Optimal Design of Dispersion Filter for Time-Domain Implementation of Split-Step Method in Optical Fiber Communication

Yang Zhu

Department of Electrical & Computer Engineering
McGill University
Montreal, Canada

September 2011

A thesis submitted to McGill University in partial fulfillment of the requirements for the degree of Master of Engineering.

2011/09/29

Abstract

The nonlinear Schrödinger equation can be solved by split-step methods, where in each step, linear dispersion and nonlinear effects are treated separately in a sequential manner. This thesis investigates the optimal design of an finite-impulse-response (FIR) filter as the time-domain implementation for the linear part. The objective is to minimize the integral of the squared error between the frequency response of the FIR filters and the desired dispersion characteristics over the band of interest. This least square (LS) problem is solved in two approaches: the normal equation approach gives an explicit solution and its Toeplitz structure enables fast computation; the singular value decomposition (SVD) approach provides geometrical, physical and numerical insights based on the theory of discrete prolate spheroidal sequence (DPSS).

A major concern is that as revealed by the theory of DPSS, this problem could be ill-conditioned, and henceforth its solution would be sensitive to small perturbations. Besides, the frequency response might exhibit singular behaviors such as overshoots. Two approaches are proposed to mitigate these shortcomings: the unconstrained LS approach adds a regularization term to the objective function, whereas the constrained approach also imposes a maximum magnitude constraint on the frequency response. The latter approach is formulated into a standard quadratically constrained quadratic programming (QCQP) problem that can be readily solved using state-of-the-art interior-point methods. Compared with previously designed FIR filters, these filters are easier to extract and the QCQP-based filter saves the filter length by at least 1/3. There is a complexity trade-off between these two filters: the unconstrained regularized LS filter is much easier to extract with the help of the modified Levinson-Durbin algorithm; the QCQP-based filter is shorter in length and saves computational complexity of the linear convolutions. The choice depends on whether the filter needs to be regenerated frequently or not. In addition, the required filter lengths for these filters are approximately linear functions of several parameters, which simplifies the task of choosing step size.

We verify the feasibility of the proposed filters in two categories of applications. Firstly, they can be used in the time-domain simulation of pulse propagation in optical fiber. For single channel and WDM channels, the proposed filters generate similar outputs as previous time-domain and frequency-domain methods, even after propagating thousands of kilometers. The proposed designs, together with overlap-add and overlap-save, can

reduce the overall computational complexity significantly. Secondly, the proposed filters can also be applied in time-domain digital backpropagation algorithms for fiber impairment compensation. Numerical simulations of a polarization division multiplexed quadrature phase-shift keying (PDM-QPSK) transmission system illustrate that the split-step methods based on the proposed filters are able to effectively mitigate the signal distortions caused by both dispersion and nonlinearities.

Sommaire

L'quation de Schrödinger non linéaire peut être résolue par des méthodes split-étape, où à chaque étape, la dispersion linéaire et les effets non linéaires sont traités séparément de manière séquentielle. Cette thèse étudie la conception optimale d'un filtre finis à réponse impulsionnelle (FIR) en tant que mise en uvre dans le domaine temporel pour la partie linéaire. L'objectif est de minimiser l'intégrale de l'erreur quadratique entre la réponse fréquentiel du FIR et les caractéristiques de dispersion souhaitées sur la bande d'intérêt. Ce problème de moindres carrés (LS) est résolu en deux approches: l'approche équation normale donne une solution explicite et sa structure de Toeplitz permet le calcul rapide; la décomposition en valeurs singulires (SVD) fournit un point de vue géométrique, physique et numérique basé sur la théorie de la séquence discrète sphéroïdale prolate (DPSS).

Une préoccupation majeure est que, comme révélé par la théorie de la DPSS, ce problème pourrait être mal conditionné, et désormais sa solution serait sensible aux petites perturbations. Par ailleurs, la réponse en fréquence peut manifester des comportements singuliers tels que les dépassements. Deux approches sont proposées pour atténuer ces lacunes: l'approche LS contrainte ajoute un terme de régularisation de la fonction objectif, alors que l'approche contrainte impose aussi une contrainte magnitude maximale sur la réponse fréquentielle. Cette dernière approche est formulée dans une norme quadratique contrainte quadratique de programmation QCQP problème qui peut être facilement résolu en utilisant l'état de l'art méthodes de points intérieurs. Comparé à des filtres FIR qui est déjà conu, ces filtres sont plus faciles à extraire et le filtre à base QCQP économise la longueur du filtre par au moins 1/3. Il y a une complexité compromis entre ces deux filtres: la contrainte régularisé LS filtre est beaucoup plus facile à extraire à l'aide de la modification algorithme de Levinson-Durbin; le filtre basé QCQP est plus court dans la longueur et la complexité des calculs sauve des circonvolutions linéaires. Le choix dépend si le filtre doit être régénéré fréquemment ou pas. En outre, les longueurs requises du filtre pour ces filtres sont des fonctions approximativement linéaires de plusieurs paramètres, ce qui simplifie la tâche du choix de la longueur de létape.

Nous vérifions la faisabilité des filtres proposés dans deux catégories d'applications suivantes. Premièrement, ils peuvent être utilisés dans la simulation dans le domaine temporel de la propagation des impulsions dans les fibres optiques. Pour un seul canal et des canaux WDM, les filtres proposés génèrent des sorties similaires à celles des méthodes précédentes

dans le domaine temporel et le domaine fréquentiel, même après la propagation de milliers de kilomètres. Les conceptions proposées, avec overlap-add et over-save peut réduire la complexité globale du calcul de faon significative. Deuxièmement, les filtres proposés peuvent également être appliqués dans le domaine temporel des algorithmes numériques rétro propagation d'indemnisation dépréciation de la fibre. Les simulations numériques d'une division polarisation en quadrature multiplexés déphasage système de transmission de saisie (PDM-QPSK) montrent que les méthodes de partage des étapes basées sur les filtres proposés sont en mesure d'atténuer efficacement les distorsions du signal causé par la dispersion et les non-linéarités.

Acknowledgments

First and foremost, I want to express my deepest gratitude to my supervisor, Prof. David V. Plant, who led me into the area of optical fiber communication, and worked with me closely throughout the last two years. He always showed full confidence in me, encouraged me and taught me how to develop immature ideas into a research paper. His engineering insights, technical depth, and critical thinking have always been the source of inspiration. His indispensable guidance and invaluable advices are throughout my graduate studies. I think, and hope what I have learned from him is not only the skill of conducting research, but also the way of living and personality.

Secondly, I would like to thank several professors for their lectures: Prof. Martin Rochette (Nonlinear optics), Prof. Mai Vu (Multiuser communications), Prof. Peter Kabal (Digital signal processing 1&2), Prof. Michael Rabbat (Telecom. network analysis), Prof. Jan Bajcsy (Introduction to digital communications) and Prof. Bruce Shepherd (Optimization). What I learned from them is invaluable for my research.

Thirdly, I am very grateful to Qunbi Zhuge, Benoît Châtelain and Zhaoyi Pan for their valuable suggestions in the simulations of long-haul transmission systems. I would also thank my colleagues at the Photonic System Group for their support and friendship, especially to Joshua Schwartz, Chen Chen, Ziad ElSahn, Jonathan Buset, Mathieu Chagnon, Mohamed Osman, Mohammad Pasandi, Bhavin Shastri, Mohammed Sowailem, Zhaobing Tian, Xian Xu, Yongyuan Zang and Christopher Rolston.

I would also like to thank my friends who has been helping, supporting and encouraging me for the past two years, especially Chao Zhao, Jian Wang, Yuling Nong, Xi Chen, Zhe Yao, Shudong Lin, and Jia Li. My gratitude also goes to my friends who are studying and working all over the world and still keep in close contact in the past two years, especially Haiding Sun, Rui Mao, Dan Hu, Jing Zhang, Zhixuan Xia, An Mao, Bin Zhang, Jing Liu, Linghui Rao, Shiqian Shao, Wei Yi, Danlu Xu, Li Cai, Zengli Yang, Ning Wang, Yao Xiao, Dong Xu, Dongxu Ji, Xi Liu, Chenguang Huang and Xiao Ma. Talking with you is always full of joy.

Last but not least, I would like to thank my parents for their support and love. Great thanks also go to all my family and friends in China.

