[\hat{R_{ss}}(N) + \hat{R_{xx}}(N)]\omega - \omega^H \hat{r_s}^H(N)\omega + \hat{r_s}\} \tag{4.6}$$

where source correlation can be written as

$$\hat{R}_{ss}(N) = \frac{1}{N}\sum_{n=0}^{N-1} s[n]s^H[n] \tag{4.7}$$

$$\hat{r}_s(N) = \frac{1}{N}\sum_{n=0}^{N-1} s[n]s^*[n] \tag{4.8}$$

where $s[k] = [s_1[k], s_2[k], \cdots, s_M[K]]^T$ are microphone observations when the calibration source signal is active alone. The observed data correlation matrix estimate $\hat{R}_{xx}(K)$ can be calculated similarly.

$$\omega_{LS}^{opt}(K) = [\hat{R}_{ss}(K) + \hat{R}_{xx}(K)]^{-1}\hat{r}_s(K)$$

4.2.2 Signal to Noise Interference Ratio Methods

Signal to noise interference ratio (SNIR) is described as

$$Q = \frac{avereage\ signal\ output\ power}{average\ noise\ pluse\ interference\ output\ power} \tag{4.9}$$

The beamformer weights design method by optimizing this ratio is known as the signal-to-noise ratio maximization method. The cross-correlation function between microphone when only the useful signal is active is denoted as

$$r_{s_is_j}(l) = E[s_i(k)_j^*(k+l)], \quad l = 0, 1, \cdots, L-1 \tag{4.10}$$

where $*$ denotes the conjugation and L is the length of the filter. Correlation matrix and thee correlation matrix between microphones when only the signal is active is

$$R_{ss} = \begin{bmatrix} R_{s_1s_1} & R_{s_1s_2} & \cdots & R_{s_1s_K} \\ R_{s_2s_1} & R_{s_2s_2} & \cdots & R_{s_1s_K} \\ \vdots & \vdots & \ddots & \vdots \\ R_{s_Ks_1} & R_{s_Ks_2} & \cdots & R_{s_Ks_K} \end{bmatrix} \tag{4.11}$$

where

$$R_{s_is_j} = \begin{bmatrix} r_{s_is_j}(0) & r_{s_is_j}(1) & r_{s_is_j}(2) & \cdots & r_{s_is_j}(L-1) \\ r^*_{s_is_j}(1) & r_{s_is_j}(0) & r_{s_is_j}(1) & \cdots & r_{s_is_j}(L-2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r^*_{s_is_j}(L-1) & r^*_{s_is_j}(L-2) & \cdots & \cdots & r_{s_is_j}(0) \end{bmatrix} \tag{4.12}$$

The output of the useful signal can be described as:

$$r_{y_sy_s} = \mathbf{w}^H R_{ss}\mathbf{w} \tag{4.13}$$

where ^{H} is hermitian transpose.

Similarly, the autocorrelation function when only the noise is active is described as

$$r_{n_in_j}(l) = E[n_i(k)^*_j(k+l)], \quad l = 0, 1, \cdots, L-1 \tag{4.14}$$

and the R_{nn} is

$$R_{nn} = \begin{bmatrix} R_{n_1n_1} & R_{n_1n_2} & \cdots & R_{n_1n_K} \\ R_{n_2n_1} & R_{n_2n_2} & \cdots & R_{n_1n_K} \\ \vdots & \vdots & \ddots & \vdots \\ R_{n_Kn_1} & R_{n_Kn_2} & \cdots & R_{n_Kn_K} \end{bmatrix} \tag{4.15}$$

where

$$R_{s_is_j} = \begin{bmatrix} r_{n_in_j}(0) & r_{n_in_j}(1) & r_{n_in_j}(2) & \cdots & r_{n_in_j}(L-1) \\ r^*_{n_in_j}(1) & r_{s_is_j}(0) & r_{s_is_j}(1) & \cdots & r_{n_in_j}(L-2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r^*_{n_in_j}(L-1) & r^*_{n_in_j}(L-2) & \cdots & \cdots & r_{n_in_j}(0) \end{bmatrix} \tag{4.16}$$

Now, the output signal of noise is

$$r_{y_n y_n} = \mathbf{w}^H R_{nn} \mathbf{w} \tag{4.17}$$

Finally, this problem can be considered as an optimal problem as

$$\mathbf{w}_{SNR}^{opt} = arg\ max[\frac{\mathbf{w}^H R_{ss} \mathbf{w}}{\mathbf{w}^H R_{nn} \mathbf{w}}] \tag{4.18}$$