Contents

List of Figures

List of Tables

List of Acronyms

ADC	analog-to-digital converter
A-SSM	asymmetric split-step method
ASE	amplified spontaneous emission
AWGN	additive white Gaussian noise
CD	chromatic dispersion
CDMA	code division multiple access
DCF	dispersion compensation fibers
DFT	discrete Fourier transform
DPSS	discrete prolate spheroidal sequence
DPSWF	discrete prolate spheroidal wave function
DSF	dispersion-shifted fibers
DSP	digital signal processing
DWDM	dense wavelength division multiplexing
EDFA	erbium-doped fiber amplifier
FFT	fast Fourier transform
FIR	finite-impulse-response
FWHM	full width at half maximum
FWM	four wave mixing
GVD	group velocity dispersion
IDFT	inverse discrete Fourier transform
IFFT	inverse fast Fourier transform
IIR	infinite-impulse-response
ISI	intersymbol interference
LO	local oscillator

LS	least square
MAC	multiply-accumulate
MLD	modified Levinson-Durbin
MZM	Mach-Zehnder modulators
NLSE	nonlinear Schrödinger equation
NZDSF	dispersion-shifted fibers
OEO	low-speed optical-electrical-optical
O/E	Optical/Electrical
PDM-QPSK	polarization division multiplexed quadrature amplitude phase-shift keying
QCQP	quadratically constrained quadratic programing
QPSK	quadrature amplitude phase-shift keying
SMF	single mode fibers
SNR	signal-to-noise ratio
SPM	self-phase modulation
SSFM	split-step Fourier method
S-SSM	symmetric split-step method
SVD	singular value decomposition
WDM	wavelength division multiplexing
XPM	cross-phase modulation

Chapter 1

Introduction

1.1 Backgrounds and Motivations

The data transmission rate in optical fiber communications has been increasing tremendously in an exponential manner since its introduction dating back to the 1970s, namely, from several Mb/s for single channel to current 100 Gb/s per channel with channel counts 80-100 and will cross the Tera milestone in the near future [1]. The secrets behind this spectacular achievement hide themselves in numerous advanced technologies such as third-window distributed feedback laser, wavelength division multiplexing (WDM), dispersion management, advanced modulation formats, effective coding schemes, electronic signal processing, Raman amplification, code division multiple access (CDMA) and multicarrier transmissions [2]. A good understanding, or more specifically, modeling of signal transmission through fibers is indispensable to design sophisticated high data-rate systems. It is well-known that the propagation of optical wave in fiber is described by the nonlinear Schrödinger equation. However, an analytical solution is generally unavailable except for some special cases (e.g. soliton) [3]

To this end, various numerical methods have been proposed to solve this partial differential equation [4, 5, 6, 7, 8, 9, 10]. Among these approaches, split-step Fourier method (SSFM) is a favorite choice because of its mathematical simplicity, conceptual clarity and numerical stability [11]. Although widely used in practice, it suffers from its high computational complexity and numerical inaccuracy due to discrete Fourier transform (DFT) and inverse discrete Fourier transform (IDFT) used. To overcome these shortcomings, a time-domain approach has been proposed and highlighted recently [6, 12, 13]. The time-domain

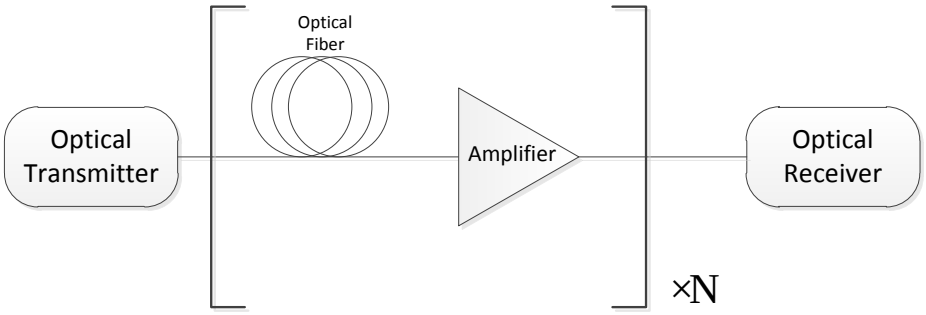

Fig. 1.1 A simple model of a fiber communication system with N spans.

approach implements the numerical calculation solely in the time domain by replacing the linear dispersion operation with a discrete-time infinite-impulse-response (IIR) or finite-impulse-response (FIR) filter.

Although previous efforts have already led to several time-domain designs with relatively low computational complexity and high accuracy, a systematic and comprehensive treatment of this topic is still missing from an optimization perspective. It is true that some ideas presented in this thesis were referred to in previous works; however, several important issues were treated casually and cannot be overemphasized.

1.1.1 Fiber-Optic Communication Systems

A typical fiber-optic communication system consists of an optical transmitter with electrical input, the fiber channel(s) and an optical receiver with electrical output, as shown in Fig 1.1. The problem of designing a fiber-optic communication system can be tackled from a layered approach — that is, to decompose the whole system into several layers as illustrated in Fig. 1.2. Herein, the block "channel" represents the equivalent channel that includes all the optical components in Fig 1.1. The combination of the pre-equalizer, pulse shaper, optical channels, filter and sampler, and post-equalizer can be viewed as another layered channel with the modulated pulses as the input signal. We denote the input signal before pre-equalization as $x(t)$ and the output signal after post-equalization as output $y(t)$. The purpose of fiber simulations is to obtain the output signal before post-equalization, based on the input signal $x(t)$ and the channel characteristics. The goal of digital backward propagation is to compensate the distortions introduced by the fiber channel so that $y(t)$ is as close to $x(t)$ as possible.

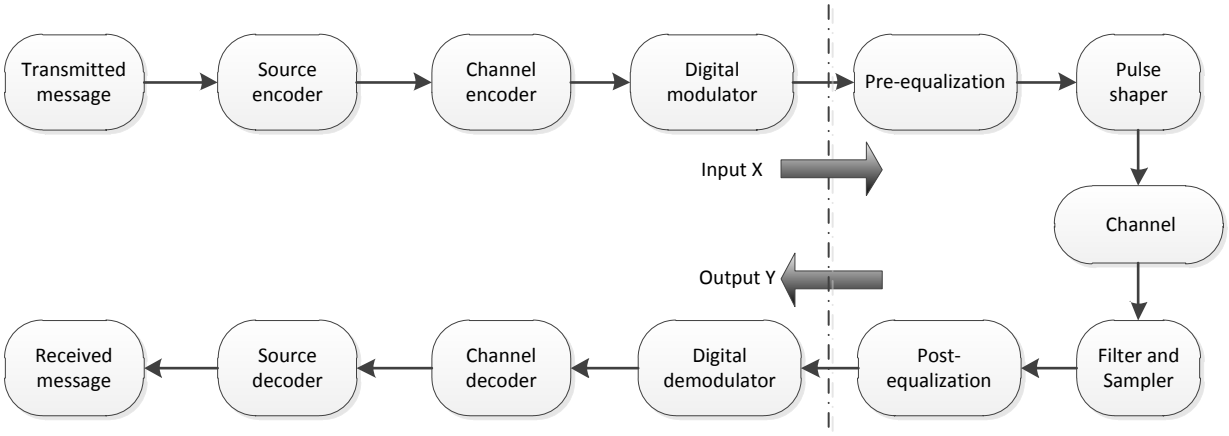

Fig. 1.2 The layered model of a typical digital communication system

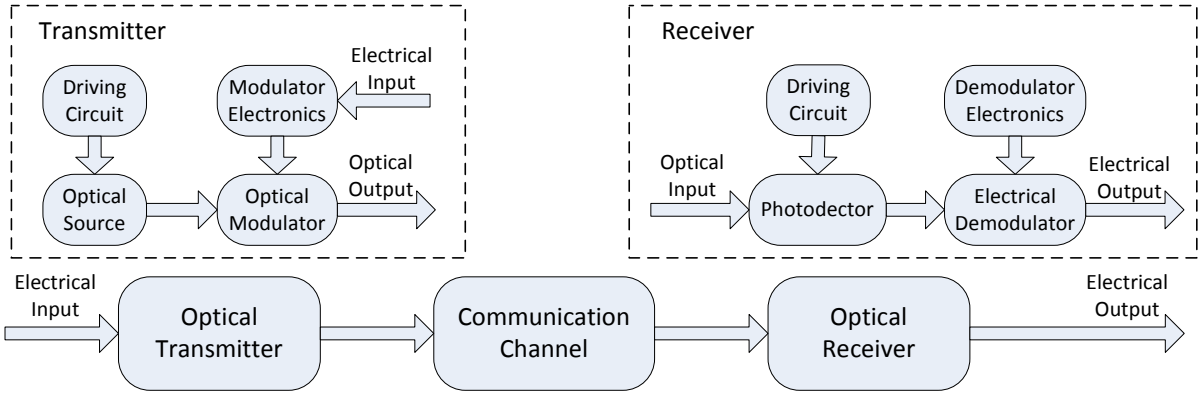

Fig. 1.3 The detailed scheme of a typical fiber-optic communication system

The detailed scheme of the optical components and channels is shown in Fig 1.3. The optical transmitter plays the role of converting electrical signal into optical signal that is suitable for fiber transmission. It includes an optical source such as laser or LED, and a modulator that can modulate the optical carrier with the electrical signal. The optical signal then propagates through the optical fibers and is received at the optical receiver. The photodetector converts the received optical signal into electrical signal and the demodulator extracts the electrical pulse train. Depending on specific modulation formates, the optical transmitter and receiver can be slightly different but with the same general structure [14].

To utilize the large bandwidth provided by optical waves, a practical fiber communication system always incorporates multiple transmission channels instead of a single channel

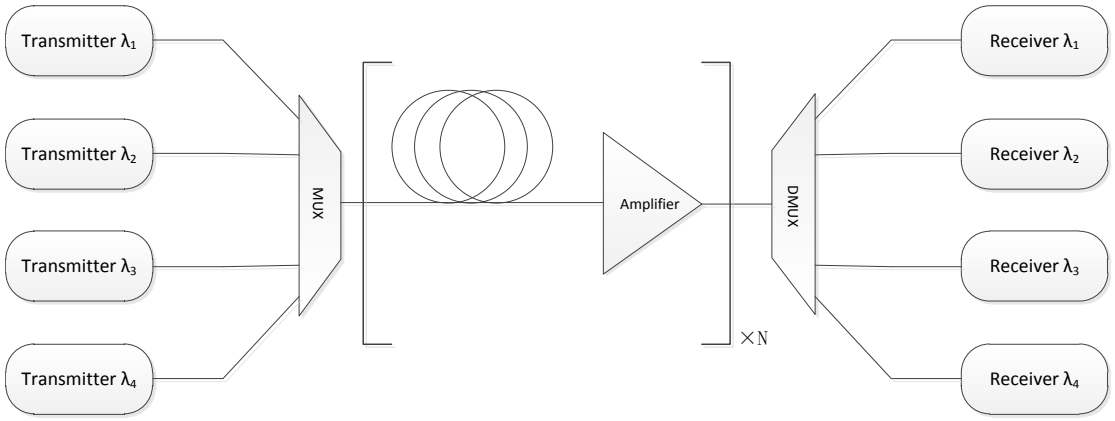

Fig. 1.4 A typical WDM system with four wavelengths

[15]. The rapid growth of information rate for optical fiber communication since 1994 was triggered by the invention of the wavelength division multiplexing (WDM) techniques, whose underlying idea is to transmit multiple data streams over different optical wavelengths in a single mode fiber [1]. A typical WDM architecture with four wavelengths is demonstrated in Fig. 1.4. Dense wavelength division multiplexing (DWDM) systems upgrade the original WDM scheme by adopting denser channel spacing in a single mode fiber [16]. The system throughput of WDM/DWDM is proportional to two factors — wavelength counts and spectral efficiency of each subchannel. On the one hand, current systems incorporate as many as 80-100 subchannels and even more. The degrees of freedom achieved by multiple channels seem unlimited but are actually constrained by practical issues such as the implementation costs in mounting multiple optical components and the physical limitations on device sizes. On the other hand, the task of increasing spectral efficiency are constantly challenging scientists and engineers to develop optoelectronic devices, sophisticated modulation and coding schemes, and digital signal processing techniques.