4.2.3 Performance Measure

Generally, there are two criteria to evaluate the performance of the beamformer, the distortion calculated by a deviation between the original speech signal and the output signal of the beamformer, and the noise suppression. To make the performance visible, some measurements are introduced. For the distortion, a normalized quantity is described as:

$$D = \frac{1}{2\pi} \int_{-\pi}^{\pi} \mid C_d \hat{P}_{y_s}(\omega) - \hat{P}_{x_s}(\omega) \mid d\omega \tag{4.19}$$

and constant C_d is

$$C_d = \frac{\int_{-\pi}^{\pi} \hat{P}_{x_s}(\omega) d\omega}{\int_{-\pi}^{\pi} \hat{P}_{y_s}(\omega) d\omega} \tag{4.20}$$

where $\omega = 2\pi f$ and f is the normalized frequency. In this equation, $\hat{P}_{x_s}(\omega)$ is a mean spectral power of the speech signal observed by microphones and $\hat{P}_{y_s}(\omega)$ is the mean spectral power of the output signal obtained from the beamformer, when the signal of interest is active alone.

To measure the noise suppression in time domain, a normalized noise suppression is denoted as

$$S_N = C_s \frac{\int_{-\pi}^{\pi} \hat{P}_{y_n}(\omega) d\omega}{\int_{-\pi}^{\pi} \hat{P}_{x_n}(\omega) d\omega} \tag{4.21}$$

where

$$C_s = \frac{1}{C_d} \tag{4.22}$$

In this equation, the $\hat{P}_{x_n}$ is the spectral power estimate of the signal obtained from the sensors and $\hat{P}_{y_n}$ is the spectral power estimate of the signal obtained from the output of the beamformer when only the noise signal is active.

Another important measure is residual noise suppression known as segmental SNR and denotes as

$$S_t = \frac{10}{M}\sum_{m=0}^{M-1}\log_{10}\frac{\sum_{n=Nm}^{Nm+N-1}x^2(n)}{\sum_{n=Nm}^{Nm+N-1}[x(n)-y(n)]^2} \tag{4.23}$$

where $x(n)$ is the original speech and $y(n)$ is the output speech of the beamformer, the N is the frame length and M is the total number of the frames.

4.2.4 Experimental Results

In this experiment, Microsemi's audio processing eval board is used, and the model is ZLE38000. There are four microphones and a USB interface converter, and then a Raspberry Pi is applied to receive signals. The data was gathered from a beamformer array with four microphones (sensors) placed in a square array with a distance between the sensors $d = 0.08m$. The environment consisted of two recorded speeches, including the source signals and noise signals. The source signals had been used which are emitted from the artificial talker and the hands-free loudspeaker individually and the same as the interference noise signals. These two kinds of speeches are recorded from five different positions and shown in Fig.4.1. The sampling frequency of the signals is 16k Hz. All of the filters are designed on the LS and SNIR. We set the filter length $N = 100$, and the signal and noise ratio was set as 0dB.

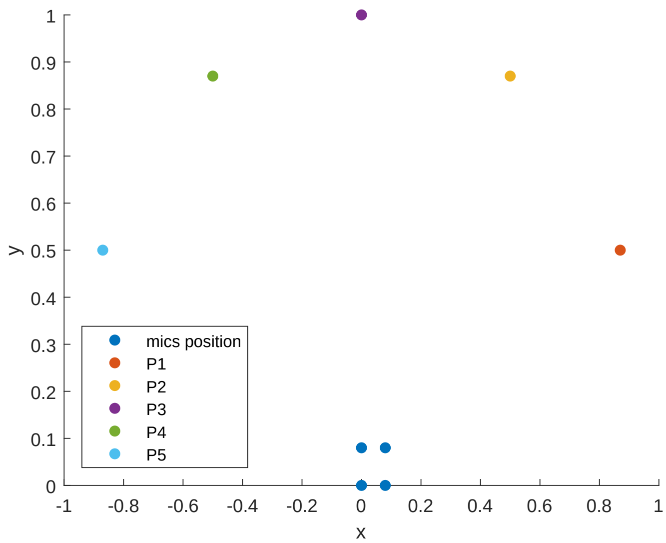


Figure 4.1: The positions of the microphone array and speech signals

Single Passband and Single stopband

The construction of a filter to make signals from a certain location pass the system, known as the passband, and to make signals from other locations be attenuated or canceled, known as stopband, is referred to as broadband beamforming. In this part, filters were designed with the single passband and single stopband, and the passband and stopband of the four beamformers are shown in Tab. 4.1

Table 4.1: Beamformer Settings

No.Beamformer	Passband	Stopband
Beamformer 1	P1	P2
Beamformer 2	P1	P3
Beamformer 3	P1	P4
Beamformer 4	P1	P5

The suppression of the signal on time domain and frequency domain and the distortion were given to illustrate the effectiveness of these two methods. The performance measurements were shown in Tab. 4.2 and Tab. 4.3, where "NS" represented the noise suppression calculated by (4.21), "SegSNR" represented residual noise suppression in (4.23) and "Distortion" denoted the signal distortion achieved by (4.19). To observe the performances more clearly, a figure of the segmental SNR of Beamformer 4 in passband, stopband and transition band was plotted. Results based on LS method and SNIR method were shown in Fig. 4.2 and Fig. 4.3, respectively.