To aim at higher spectral efficiency, most recent fiber systems are based on coherent detection. Different from noncoherent detection methods which make decisions solely based on the amplitudes of signals, coherent detection retains both the amplitude and phase (complex envelop) of the optical field through the use of in-phase (I) and quadrature (Q) demodulation [17]. The degree of freedom added in the Q channel allows for a two-fold spectral efficiency boost by using advanced IQ-based modulation formats such as QPSK and M ary-QAM instead of OOK. Besides of its advantage in spectral efficiency, coherent detec-

tion can increase the receiver sensitivity over non-coherent detection, thereby increasing the maximum transmission distance. Moreover, coherent detection recovers full information of the optical fields when transforming optical signals into electrical signals. The linear and nonlinear impairments induced in the transmission process including chromatic dispersion and Kerr nonlinearities can be compensated by powerful, versatile and inexpensive digital signal processing (DSP) chips instead of restrictive and costly optical devices [18].

1.1.2 Nonlinear Schrödinger Equation

The propagation of optical pulses through a single mode fiber is governed by the nonlinear Schrödinger equation (NLSE) [3]

$$\frac{\partial A(z,t)}{\partial z} = -\frac{\alpha}{2}A(z,t) - \beta_1\frac{\partial A(z,t)}{\partial t} - \frac{j}{2}\beta_2\frac{\partial^2 A(z,t)}{\partial t^2} + \frac{1}{6}\beta_3\frac{\partial^3 A(z,t)}{\partial t^3} + j\gamma|A(z,t)|^2 A(z,t)\,, \tag{1.1}$$

where $A(z,t)$ is the complex envelope of the slowly varying optical field and z is the propagation distance. Parameters α is the propagation loss, β_1 is the inverse of group velocity, β_2 is the group velocity dispersion (GVD) constant, β_3 is the third-order dispersion, and γ is the Kerr coefficient representing the Kerr nonlinearities in optical fiber. As usual, higher-order dispersion terms are neglected in this equation. To better understand the propagating behaviors of optical waves in fiber, we explain the physical meanings of the above parameters in detail.

The propagation loss factor α measures the loss in power as the optical signal propagates inside the fiber. The power of the output signal P_{out} can be expressed as a function of the power of the input signal P_{in} and the propagation distance L by

$$P_{out} = P_{in}\exp(-\alpha L)\,, \tag{1.2}$$

which means that the intensity of light is attenuated as a function of transmission distance. This propagation loss primarily comes from the material absorption of silica and Rayleigh scattering. Other factors that contribute to the fiber loss include the bending of fiber and light scattering at the core-cladding interface. In early fiber transmission systems, transmission distance is mostly limited by the fiber loss. However, technical achievements

in this field has already reduced this loss from 2 dB/km near 850 nm wavelength for the first generation fiber-optic transmission systems to 0.2 dB/km near 1500 nm wavelength for modern fiber-optic transmission systems. Moreover, Erbium-doped fiber amplifiers (EDFA) are inserted periodically to combat power attenuation in current transmission systems, which enables transmission distances longer than 100 km and avoids the low-speed optical-electrical-optical (OEO) conversion. The introduction of EDFA also enables the WDM technology due its broadband property [19]. Therefore, propagation loss is no longer a major issue for modern fiber-optic transmission systems.

The terms β_1, β_2 and β_3 come from the Taylor series expansion of the mode-propagation constant β at frequency ω_0, that is,

$$\beta(\omega) = n(\omega)\frac{\omega}{c} = \beta_0 + \beta_1(\omega - \omega_0) + \frac{1}{2}\beta_2(\omega - \omega_0)^2 + \frac{1}{6}\beta_3(\omega - \omega_0)^3 + \cdots , \qquad (1.3)$$

where c is speed of light in vacuum, $n(\omega)$ is frequency-dependent refractive index and

$$\beta_m = \left(\frac{\mathrm{d}^m\beta}{\mathrm{d}\omega^m}\right)_{\omega=\omega_0}, \quad m = 0, 1, 2, \cdots . \qquad (1.4)$$

The first-order derivative term β_1 is the inverse of group velocity measuring the moving speed of the pulse envelop. The second-order derivative term β_2, called group velocity dispersion constant, produces a symmetrical broadening of the pulse. The third-order derivation term β_3 causes asymmetrical distortion of pulses and is refereed to as third-order dispersion.

Dispersion, especially the group velocity dispersion that is also refereed to as chromatic dispersion (CD), plays an important role in fiber-optic communication systems. CD causes frequency-dependent group velocity which means that different spectral components travel at different speeds and the optical signal becomes more spread in time. The time-domain broadening of the pulses leads to the overlapping of neighboring symbols, namely, intersymbol interference (ISI). Although dispersion-shifted fibers (DCF) can nullify the dispersion at 1550 nm, zero dispersion is undesired because the fiber nonlinearities would be high. Later, a nonzero dispersion-shifted fiber (NZDSF) was designed to maintain a small amount of residual dispersion at 1550 nm, but its poor performance in nonlinearity tolerance gives rise to standard single mode fibers (SMF) that is already widely used. Various compensation techniques, whether in the optical domain or electrical domain, can be used to further

mitigate the distortions caused by dispersion. Initially, optical approaches such as dispersion compensation fibers (DCF) were popular; however, the emergence of coherent receiver shifts the major interests from optical compensation to electrical compensation in which chromatic dispersion and third-order dispersion can be simultaneously compensated using cost-effective digital signal processing (DSP) algorithms [20].

Indicated by the Kerr coefficient γ, Kerr effect describes the dependence of the refraction index on the instantaneous signal intensity ($|A|^2$), which is the dominant nonlinearity in optical fibers. More specifically, Kerr effect can be decomposed into three general nonlinear effects: self-phase modulation (SPM), cross-phase modulation (XPM) and four wave mixing (FWM). In SPM, the field intensity of the optical field affects its own phase, whereas XPM arises when optical fields with different wavelengths or different polarizations influence the each other's phase. Moreover, three optical fields can interact with each other to generate the fourth wavelength, which is called FWM. A single channel has only SPM effect, whereas the multichannel or multiuser channels suffer from all these interferences due to multi-wavelength multiplexing and co-propagation in the same optical fiber. As signal pulses propagate through the fiber channel, the Kerr effect and optical dispersion interact with each other all the way, causing distortion to the pulse shape and optical field spectrum. These distortions, if casually treated at the receiver, can lead to significant interferences to other wavelengths and successive symbols [21].

For fiber-optic communication systems, Kerr nonlinearities degrades the channel capacity dramatically. As early as in the 1940s, Shannon established the mathematical foundation for modern communication theory, opening a new area of scientific research called information theory [22]. For a particular channel, there is a maximum transmission rate named as capacity under which an arbitrarily low probability of error is obtainable. Shannon proved that the capacity of a discrete-time memoryless additive white Gaussian noise (AWGN) channel is

$$C = W \log_2 \left(1 + \frac{S}{N}\right), \tag{1.5}$$

where W is the bandwidth, S is the transmitted power, N is the noise power and S/N is denoted as signal-to-noise ratio (SNR). The channel capacity increases logarithmically for the AWGN channel as its input power S increases. However, this is not true for the nonlinear fiber channel as show in Fig. 1.5 reproduced from [23]. The capacity is plotted based on the theoretical results derived in [23] (additive noise power I_n is set to 0.026504mW

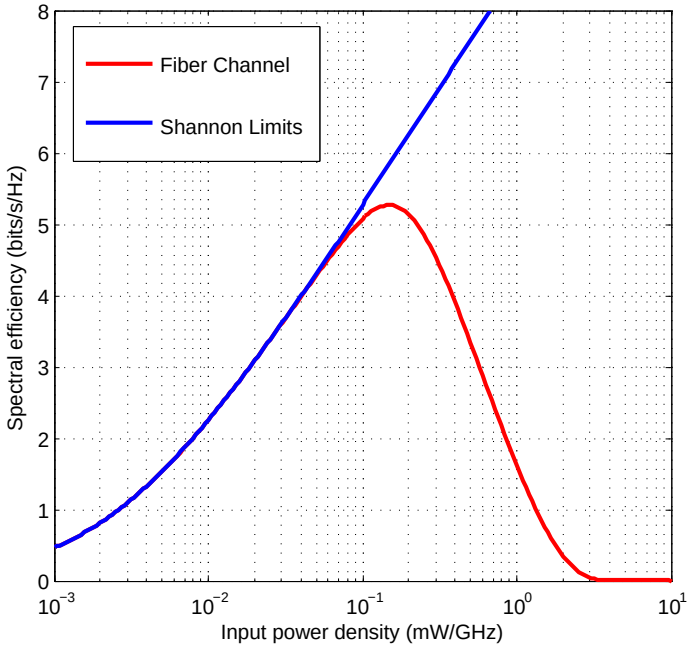

Fig. 1.5 Capacity of a WDM system

and nonlinear intensity scale I_0 is set to 16mW). For fiber-optic communication systems, nonlinear channel capacity has a maximum value at a certain power level instead of going to infinity and beyond as power increases. Instead, large power under nonlinearity can saturate the optical fiber channels and the channel capacity would degrade and even decay to zero.