Table 4.2: Performance measures for beamformers Based on LS

No.Beamformer	NS	SegSNR	Distortion
Beamformer 1	5.7119	6.7016	-3.4152
Beamformer 2	5.2048	7.8192	-3.6341
Beamformer 3	6.1460	8.4598	-3.6699
Beamformer 4	9.4067	7.9828	-3.9919

Table 4.3: Performance measures for beamformers Based on SNIR

No.Beamformer	NS	SegSNR	Distortion
Beamformer 1	19.8517	18.1895	1.3302
Beamformer 2	17.2155	17.1881	0.4241
Beamformer 3	14.0220	17.9572	2.0889
Beamformer 4	16.6900	18.6112	1.0012

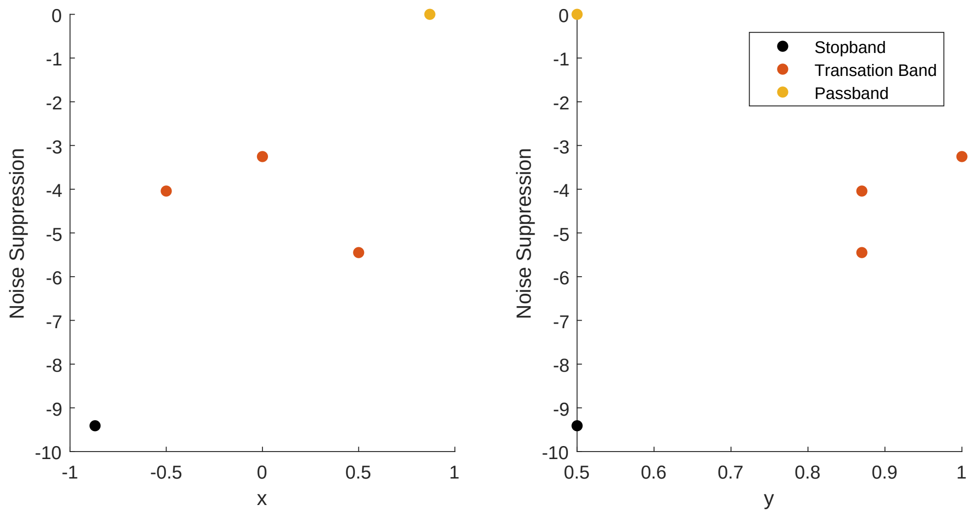


Figure 4.2: Results of Beamformer 4 Designed by LS

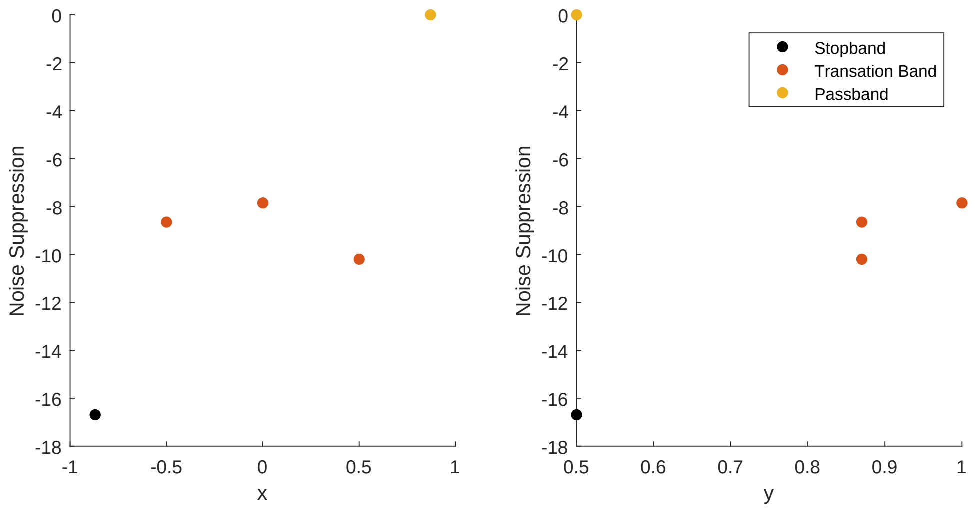


Figure 4.3: Results of Beamformer 4 designed by SNIR

Multi Passband and Multi Stopband

In this part, two beamformers are designed, and the passbands and stopbands were given in Tab.4.4. Values of noise suppression were plotted in Fig.4.4 and Fig.4.5.

Table 4.4: Beamformer Settings

No.Beamformer	Passband	Stopband
Beamformer 5	P1	P2 P3 P4 P5
Beamformer 6	P1 P2	P3 P4 P5

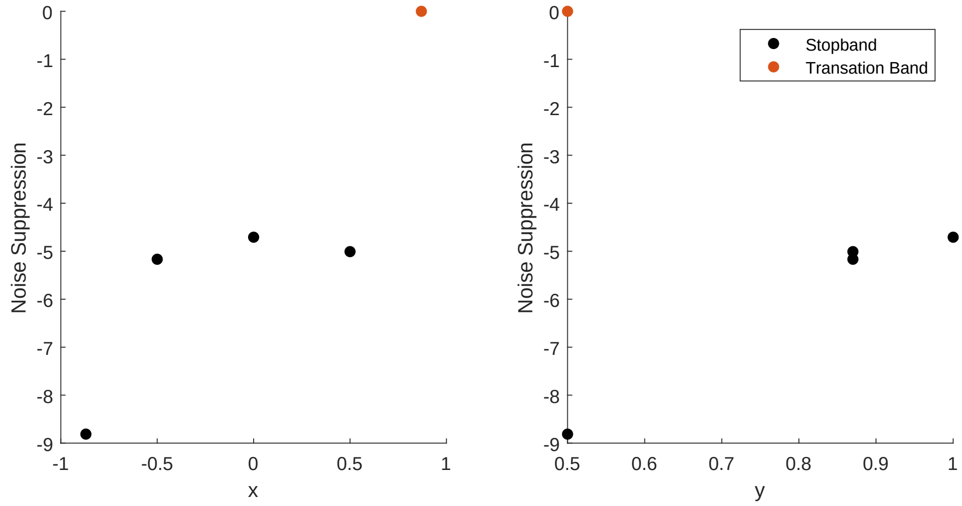


Figure 4.4: Results of Beamformer 5 Designed by LS

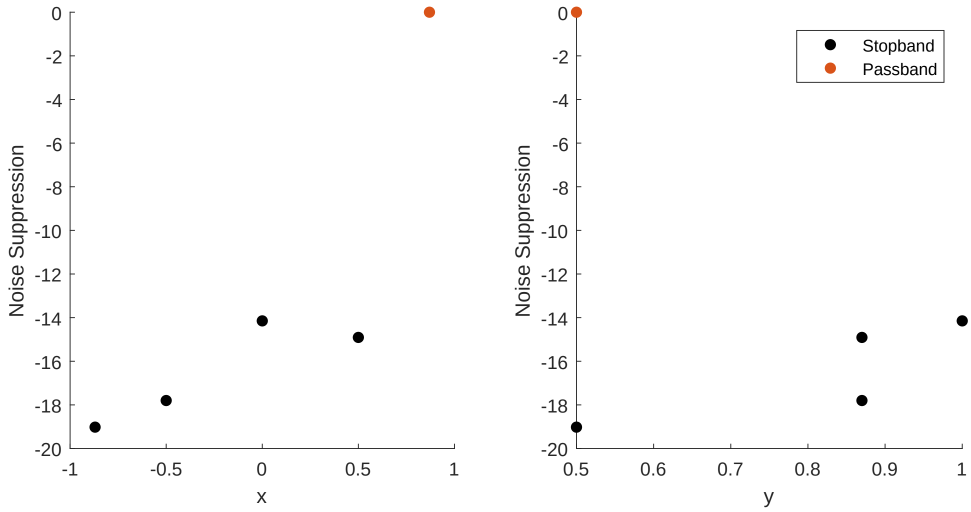


Figure 4.5: Results of Beamformer 5 designed by SNIR

4.3 Method Based on Transfer Function

4.3.1 Signal Model

The performance and theory of adaptive arrays always demand a careful calibration of the array elements, especially for broadband arrays. For the LS method proposed in 4.2, the calibration signals are recorded by a reference sensor in a real situation. These are recorded from both the target position and noisy position and stored in memory for later use. This method has a high demand for prior information. However, in this part, a transfer function is applied to gather the calibration signals.