Apart from Kerr effect, higher order nonlinear effects like self-steepening and Raman scattering also affect pulse propagation, but they are not considered in (1.1). Moreover, amplified spontaneous emission (ASE) noise can affect the performance of fiber communication systems as well. To incorporate all these effects, a generalized NLSE should be used [3]. Nevertheless, for simplicity, we restricted the discussions in this thesis to the most important effects: dispersion and Kerr nonlinearities.

From the NLSE, an analytical solution is only possible under unrealistic assumptions including zero dispersion or zero nonlinearity, or for special waves such as Soliton. Coupled with all the above linear and nonlinear effects, practical fiber-optical communication channels cannot be expressed in closed forms, which sets formidable barriers to the modeling of pulse propagation in optical fiber. However, we can solve the optical field at time t and

distance point z due to an arbitrary optical input by virtue of *numerical* methods such as SSFM and split-step time-domain method.

1.1.3 Split-Step Fourier Method

Let ν_g be the group velocity. By selecting $T = t - z/\nu_g$ as the retarded time that uses the moving pulse as reference, (1.1) becomes

$$\frac{\partial A(z,T)}{\partial z} = (\hat{D} + \hat{N})A(z,T)\,, \tag{1.6}$$

where the linear operator $\hat{D}$ and the nonlinear operator $\hat{N}$ are defined as

$$\hat{D} = -\frac{\alpha}{2} - \frac{j}{2}\beta_2\frac{\partial^2}{\partial T^2} + \frac{1}{6}\beta_3\frac{\partial^3}{\partial T^3}\,, \tag{1.7}$$

$$\hat{N} = j\gamma|A(z,T)|^2\,. \tag{1.8}$$

Generally speaking, numerical approaches to solve NLSE fall in two categories: finite-difference methods and pseudospectral methods. Pseudospectral methods are more efficient and popular. SSFM, as one of the pseudospectral methods, is the most favorable one due to its mathematical simplicity, conceptual clarity and numerical stability [11]. It approximately solves the NLSE based on the premise that within a small distance, the linear operator and the nonlinear operator can be treated independently in a sequential manner, yet with acceptable error. Denoting the small distance as Δz, two implementation methods are usually employed, namely, asymmetric split-step method (A-SSM)

$$A(z+\Delta z,T) \approx \exp(\Delta z\hat{D})\exp(\Delta z\hat{N})A(z,T)\,, \tag{1.9}$$

and symmetric split-step method (S-SSM)

$$A(z+\Delta z,T) \approx \exp(\Delta z\hat{D}/2)\exp(\Delta z\hat{N})\exp(\Delta z\hat{D}/2)A(z,T)\,. \tag{1.10}$$

To compare the accuracy of these two methods, we introduce the Baker-Hausdorff formula [24]

$$\exp(\hat{a})\exp(\hat{b}) = \exp\left(\hat{a}+\hat{b}+\frac{1}{2}[\hat{a},\hat{b}]+\frac{1}{12}[\hat{a}-\hat{b},[\hat{a}.\hat{b}]]+\cdots\right)\,, \tag{1.11}$$

where $[\hat{a}, \hat{b}] = \hat{a}\hat{b} - \hat{b}\hat{a}$ is the commutator. If we expand the linear operator and nonlinear operator in A-SSM and S-SSM, we can readily see

$$e^{\Delta z \hat{D}} e^{\Delta z \hat{N}} = \exp\left(\Delta z(\hat{D} + \hat{N}) + \frac{1}{2}\Delta z^2 \left[\hat{D}, \hat{N}\right] + \cdots\right) , \tag{1.12}$$

$$e^{\Delta z \hat{D}/2} e^{\Delta z \hat{N}} e^{\Delta z \hat{D}/2} = \exp\left(\Delta z(\hat{D} + \hat{N}) + \frac{1}{6}\Delta z^3 \left[\hat{N} + \frac{\hat{D}}{2}, \left[\hat{N}, \frac{\hat{D}}{2}\right]\right] + \cdots\right) . \tag{1.13}$$

It can be seen all other terms except the first one are error terms. Therefore, when the computation requires high accuracy, S-SSM is more accurate than A-SSM since its major error term is a third-order function of Δz.

The accuracy of the SSFM can be further improved by using an iterative procedure to calculate the nonlinear operator [3]. Hereafter, for simplicity, we restrict our discussion to non-iterative S-SSM. Non-iterative symmetric SSFM takes a half step implementing dispersion, and then takes a full step adding nonlinear effects, and finally takes another half step implementing dispersion again [5]:

$$A(z + \Delta z, T) \approx \exp(\Delta z \hat{D}/2) \exp(\Delta z \hat{N}(z + \Delta z/2)) \exp(\Delta z \hat{D}/2) A(z, T) . \tag{1.14}$$

In non-iterative S-SSM, the linear and nonlinear operators are defined as

$$\mathscr{F}\left\{\exp(\Delta z/2\hat{D}) A(z, T)\right\} = \exp\left[-\frac{a\Delta z}{4} + j\left(\frac{\beta_2\omega^2}{2} - \frac{\beta_3\omega^3}{6}\right)\frac{\Delta z}{2}\right] A(z, \omega) , \tag{1.15}$$

$$\hat{N}(z + \Delta z/2) = j\gamma |A(z + \Delta z/2, T)|^2 . \tag{1.16}$$

Accordingly, there are one back-and-forth Fourier transforms during each step adaptation [3].

$$\hat{A}(z + \Delta z/2, \omega) = e^{-\alpha \Delta z/4} H_D(\omega) \mathscr{F}[A(z, T)] , \tag{1.17a}$$

$$\hat{A}_1(z + \Delta z/2, T) = \mathscr{F}^{-1}[\hat{A}(z + \Delta z/2, \omega)] , \tag{1.17b}$$

$$\hat{A}_2(z + \Delta z/2, T) = \exp(j\Delta z\gamma |\hat{A}_1(z + \Delta z/2, T)|^2) \times \hat{A}_1(z + \Delta z/2, T) , \tag{1.17c}$$

$$A(z + \Delta z, \omega) = e^{-\alpha \Delta z/4} H_D(\omega) \mathscr{F}[\hat{A}_2(z + \Delta z/2, T)]. \tag{1.17d}$$

where $\mathscr{F}$ is the Fourier transform operator, $\mathscr{F}^{-1}$ is the inverse Fourier transform operator, and $H_D(\omega)$ is the frequency response including dispersion-related effects, that is,

$$H_D(\omega) = \exp\left[j\left(\frac{\beta_2\omega^2}{2} - \frac{\beta_3\omega^3}{6}\right)\frac{\Delta z}{2}\right], \quad -\omega_s/2 \leq \omega < \omega_s/2\ . \tag{1.18}$$

where ω_s is the sampling frequency. Although Fourier transform and inverse Fourier transform can be implemented efficiently by virtue of FFT and IFFT, the computational cost is still high when a large of number of input samples are processed. In the next section, we will introduce the time-domain split-step approach, which avoids back-and-forth Fourier transforms and therefore reduces computational complexity while maintaining high accuracy.

1.2 Time-Domain Split-Step Methods

Although SSFM is widely used in practice, it suffers from high computational overhead, huge memory usage and time aliasing [12]. Time-domain split-step methods were proposed to overcome these shortcomings.

Firstly, SSFM becomes computationally expensive when it incorporates many steps and for each step, the discrete-time input signal consists of a large number of samples. Even if fast Fourier transforms (FFT) and inverse fast Fourier transforms (IFFT) are used to compute DFT and IDFT, they can still place much overhead on float-point units and system memory. On the contrary, time-domain approaches are exclusively implemented in the time domain instead of toing and froing between the two domains. Therefore, it saves system memory and computer operations by replacing large-point FFT and IFFT with a digital filter that is significantly shorter than the input signal.

Secondly, SSFM has the problem of time aliasing unless it takes large-point DFT's and IDFT's, which also cause high computational complexity and memory usage. After multiplying the DFT of the input sequence with the dispersion filter, the output sequence from IDFT is essentially a circular convolution instead of a linear convolution. The output signal of the latter is always longer than that of the former even if zero-padding is used(The difference is actually equal to the length of the input signal minus one). Therefore, these two convolutions are never identical to each other, implying that time aliasing is inevitable.

It is true that increasing the point number of DFT (by zero-padding) and IDFT (by taking more frequency samples) can overcome this weakness to some extent; however, this would in turn generate more time samples in each step, and therefore causes higher computational complexity and memory usage. In contrast, time-domain approaches work in the time domain exclusively, avoiding the time-aliasing problem accompanied with frequency-domain approaches.

Therefore, time-domain split-step methods are superior to SSFM in computational complexity and numerical accuracy. Time-domain methods replace the dispersion operator with an FIR filter, which means that (1.17a), (1.17b) and (1.17d) are respectively replaced by

$$\hat{A}_1(z+\Delta z/2,T) = h_D(T) \otimes A(z,T)\exp(-\alpha\Delta z/4)\ , \tag{1.19}$$

$$A(z+\Delta z,T) = h_D(T) \otimes \hat{A}_2(z+\Delta z/2,T)\exp(-\alpha\Delta z/4)\ . \tag{1.20}$$

Here, the operator $\otimes$ represents convolution and $h_D(T)$ is a time-domain filter which has the same effects as the frequency response in (1.18). The back-and-forth Fourier transforms and frequency-domain multiplications have been replaced by two time-domain convolutions. In practice, this is implemented in the discrete-time domain, where convolution reduces to convolution sum involving shifts and multiplications [12].

1.3 Contributions in This Thesis

Since all split-step methods for fiber simulations and digital backpropagation algorithms use the same nonlinear models, the problem of designing a time-domain method is reduced to that of finding an optimal or suboptimal discrete-time linear filter with low complexity and yet small error.

An IIR filter takes a restricted form of rational fraction and therefore only allows for either a very small step size or a narrow fitted bandwidth [6]. A "broad-band" FIR filter proposed in [12] matches the desired response over a wider band in a least-square sense, with the error measured at a discrete set of frequency points. Theoretically, the fitting error can be made arbitrarily small by increasing the filter order, so that the step size only needs to ensure that "split-step" itself is reliable. However, this approach measures the errors at a discrete set of frequency points instead of a continuous interval, which introduces an error floor itself. thereby placing a fundamental limit on error performance. The level of

this error floor is determined by the number of points taken in the frequency domain, no matter how large the order of the FIR filter grows. By taking more discrete frequency points, one can lower this error floor but would also increase complexity. A recent method based on Tukey windows applies a Tukey window in the frequency domain, transforms back to the time domain and then multiplies another discrete-time Tukey window [13]. This controls the fitting error efficiently and reduces the filter order at the same time Furthermore, the step size and the filter length can be optimized to minimize the overall computational complexity. Nonetheless, this approach takes the restricted forms of double Tukey windows, which limits the freedom of design to two parameters. Since the error is a highly nonlinear function of these parameters, no closed-form solution exists and an exhaustive search can be time-consuming.

In this thesis, we focus on the optimal design of an FIR filter used as the time-domain implementation for the dispersion and dispersion slope characteristics. Our objective is to minimize the integral of the squared error between the frequency response of the FIR filter and the desired response over the band of interest. Unlike the work in [12], the sum of errors taken at discrete points is replaced by a numerical integral, which can be computed based on adaptive integration techniques such as Gauss-Kronrod quadrature formula [25]. This reduces the error floor to the order of 10^{-15}. Moreover, since no structural constraint is imposed on the FIR filter, this approach can explore all the degrees of freedom provided by the FIR filter. This least-square problem can be solved in two different approaches.

In the first approach, this problem is reduced to that of solving a system of linear equations, i.e., the normal equation. Its Toeplitz structure enables fast computation based on a recursive Levinsion-Durbin-like algorithm. This algorithm explicitly generates the optimal filters of order $1, 2, \ldots, n$ in an iterative manner, with a total computational complexity of $O(n^2)$. Henceforth, an implicit search for optimal order is naturally included, whereas searching from 1 to n based on the Gaussian elimination method requires $O(n^4)$. In the second approach, the solution is derived based on the singular value decomposition (SVD) of a quasi-matrix. This approach provides geometrical, physical and mathematical insights into this problem and its solution. Geometrically, we find that the frequency response of the optimal filter is the orthogonal projection from the desired dispersion filter to the subspace spanned by a subset of discrete prolate spheroidal wave functions (DPSWF). In parallel, the optimal filter is a linear combination of the time-domain counterparts of these DPSWFs, namely, a set of index-limited discrete prolate spheroidal sequences (DPSS).

Mathematically, the theory of DPSS and DPSWF reveals that in the previous normal equation approach, the system of linear equations is usually ill-conditioned, which is not recognized in previous works such as [12]. Under such circumstances, the solution could be sensitive to numerical errors and could also generate overshoots outside the band of interest, which can be transformed back in band by the nonlinear operations.

With this consideration in mind, we proposed two approaches to mitigate these shortcomings. Firstly, we add a regularization term to the objective function to provide robustness for the solution. The resulting filter can also suppress overshoots by increasing its length; however, we improve it by imposing a maximum magnitude constraint on the frequency response to control overshoots more efficiently. The resulting quadratically constrained quadratic programming (QCQP) problem can be readily solved by state-of-the-art interior-point methods [26, 27]. The single channel and wavelength-division multiplexing (WDM) simulations verify that the output signals generated by the proposed regularized LS filter and QCQP-based filter are almost the same as those by SSFM, even after propagating thousands of kilometers. Afterwards, the proposed filters are used in time-domain digital backpropagation algorithms to mitigate the impairments caused by dispersion and nonlinearities [28, 29, 30].

For a given error tolerance, we establish the relationship between the required filter order and several parameters both theoretically and numerically. Based on the one-to-one correspondence between group delay and instantaneous frequency, we derive a tight lower bound of the filter order as a linear function of the step size, whose validity is also verified by numerical experiments. This can simplify the task of choosing the step size from the perspective of reducing computational complexity. As the step size increases, the filter order also increases and the total number of split steps decreases. Consequently, the total computational complexity of linear convolutions is approximately a constant or more strictly, on the same order. Therefore, a constant step can be simply chosen as the maximum value allowed by the "split-step" itself [3, 5].

The proposed optimal filters reduce the total computational complexity, both when extracting the filter and implementing linear convolutions. On the one hand, the unconstrained regularized LS filter is the solution of a Toeplitz system. This enables a fast modified Levinson-Durbin algorithm with the complexity of $O(n^2)$. The QCQP-based filter can be computed with efficient interior-point methods. On the other hand, the computational complexity of linear convolutions depends exclusively on the filter length. Numerical

simulations show that the QCQP-based filter saves at least 1/3 of the total filter order when compared with most recent work [13]. Moreover, there is a complexity trade-off between the unconstrained regularized LS filter and QCQP-based filter if overshoot control is required: the former is easier to extract but the latter is shorter. The choice depends on whether the filter needs to be regenerated frequently or not. In addition, we also introduce the overlap-add method that can reduce the computational complexity of the linear convolutions from $O(PM')$ to $O(P(\log M'))$, where P is the length of input signal and M' is the filter order.

The following notations are used throughout this thesis: an italic letter represents a scalar; a boldface lowercase letter refers to a vector; a boldface uppercase letter denotes a matrix. $(\cdot)^T$ or $(\cdot)^H$ represents transpose or conjugate transpose of a matrix. $\|\cdot\|$ means the norm of a vector or matrix. $\mathbf{I}$ is an identity matrix of appropriate dimension. The operator $\otimes$ represents convolution. The imaginary unit is denoted by j; $\Re(\cdot)$ and $\Im(\cdot)$ are real and imaginary parts of a complex number; $|\cdot|$ and $\angle(\cdot)$ represents the magnitude and phase.

1.4 Organization of Thesis

The thesis is organized as follows:

- **Chapter 2: Optimal Design of the Dispersion FIR Filter**

  In this chapter, we formulate the optimization problem and develop two different methods to solve this problem: one is the normal equation approach and the other is based on the singular value decomposition (SVD). The former is more suitable for implementation, whereas the latter provides geometrical and mathematical insights into the solution of this problem based on the theory of DPSS and DPSWF.

- **Chapter 3: Numerical Issues and Modified Filters**

  This chapter begins with theoretical and numerical experiments that reveal the ill-conditioned property of the LS problem in certain instances. The LS filter could also generate overshoots that lead to unreliable results after propagating through long distances. Two modified filters are introduced to mitigate these numerical problems. We first add a regularization term to the objective function to provide robustness and

introduce the fast MLD algorithm. Then, to suppress overshoots more efficiently, a maximum magnitude constraint is enforced on the frequency response, which can be formulated into a standard QCQP problem. The issue of filter order is discussed at the end of this chapter.

- **Chapter 4: Time-Domain Simulations of Pulse Propagation in Optical Fiber**

  The proposed filters are verified based on several simulations of single channel and WDM systems. We also introduce the overlap-add and overlap-save method that can reduce the computational complexity significantly.

- **Chapter 5: Time-Domain Backpropagation for Fiber Impairment Compensation**

  In this chapter, the proposed filters are used to design time-domain digital backpropagation algorithms. The simulation results show that these algorithms can effectively compensate dispersion and nonlinearities, thereby improve the performance of fiber-optic communication systems.

- **Chapter 6: Conclusions**

Chapter 2

Optimal Design of the Dispersion FIR Filter

In this chapter, we consider the optimal design of the dispersion FIR filter in a least square sense. We will consider two approaches to solve this problem, one is normal equation and the other is based on the quasi-matrix SVD. The resulting optimal solution is discussed and analyzed based on the theory of DPSS.