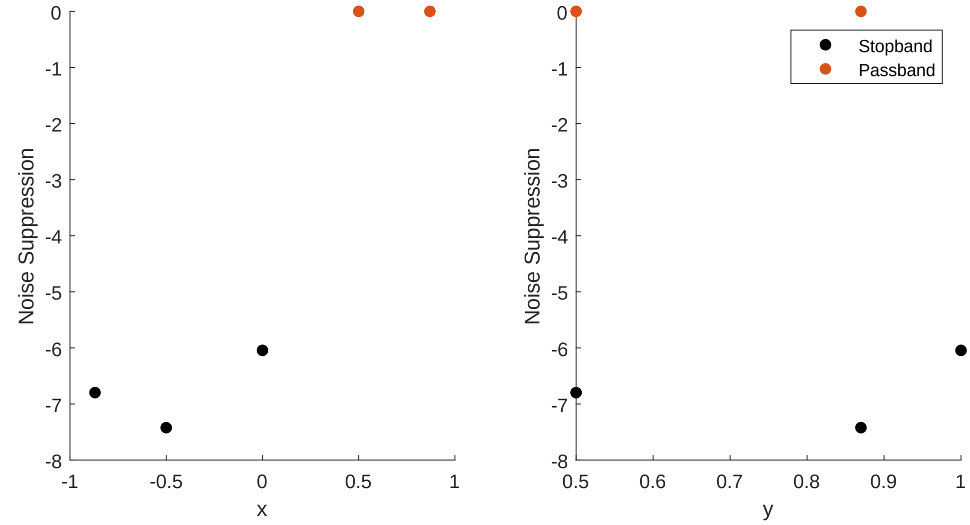


Figure 4.6: Results of Beamformer 6 Designed by LS

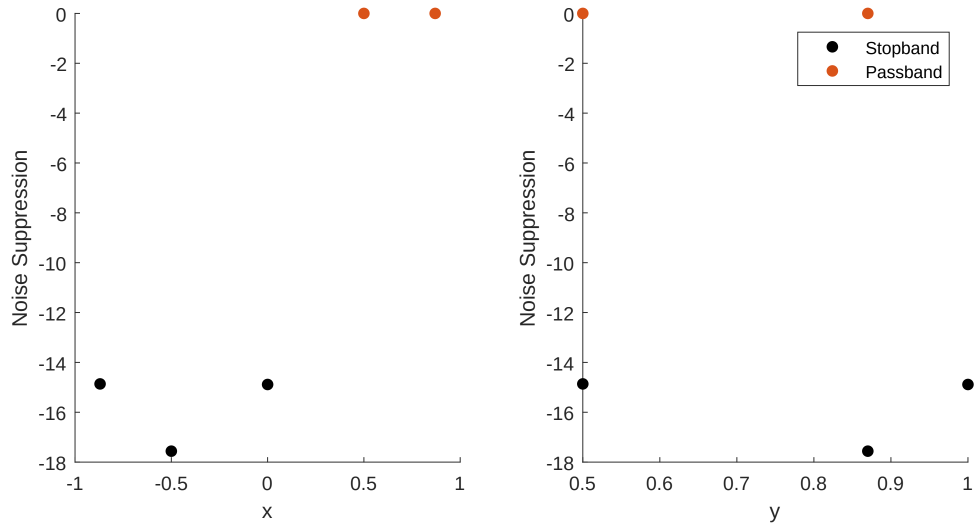


Figure 4.7: Results of Beamformer 6 designed by SNIR

For a given placement of the microphone array design problem, the configuration of the microphone array is the available and only problem is to find the coefficients of filters so that the output of the beamformer could satisfy certain criteria. In the methods proposed in last subsection, the main problem is the estimation of source correlation matrix $\hat{R}_{ss}(K)$, the calibration sequences $s_r[n]$ and the observe data correlation matrix $\hat{R}_{xx}(K)$. In the broadband beamforming design, the aim is to make signals from a certain location pass and cancel or attenuate signals from other places. Assuming that the point source is fixed in a known location, the estimated source signal can be obtained by the transfer function of sound propagation and the calibration sequences is achieved by a reference channel. In the same way, the observed signal can be calculated.

In the nearfield, the transfer functions from the sound source to the sensors are given as

$$A_i(r, f) = \frac{1}{\parallel r - r_i \parallel} e^{\frac{-j2\pi f \parallel r - r_i \parallel}{c}}, \quad i = 1, \cdots, N \tag{4.24}$$

where N is the total number of the microphones, c is the speed of sound in air, r is the location of the sound source, and r_i is the location of the ith sensor. However, the transfer function is given in the frequency domain and in the applied methods, the signals have proceeded in the time domain. Fourier transform and inverse Fourier transform are needed. Then, if the positions of the sound source and the microphones array are known, the signals received by sensors could be evaluated.

However, there would be a deviation between the actual value and measured value of the distance from the microphone to the sound source. In addition, because of the existence of reverberation and echo during the process of sound transmission, there would also be a deviation of the transfer function proposed in the last subsection. In order to reduce deviation and make the estimation to be more robust, more points around the target location could be considered when designing the beamformer.

4.3.2 Experimental Results

In this experimental part, the calibration signals were created by the transfer function according to the configuration of the microphone array and the locations of the source speech and noise speech. Real data were collected by the multichannel digital audio tap recorder with a sample rate of 16kHz. Both noises and speech signals had been used which are emitted individually from the artificial talker and hands-free loudspeaker. The recordings with real speech serve as the point source with a bandwidth of 300-3400 Hz which covers the frequency range of human voice.

The environment contained one speech source and one noise source, and the locations were shown in Fig. 4.8. The four microphones (sensors) placed in a square array. According to the positions, the signals received by the four microphones could be estimated. All the estimation signals could be achieved by the transfer function in (4.24). The beamformers were designed by these estimation signals and tested by the real data recorded by the microphone array. The performances of beamformers designed by LS and SNIR based on these data were given in Tab. 4.5.

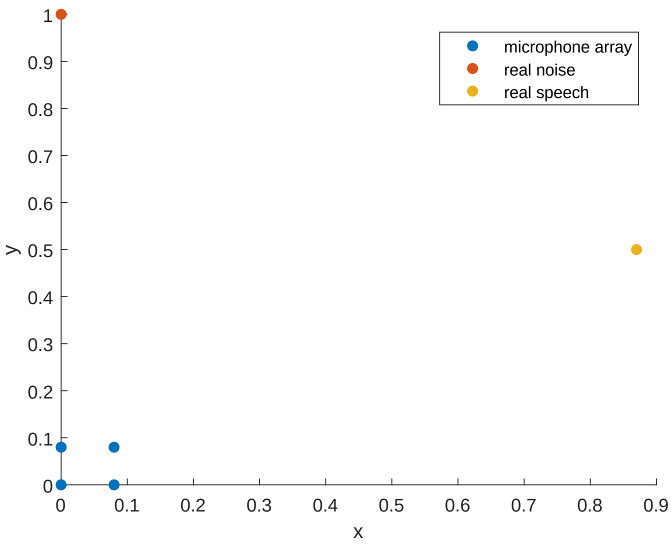


Figure 4.8: The positions of the microphone array and signals

Table 4.5: Performance Measures of Beamformers

Methods	NS	SegSNR	Distortion
LS	1.6466	1.8814	0.4463
SNIR	2.4422	2.4879	2.0889

To enhance robustness, four more signals whose locations surrounded the speech signal were created, and another four signals surrounded noisy position were also estimated and Fig. 4.9 showed the positions. By adding the surrounding points, the passband and stopband were extended, which may reduce the deviation caused by distance measurement. The performances tested by real data were presented in Tab. 4.6

Table 4.6: Performance Measures of Beamformers

Methods	NS	SegSNR	Distortion
LS	2.1466	2.4814	-0.2063
SNIR	12.3422	11.9879	2.7758

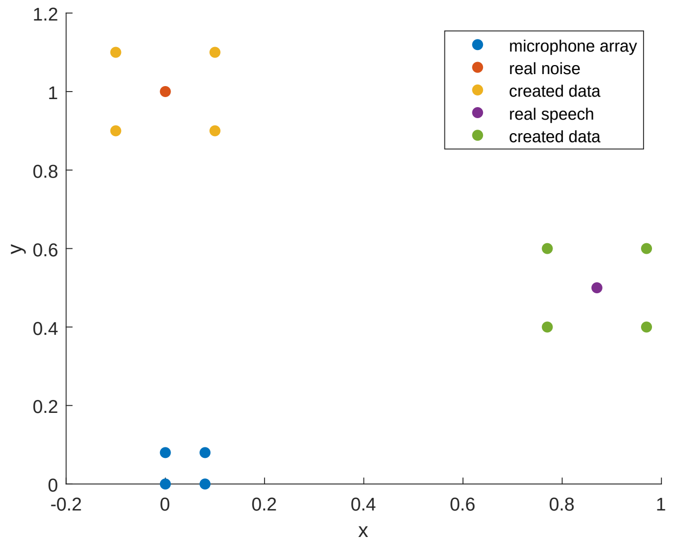


Figure 4.9: The positions of the microphone array and signals

Chapter 5

Summary and Future Works

This chapter will summarize the works studied in this thesis and suggest some possible directions for further research.

5.1 Summary

The focus of this thesis is placed on various optimization problems in speech signal processing, including IIR design problems by a decomposition method, fixed equalizer design by optimizing the objective measures and beamformer design problems by optimizing filter weights. Specifically, three research works have been studied in the following.