2.1 Problem Formulation

Since nonlinear processing does not vary from one to another approach in time-domain split-step methods, the main problem is to design a discrete-time filter which mimics the dispersion and dispersion slope characteristics in (1.18). In other words, here we want to design an FIR filter with unit response $h(n)$ whose frequency response is close to that of desired response $h_D(n)$. For notational simplicity and discussion convenience, we assume that the index n takes the integer values from $-M$ to M and thus the number of order is $2M+1$. This is from the consideration that second-order dispersion often dominates over third-order terms and therefore $h(n)$ is close to being symmetric, that is, $h(-n) \approx h(n)$. The discrete-time Fourier transform (DTFT) of $h(n)$ is

$$H(e^{j\omega}) = \sum_{n=-M}^{M} h(n) e^{-jn\omega} , \quad -\pi \leq \omega < \pi . \tag{2.1}$$

This expression is with respect to the normalized frequency ranging from $-\pi$ to π. To be consistent with the literature, we rewrite it as a function of the physical frequency as follows

$$H(e^{j\omega T_s}) = \sum_{n=-M}^{M} h(n)e^{-jn\omega T_s} = \mathbf{a}^H(\omega)\,\mathbf{h}\,, -\frac{\omega_s}{2} \le \omega < \frac{\omega_s}{2}\,, \tag{2.2}$$

where T_s is the sampling period, and ω_s is the sampling frequency satisfying $\omega_s T_s = 2\pi$. We define

$$\mathbf{h} = [h(-M), h(-M+1), \cdots, h(M)]^T\,, \tag{2.3}$$

$$\mathbf{a}(\omega) = [e^{-j\omega MT_s}, e^{-j\omega(M-1)T_s}, \cdots, e^{j\omega MT_s}]^T\,, \tag{2.4}$$

where $(\cdot)^T$ represents transpose of a matrix. A good FIR filter should match the desired dispersion and dispersion slope characteristics as much as possible. Herein, we are interested in a partial band, $[-\omega_c, \omega_c]$, instead of the whole band $[-\omega_s/2, \omega_s/2]$. The effective bandwidth ratio is defined as $\mu = 2\omega_c/\omega_s \le 1$ and this versatile formulation includes the whole band as a special case when $\mu = 1$. The rationale of this "partial-band" approach is to "squeeze" the ripples out of the band of interest by sacrificing the uninterested band, which will become evident later.

In order to formulate a mathematically tractable problem, we use the squared error as the measure of the difference between the frequency response in (2.2) and the desired response $H_d(\omega)$ in (1.18). Our objective is to minimize the integral of this squared error over the frequency range of interest, that is

$$\begin{aligned} E_{\mathrm{LS}}(\mathbf{h}) &= \frac{1}{2\omega_c}\int_{-\omega_c}^{\omega_c} \left|H_D(\omega) - H(e^{j\omega T_s})\right|^2 d\omega \\ &= 1 - \mathbf{b}^H\mathbf{h} - \mathbf{h}^H\mathbf{b} + \mathbf{h}^H\mathbf{A}\mathbf{h}\,, \end{aligned} \tag{2.5}$$

where $\mathbf{b} \in \mathbb{C}^{(2M+1)\times 1}$ and $\mathbf{A} \in \mathbb{C}^{(2M+1)\times(2M+1)}$ are defined as follows

$$\mathbf{b} = \frac{1}{2\omega_c}\int_{-\omega_c}^{\omega_c} H_D(\omega)\mathbf{a}(\omega)d\omega\,, \tag{2.6}$$

$$\mathbf{A} = \frac{1}{2\omega_c}\int_{-\omega_c}^{\omega_c} \mathbf{a}(\omega)\mathbf{a}^H(\omega)d\omega\,. \tag{2.7}$$

Note that $\mathbf{b}$ is actually equal to $1/\mu$ times the inverse DTFT of $H_D(\omega)$, which implies that

it is a truncation of the sequence whose spectrum is the desired response, except by a scale factor of $1/\mu$. We also introduce an alternative notation for this problem based on the concept of quasi-matrix from Stewart[31], Battles and Trefethen[32, 33]. Herein, we define an "$\infty \times (2M+1)$" matrix $\mathbf{F}$ whose "columns" are the continuous function, $\mathbf{f}_k(\omega)$, namely,

$$\mathbf{F} = [\mathbf{f}_{-M}, \mathbf{f}_{-M+1}, \cdots, \mathbf{f}_M] \ , \tag{2.8}$$

where $\mathbf{f}_k$ is not a vector in the Euclidean space, but an infinite-dimensional "vector" whose entries are all the function values of $f_k(\omega)$ in the interval $[-\omega_c, \omega_c]$:

$$f_k(\omega) = e^{-jk\omega T_s}, \quad -\omega_c \leq \omega \leq \omega_c \ . \tag{2.9}$$

These "vectors" satisfy

$$\|\mathbf{f}_k\|_2^2 = \int_{-\omega_c}^{\omega_c} |f_k(\omega)|^2 \, \mathrm{d}\omega = \int_{-\omega_c}^{\omega_c} f_k(\omega) f_k^*(\omega) \mathrm{d}\omega < \infty \ , \tag{2.10}$$

and form a subspace of the Hilbert space $\mathcal{L}^2$ (the space of complex signals with finite energy on the interval $[-\omega_c, \omega_c]$), denoted as $\mathcal{F}^2$. The inner product of two infinite-dimensional vectors, $\mathbf{g}_1$ and $\mathbf{g}_2$, is defined as

$$(\mathbf{g}_1, \mathbf{g}_2) = \mathbf{g}_2^H \mathbf{g}_1 = \int_{-\omega_c}^{\omega_c} g_1(\omega) g_2^*(\omega) \mathrm{d}\omega \ . \tag{2.11}$$

Similarly, $\mathbf{H}_D$ is a "$\infty \times 1$" quasi-vector whose "column" is the desired response $H_D(e^{j\omega})$, where $\omega \in [-\omega_c, \omega_c]$. Therefore, the problem of designing the FIR filter reduces to that of finding an $\mathbf{h}$ such that

$$\mathbf{F}\mathbf{h} = \mathbf{H}_D \ . \tag{2.12}$$

This is an overdetermined equation, i.e., there are a finite number of unknowns but an infinite number of equations. Two cases can arise for the solutions:

1. If $\mathbf{H}_D \in \mathcal{F}^2$, there exists an exact solution to this problem. However, this rarely holds for overdetermined problems, especially infinite-dimensional ones.

2. If $\mathbf{H}_D \notin \mathcal{F}^2$, no exact solution exists for this problem. In this case, we can only find a $\mathbf{h}$ such that the two sides of this equation is as close as possible.

If we use the 2-norm of the error vector to measure the difference, the error function can be rewritten as

$$E_{LS}(\mathbf{h}) = \frac{1}{2\omega_c}\|\mathbf{H}_D - \mathbf{F}\mathbf{h}\|_2^2 \,, \tag{2.13}$$

which is in essence the same as (2.5). The relationship between between $\mathbf{A}$, $\mathbf{b}$ and $\mathbf{F}$, $\mathbf{H}_D$ are

$$\mathbf{A} = \frac{1}{2\omega_c}\mathbf{F}^H\mathbf{F} \,, \tag{2.14}$$

$$\mathbf{b} = \frac{1}{2\omega_c}\mathbf{F}^H\mathbf{H}_D \,. \tag{2.15}$$

Note that the above quasi-matrix descriptions are for notational convenience only, more rigorous mathematical formulation can be built based on the SVD of bounded operators on Hilbert spaces[34].

Therefore, the problem of FIR filter design reduces to an unconstrained optimization problem, namely,

$$\min_{\mathbf{h}} E_{LS}(\mathbf{h}) \,. \tag{2.16}$$

It is worth mentioning that the "broad-band" FIR approach proposed in [12] solves a similar problem. However, the objective function is taken as the sum of squared errors at a set of uniformly sampled frequency points, which can introduce an error floor itself, thereby placing a fundamental limit on the error performance. Our approach uses the integral over the band of interest which is more accurate. More importantly, we realize the importance of DPSWF and DPSS in this problem, explain the physical meanings behind its solution, and discover the ill-conditioned nature of this problem, which is missing from previous work. We also recognize the Toeplitz structure and provide a significantly faster recursive implementation.

2.2 Optimal Solution of Unconstrained Filter Design

2.2.1 Normal Equation Approach

The unconstrained minimization problem in (2.16) is a convex optimization problem since $\mathbf{A}$ is Hermitian positive definite (as shown later). Henceforth, its local minimizer (stationary

point) is also the global minimizer, i.e., we can find the solution by taking the derivative of (2.5) with respect to $\mathbf{h}^*$ and setting it to zero[35]. From this, the optimal solution satisfies the following *normal equation*,

$$\mathbf{A}\mathbf{h}_o = \mathbf{b} \ , \tag{2.17}$$

and the error is

$$E_{\min} = 1 - \mathbf{b}^H \mathbf{A}^{-1} \mathbf{b} \ . \tag{2.18}$$

The equation in (2.17) has a unique solution because $\mathbf{A}$ is nonsingular. However, $\mathbf{b}$ can only be calculated numerically and is thus subject to small perturbation. If $\mathbf{A}$ is well-conditioned, the solution can be obtained as $\mathbf{h}_o = \mathbf{A}^{-1}\mathbf{b}$, based on any standard method of solving linear equations. Nonetheless, if $\mathbf{A}$ is ill-conditioned, the solution may deviate from the exact solution dramatically. The entries of $\mathbf{A}$ can be derived from (2.7), expressed as

$$\begin{gathered} A_{mn} = \frac{1}{2\omega_c} \int_{-\omega_c}^{\omega_c} a_m(\omega) a_n^*(\omega) d\omega = \frac{\sin[\omega_c (m-n) T_s]}{\omega_c (m-n) T_s} \ , \\ m, n = 1, 2, \cdots, 2M+1 \ , \end{gathered} \tag{2.19}$$

where $(\cdot)_{mn}$ represents the entry of a matrix in the mth row and nth column. Obviously, $\mathbf{A}$ is a Hermitian (symmetric) Toeplitz matrix. As it will be shown later, this Toeplitz property enables an implementation of solving (2.17) that is dramatically faster than Gaussian elimination. From (2.6) and (2.4), each entry of the column vector $\mathbf{b}$ is given by

$$\begin{gathered} b_k = \frac{1}{2\omega_c} \int_{-\omega_c}^{\omega_c} \exp\left[\left(j\frac{\beta_2 \omega^2}{2} - j\frac{\beta_3 \omega^3}{6} \right) \Delta z + j\omega k T_s \right] d\omega \ , \\ k = -M, -(M-1), \cdots, M-1, M \ . \end{gathered} \tag{2.20}$$

A broad family of algorithms are available for calculating the integral numerically[25]. We use a high-order global adaptive quadrature method with a given (relative) error tolerance as low as 10^{-15}. Specifically, Gauss-Kronrod quadrature formulas are used, which is the most efficient for oscillatory integrands. This approach reduces the numerical error, while at the same time reduces computational complexity when compared to the discrete-frequency approach in [12] which is actually a Riemann integral.