1. A decomposition method for designing IIR digital filter by optimizing the poles is presented in Chapter 2. The direct form of an IIR was translated into many parallel forms by fraction decomposition and simplify into only one form of the transfer function. This poles design method achieves advantages such as better stability conditions, and according to the simulation result, this method performs well, especially in the situation of the high order system.

2. In Chapter 3, a mew fixed equalizer design method based on optimization of Perceptual Evaluation of Speech Quality (PESQ) and Short-Time Objective Intelligibility (STOI was proposed. In the experimental part, the equalizer has proven to be effective and robust.

3. Chapter 4 presents two popular beamformer design methods in hands-free devices, by minimizing the Least Square and optimizing the SNR and some results based on real data are given, which is important for its practical applications.

5.2 Future Works

Each of the three problems discussed in this thesis is a rich subject of research. As for the beamformer design, just two traditional beamformer design methods are discussed in this thesis. However, with the increasing demands of speech quality and the development of both size and power-efficient computing, a single microphone array can be extended into distributed arrays. In future work, we could consider applying the traditional beamforming algorithms on the distributed systems.

Bibliography

[1] Oscar Agazzi, D Messerschmitt, and D Hodges. Nonlinear echo cancellation of data signals. *IEEE Transactions on Communications*, 30(11):2421–2433, 1982.

[2] A. Antoniou. *Digital Filters: Analysis and Design.* New York:McGraw-Hill, 1993.

[3] S.D. Stearns B. Widrow. *Adaptive Signal Processing.* Prentice Hall, New Jersey, 1985.

[4] RB Blackman, JW Tukey, and T Teichmann. The measurement of power spectra. *Physics Today*, 13:52, 1960.

[5] S.F. Boll. Suppression of acoustic noise in speech using spectral subtraction. *IEEE Trans. Acoust. Speech Signal Process*, ASSP-27:113–120, 1979.

[6] Andrew Deczky. Synthesis of recursive digital filters using the minimum p-error criterion. *IEEE Transactions on Audio and Electroacoustics*, 20(4):257–263, 1972.

[7] Z. G. Feng and K. L. Teo. A discrete filled function method for the design of fir filters with signed-powers-of-two coefficients. *IEEE Trans. Signal Process.*

[8] Otis Lamont Frost. An algorithm for linearly constrained adaptive array processing. *Proceedings of the IEEE*, 60(8):926–935, 1972.

[9] Otis Lamont Frost. An algorithm for linearly constrained adaptive array processing. *Proceedings of the IEEE*, 60(8):926–935, 1972.

[10] EN Gilbert and SP Morgan. Optimum design of directive antenna arrays subject to random variations. *Bell System Technical Journal*, 34(3):637–663, 1955.

[11] Bernard Gold and K Jordan. A direct search procedure for designing finite duration impulse response filters. *IEEE Transactions on Audio and Electroacoustics*, 17(1):33–36, 1969.

[12] Marvin HJ Gruber. Statistical digital signal processing and modeling, 1997.

[13] S. Haykin. *Adaptive Filter Theory.* Prentice Hall, New Jersey, 1996.

[14] Howard Helms. Nonrecursive digital filters: Design methods for achieving specifications on frequency response. *IEEE Transactions on Audio and Electroacoustics*, 16(3):336–342, 1968.

[15] O Herrmann. Design of nonrecursive digital filters with linear phase. *Electronics Letters*, 6(11):328–329, 1970.

[16] Edward M Hofstetter. A new technique for the design of non-recursive digital filters. Technical report, MASSACHUSETTS INST OF TECH LEXINGTON LINCOLN LAB, 1970.

[17] Ruili Hu and Hassan M Ahmed. Echo cancellation in high speed data transmission systems using adaptive layered bilinear filters. *IEEE transactions on communications*, 42(234):655–663, 1994.

[18] J. Chen J. Tan. Low complexity and quasi-linear phase iir filters design based on iterative convex optimization. *Circuits and Systems (APCCAS), 2016 IEEE Asia Pacific Conference on.*

[19] Leland B Jackson. An improved martinet/parks algorithm for iir design with unequal numbers of poles and zeros. *IEEE Transactions on Signal Processing*, 42(5):1234–1238, 1994.

[20] A. Jiang and H. K. Kwan. Iir digital filter design with new stability constraint based on argument principle. *IEEE Trans. Circuits Syst. I, Reg. Papers*, 56(3):583593, Mar.2009.

[21] A. Jiang and H.K. Kwan. Minimax design of iir digital filters using iterative socp. *IEEE Trans. Circuits Syst. I*, 57:1326–1337, June.2010.

[22] J.H.L. Hansen J.R. Deller Jr., J.G. Proakis. Discrete time processing of speech signals. *Macmillan Publishing Company*, 1993.