2.2.2 Quasi-Matrix SVD Approach

The singular value decomposition (SVD) can be generalized to the quasi-matrix $\mathbf{F}$, which is in essence the SVD of the bounded operator $\mathbf{F}$ on a Hilbert space. Here, we explain this concept in a leisure style. The quasi-matrix $\mathbf{F}$ can take the following form[33],

$$\mathbf{F} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H \ , \tag{2.21}$$

where $\mathbf{U}$ is an "$\infty \times (2M+1)$" quasi-matrix, $\mathbf{V}$ is a "$(2M+1) \times (2M+1)$" unitary matrix, and $\mathbf{\Sigma}$ is an "$(2M+1) \times (2M+1)$" diagonal matrix with its diagonal entries, $i \neq j$, and $\langle \mathbf{\Sigma} \rangle_{ii} = \sigma_i \geq 0$, $i = 1, 2, \cdots, (2M+1)$, sorted in a decreasing order. The column "vectors" of the quasi-matrix $\mathbf{U}$ are a set of orthogonal basis functions $U_0(\omega), \cdots, U_{2M}(\omega)$ (or interchangeably, $\mathbf{U}_0, \ldots, \mathbf{U}_{2M}$) for $\mathcal{F}^2$, i.e., the subspace spanned by the column vectors of $\mathbf{F}$. Then we define orthogonal complement of this subspace with respect to $\mathcal{L}^2$ as $\mathcal{O}^2$. The columns of the matrix $\mathbf{V}$, namely, $\mathbf{v}_0, \cdots, \mathbf{v}_{2M}$, are a set of orthonormal index-limited sequences. The singular values σ_i are all positive and we postpone the explanation later. The relationship between the sequence $\mathbf{v}_k$ and the function U_k is

$$\mathbf{F}\mathbf{v}_k = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H\mathbf{v}_k = \sigma_k \mathbf{U}_k \ . \tag{2.22}$$

This means that $U_k(\omega)$ is the discrete-time Fourier transform of $\mathbf{v}_k$ scaled by a factor of $1/\sigma_k$.

It is well-known that the solution of the LS problem in (2.13) based on the SVD is[36, 33]

$$\mathbf{h}_o = \mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{U}^H\mathbf{H}_D \ . \tag{2.23}$$

Here, we have used the fact that $\mathbf{\Sigma}$ is of full rank. Substituting this solution into (2.13), the corresponding error becomes

$$E_{LS}(\mathbf{h}_o) = \frac{1}{2\omega_c} \|(\mathbf{I} - \mathbf{U}\mathbf{U}^H)\mathbf{H}_D\|_2^2 \ , \tag{2.24}$$

where $\mathbf{I}$ is an "$\infty \times \infty$" quasi-matrix. We define the residual vector as

$$\mathbf{r} = \mathbf{H}_D - \mathbf{F}\mathbf{h}_o = (\mathbf{I} - \mathbf{U}\mathbf{U}^H)\mathbf{H}_D \ . \tag{2.25}$$

Here, we have used the fact that $\mathbf{\Sigma}$ is of full rank. Based upon the quasi-matrix SVD of $\mathbf{F}$, we can obtain

$$\mathbf{A} = \frac{1}{2\omega_c}\mathbf{F}^H\mathbf{F} = \frac{1}{2\omega_c}\mathbf{V}\mathbf{\Sigma}^H\mathbf{\Sigma}\mathbf{V}^H \ , \tag{2.26}$$

$$\mathbf{b} = \frac{1}{2\omega_c}\mathbf{F}^H\mathbf{H}_D = \frac{1}{2\omega_c}\mathbf{V}\mathbf{\Sigma}\mathbf{U}^H\mathbf{H}_D \ . \tag{2.27}$$

Substituting (2.26) and (2.27) into the solution of the normal equation $\mathbf{h}_o = \mathbf{A}^{-1}\mathbf{b}$, we can observe that these two approaches lead to exactly the same solution. Actually, these two methods are related to each other. The merit of this SVD approach will not become evident until we introduce the concept of DPSS and DPSWF.

2.3 DPSS's and DPSWF's

Although the optimal solution has been derived based on the SVD of the quasi-matrix $\mathbf{F}$, the matrices $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{\Sigma}$ remain unknown. Interestingly, these matrices are expressed by the famous discrete prolate spheroidal sequences (DPSS's) and discrete prolate spheroidal wave functions (DPSWF's). Fourier theory establishes a fact that, except for the all-zero sequence, a sequence can not be both index-limited and band-limited. However, this property is often desirable in many applications. As early as in 1978, Slepian investigated the extent to which a time-domain sequence can be concentrated both in a finite index set and in a subinterval of the fundamental period of the spectrum[37]. This leads to the theory of DPSS and DPSWF, which explains the fundamental aspects of our problem.

The theoretical framework of DPSS starts from solving an optimization problem. In a strict sense, a time-domain sequence is band-limited with a bandwidth of $2\omega_c$ if its spectrum vanishes for $\omega_c < |\omega| < \omega_s/2$. Since the sequence $h(n)$ is index-limited in the interval $-M \leq n \leq M$, it is impossible for its frequency spectrum (response) $H(e^{j\omega T_s})$ as defined by (2.2) to satisfy this condition. A parameter λ can be introduced to describe the extent to which $H(e^{j\omega T_s})$ is concentrated within the frequency interval $[-\omega_c, \omega_c]$. It is defined as the ratio of the energy in this band and the total energy in the band $[-\omega_s/2, \omega_s/2]$, that is,

$$\lambda = \frac{\int_{-\omega_c}^{\omega_c} |H(e^{j\omega T_s})|^2 d\omega}{\int_{-\omega_s/2}^{\omega_s/2} |H(e^{j\omega T_s})|^2 d\omega} \ . \tag{2.28}$$

The objective is to find an index-limited sequence $h(n)$, so as to maximize this spectral concentration factor λ.

The answer turns out to be elegant and insightful. It not only solves this single problem, but also provides the framework to tackle a variety of problems. Specifically, it generates the SVD of the quasi-matrix $\mathbf{F}$ and gives some insights into the behaviors of the singular values of $\mathbf{F}$ and the eigenvalues of $\mathbf{A}$, which would otherwise not be obvious. The key to understanding this problem is finally stated in terms of $2M+1$ nonzero concentration factors, λ_k, their associated index-limited DPSS's, $v_k(n)$, and DPSWF's, $\psi_k(\omega)$, for $k = 0, 1, \cdots, 2M$. These values, sequences and functions satisfy the following attractive properties:

1. There are totally $2M+1$ distinct, real and positive concentration factors, all between 0 and 1, expressed in order as

$$1 > \lambda_0 > \lambda_1 > \cdots \gg \cdots > \lambda_{2M} > 0 \, . \tag{2.29}$$

   More importantly, a majority of them are distributed in two clusters, approximately $\lceil (2M+1)\mu \rceil$ of them stay near 1 and the others near 0, where $\lceil x \rceil$ is the smallest integer not less than x. A very small number of λ's are intermediate values.

2. The DPSS's $v_k(n)$ for $k = 0, 1, \cdots, 2M$, are a set of orthonormal *real* basis sequences for the $(2M+1)$-dimensional sequence space, that is,

$$\mathbf{v}_p^H \mathbf{v}_q = \sum_{n=-M}^{M} v_p(n) v_q(n) = \delta_{pq} \, , \tag{2.30}$$

   for any two integers p, $q = 0, 1, \cdots, 2M$, and $\mathbf{v}_k = [v_k(-M), v_k(-M+1), \cdots, v_k(M)]^T$. The function δ_{pq} is the Kronecker delta whose value is 1 if $p = q$ and zero otherwise. The concentration factor associated with the band $[-\omega_c, \omega_c]$ for the sequence $v_k(n)$ is equal to λ_k. This implies a surprising but simple fact: the space of index-limited sequences who are *almost* band-limited, i.e., constrained in the time-frequency box $[-MT_s, MT_s] \times [-\omega_c, \omega_c]$, is not a trivial subspace (zero subspace), but has an approximate dimension of $\lceil (2M+1)\mu \rceil$. The first $\lceil (2M+1)\mu \rceil$ DPSS's whose λ values are close to 1 serve as a set of basis sequences for this subspace.