[23] L. E. McBride K. Steiglitz. A technique for the identification of linear systems. *IEEE Trans. Autom. Control*, 10(4):461–464, Oct. 1965.

[24] JF Kaiser. Design methods for sampled data filters. In *Proc. 1st Allerton Conf. Circuit and System Theory*, pages 221–236, 1963.

[25] JF Kaiser. Digital filters. kapitel 7 in system analysis by digital computers, herausgegeben von kuo und kaiser, 1966.

[26] Kunio Kashino, Kazuhiro Nakadai, Tomoyoshi Kinoshita, and Hidehiko Tanaka. Organization of hierarchical perceptual sounds. In *Proc. 14th Int. conf. On Artificial Intelligence*, volume 1, pages 158–164, 1995.

[27] X. P. Lai and Z. P. Lin. Minimax design of iir digital filters using a sequential constrained least-squares method. *IEEE Trans. Signal Process.*, 58:3901–3906, 2010.

[28] X. P. Lai and Z. P. Lin. A sequential minimization procedure for minimax design of iir filters based on second-order factor updates. *IEEE Trans. Circuits Syst. II*, 58(1), Jan. 2011.

[29] M. C. Lang. Least-squares design of iir filters with prescribed magnitude and phase responses and a pole radius constraint,. *IEEE Trans.Signal Process.*, 48(11):31093121, Nov. 2000.

[30] W.-S. Lu. Design of stable minimax iir digital filters using semidefinite programming. *IEEE Signal Processing Letters*, 9(7), 2002.

[31] W.-S. Lu. Design of stable iir digital filters with equiripple passbands and peak-constrained least-squares stopbands. *IEEE Trans. Circuits Syst. II*, 46:14211426, Nov.1999.

[32] W. S. Lu and T. Hinamoto. Optimal design of iir digital filters with robust stability using conic-quadratic-programming updates. *IEEE Trans. Signal Process.*, 51:1581–1592, 2003.

[33] Rainer Martin. Speech enhancement using mmse short time spectral estimation with gamma distributed speech priors. In *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 1, pages I–253. IEEE, 2002.

[34] T Parks and James McClellan. Chebyshev approximation for nonrecursive digital filters with linear phase. *IEEE Transactions on Circuit Theory*, 19(2):189–194, 1972.

[35] Thomas W Parsons. Separation of speech from interfering speech by means of harmonic selection. *The Journal of the Acoustical Society of America*, 60(4):911–918, 1976.

[36] F Pirz. Design of a wideband, constant beamwidth, array microphone for use in the near field. *Bell System Technical Journal*, 58(8):1839–1850, 1979.

[37] L Rabiner, Bernard Gold, and C McGonegal. An approach to the approximation problem for nonrecursive digital filters. *IEEE Transactions on Audio and Electroacoustics*, 18(2):83–106, 1970.

[38] Lawrence R Rabiner. The design of finite impulse response digital filters using linear programming techniques. *Bell System Technical Journal*, 51(6):1177–1198, 1972.

[39] Lawrence R Rabiner, James H McClellan, and Thomas W Parks. Fir digital filter design techniques using weighted chebyshev approximation. *Proceedings of the IEEE*, 63(4):595–610, 1975.

[40] ITU-T Recommendation. Perceptual evaluation of speech quality (pesq): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs. *Rec. ITU-T P. 862*, 2001.

[41] Cees H Taal, Richard C Hendriks, Richard Heusdens, and Jesper Jensen. An algorithm for intelligibility prediction of time–frequency weighted noisy speech. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(7):2125–2136, 2011.

[42] J-P Thiran. Recursive digital filters with maximally flat group delay. *IEEE Transactions on Circuit Theory*, 18(6):659–664, 1971.

[43] C. Tseng. Design of stable iir digital filter based on least p-power error criterion. *IEEE Trans. Circuits Syst. I, Reg. Papers*, 51(9):18791888, Sep.2004.

[44] C. Tseng and S. Lee. Minimax design of stable iir digital filter with prescribed magnitude and phase responses. *IEEE Trans. Circuits Syst.I, Reg. Papers*, 49(4):547–551, Apr.2002.

[45] Masashi Unoki and Masato Akagi. A method of signal extraction from noisy signal based on auditory scene analysis. *Speech Communication*, 27(3-4):261–279, 1999.

[46] C. Tseng W.-S. Lu, S. Pei. A weighted least-squares method for the design of stable 1-d and 2-d iir digital filters. *IEEE Trans. Signal Process*, 46(1):1–10, Jan.1998.

[47] D. Malah Y. Ephraim. Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator. *IEEE Trans. Acoust. Speech Signal Process.*, ASSP-32:11091121, 1984.