3. The DPSWF $\psi_k(\omega)$ is the DTFT of the corresponding DPSS $v_k(n)$, up to a complex scalar, namely,

$$\psi_k(\omega) = \epsilon_k \sum_{n=-M}^{M} v_k(n) e^{jn\omega T_s} \ , \tag{2.31}$$

where $\epsilon_k = j$ (imaginary unit) for k odd, and $\epsilon_k = 1$ for k even and this plays the role in making both $\mathbf{v}_k$ and $\psi_k(\omega)$ real. The IDTFT expression is

$$v_k(n) = \frac{1}{\epsilon_k \omega_s} \int_{-\omega_s/2}^{\omega_s/2} \psi_k(\omega) e^{-jn\omega T_s} d\omega \ . \tag{2.32}$$

Since the DPSS's are orthonormal to each other, the DPSWF's satisfy the similar property because Fourier transform preserves orthogonality, that is

$$\frac{1}{\omega_s} \int_{-\omega_s/2}^{\omega_s/2} \psi_p(\omega)\psi_q(\omega) \mathrm{d}\omega = \delta_{pq} \ . \tag{2.33}$$

Moreover, the DPSWF's are also orthogonal in the interval of $[-\omega_c, \omega_c]$, which implies

$$\frac{1}{\omega_s} \int_{-\omega_c}^{\omega_c} \psi_p(\omega)\psi_q(\omega) \mathrm{d}\omega = \lambda_p \delta_{pq} \ . \tag{2.34}$$

4. The DPSS's and DPSWF's satisfy the following sum and integral equations:

$$\sum_{m=-M}^{M} \frac{\sin[\omega_c(n-m)T_s]}{\frac{\omega_s}{2}(n-m)T_s} v_k(m) = \lambda_k v_k(n) \ , \tag{2.35}$$

$$\frac{1}{\omega_s} \int_{-\omega_c}^{\omega_c} \frac{\sin\left[T_s(\omega-\omega')\frac{2M+1}{2}\right]}{\sin\left[T_s(\omega-\omega')\frac{1}{2}\right]} \psi_k(\omega') d\omega' = \lambda_k \psi_k(\omega) \ . \tag{2.36}$$

More specifically, the DPSS's are the eigenvectors of a matrix (actually, a scaled version of $\mathbf{A}$) and the DPSWF's are the eigenfunctions of the integral equation.

From (2.34), we know that the DPSWF's form a set of orthogonal basis functions for the subspace $\mathcal{F}^2$. They can be normalized to a set of orthonormal basis functions $U_k(\omega)$ in

the following form

$$U_k(\omega) = \frac{1}{\epsilon_k \sqrt{\omega_s}} \frac{\psi_k(\omega)}{\sqrt{\lambda_k}} , \quad -\omega_c \leq \omega \leq \omega_c . \tag{2.37}$$

This functional subspace is the frequency-domain counterpart of the $(2M+1)$-dimensional space of index-limited sequences. The Fourier-pair relationship in (2.31) can be written in a compact form

$$\mathbf{F}[\mathbf{v}_0, \ldots, \mathbf{v}_{2M}] = \sqrt{\omega_s}[\mathbf{U}_0, \cdots, \mathbf{U}_{2M}]\mathbf{\Lambda}^{1/2} , \tag{2.38}$$

where $\mathbf{\Lambda} = \text{diag}\{\lambda_o, \ldots, \lambda_{2M}\}$ is a diagonal matrix with λ_k as its diagonal entries, and $\mathbf{U}_k$ is the quasi-vector containing all the function values of $U_k(\omega)$ in $[-\omega_c, \omega_c]$. The square root of a diagonal matrix is defined by taking the squared root of its diagonal entries. Define $\mathbf{V} = [\mathbf{v}_0, \ldots, \mathbf{v}_{2M}]$ and $\mathbf{U} = [\mathbf{U}_0, \ldots, \mathbf{U}_{2M}]$, and right multiply $\mathbf{V}^H$ to both sides of (2.38), we can obtain the final expression for the SVD of $\mathbf{F}$ as

$$\mathbf{F} = \sqrt{\omega_s}\, \mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{V}^H . \tag{2.39}$$

The singular matrix $\mathbf{\Sigma}$ defined in (2.21) is equal to $\sqrt{\omega_s}\mathbf{\Lambda}^{1/2}$. Hence, we have already derived the explicit expression for the SVD of the quasi-matrix $\mathbf{F}$, based on the properties of DPSS's and DPSWF's.[1] From either (2.19) and (2.35), or the above SVD form and (2.26), we also have

$$\mathbf{A}\mathbf{v}_k = \frac{\lambda_k}{\mu}\mathbf{v}_k , \tag{2.40}$$

which means that $\mathbf{v}_k$'s are the eigenvectors of matrix $\mathbf{A}$ in (2.7), and λ_k/μ are the corresponding eigenvalues. This means that $\mathbf{A}$ is Hermitian positive definite.

2.4 Geometrical Explanation of the Optimal Solution

Thus far, we can rewrite the optimal solution in (2.23) as

$$\mathbf{h}_o = \sum_{k=0}^{2M} \left(\mathbf{U}_k^H \mathbf{H}_D\right) \times \frac{\mathbf{v}_k}{\sqrt{\omega_s \lambda_k}} = \sum_{k=0}^{2M} h'_k \times \frac{\mathbf{v}_k}{\sqrt{\omega_s \lambda_k}} . \tag{2.41}$$

[1] However, the DPSS's and DPSWF's themselves do not have closed-form expressions except for empirical formulas. Therefore, the SVD approach can only provides geometrical and mathematical insights. Computing the optimal filter still requires solving the normal equation.

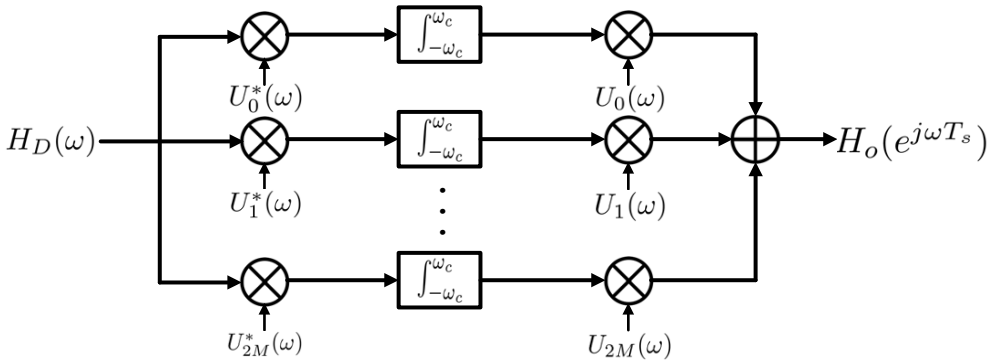

Fig. 2.1 Schematic block diagram of the least square filter design with $H_D(\omega)$ as input and $H_o(e^{j\omega T_s})$ as output, which is the frequency response of the optimal filter.

The quantity $h'_k = \mathbf{U}_k^H \mathbf{H}_D$ represents the kth coordinate of $\mathbf{H}_D$ along the direction of $\mathbf{U}_k$, which is obtained by computing the inner product of the desired vector $\mathbf{H}_D$ and the unit vector $\mathbf{U}_k(\omega)$. The vector $\mathbf{v}_k/\sqrt{\omega_s \lambda_k}$ is the time-domain corresponding vector of the frequency response $\mathbf{U}_k$. Its geometrical meaning becomes more evident when considering the frequency response of $\mathbf{h}_o$, which is

$$\begin{aligned} \mathbf{H}_o(e^{j\omega T_s}) &= \mathbf{F}\mathbf{h}_o = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H\mathbf{V}\boldsymbol{\Sigma}^{-1}\mathbf{U}^H\mathbf{H}_D(\omega) \\ &= \mathbf{U}\mathbf{U}^H\mathbf{H}_D(\omega) = \sum_{k=0}^{2M} \left(\mathbf{U}_k^H \mathbf{H}_D\right) \times \mathbf{U}_k \ , \end{aligned} \tag{2.42}$$

where the sum expression is the frequency domain counterpart of (2.41). We recognize that the matrix $\mathbf{U}\mathbf{U}^H$ is the projection matrix onto the subspace $\mathcal{F}^2 = \text{span}\{\mathbf{f}_{-M}, \mathbf{f}_{-M+1}, \cdots, \mathbf{f}_M\}$. In essence, the frequency response of the least-square solution is the orthogonal projection of desired frequency response $\mathbf{H}_D$ onto the set of orthonormal basis frequency responses, i.e, the normalized DPSWFs $U_0(\omega), \cdots, U_{2M}(\omega)$ of the subspace $\mathcal{F}^2[-\omega_c, \omega_c]$; and the optimal filter is the corresponding linear combination of the corresponding sequences. This process is illustrated by the block diagram in Fig. 2.1. The DPSSs can be viewed as a basis and the coordinates serve as the weights to construct the optimal $\mathbf{h}_o$. Substituting (2.14) and (2.15) into the normal equation (2.17), we have

$$\mathbf{F}^H(\mathbf{H}_D(\omega) - \mathbf{F}\mathbf{h}_o) = \mathbf{F}^H\mathbf{r} = 0 \ , \tag{2.43}$$

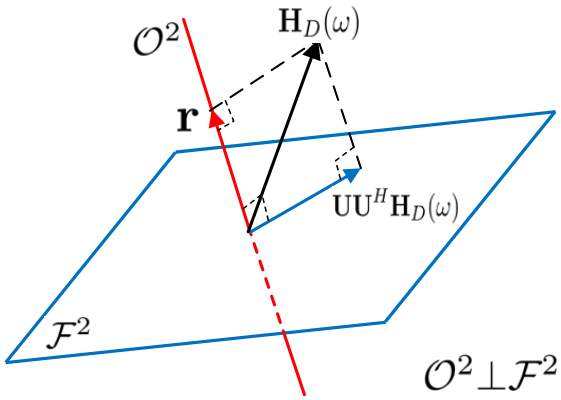

Fig. 2.2 Geometrical interpretation of optimal solution. $\mathcal{L}^2$ is the complete space of frequency responses, and $\mathcal{F}^2$ is a subspace of $\mathcal{L}^2$ spanned by the DPSWFs whose energy in the interval of $[-\omega_c, \omega_c]$ have been normalized.

where $\mathbf{r} = \mathbf{H}_D(\omega) - \mathbf{F}\mathbf{h}_o$ is defined as the residue error quasi-vector (or response) between the desired frequency response and the optimal response. This expression in (2.43) means that the residue error is orthogonal to the subspace $\mathcal{F}^2$, which is geometrically illustrated in Fig. 2.2. The error vector $\mathbf{r}$ is perpendicular to the hyperplane representing $\mathcal{F}^2$ and therefore the optimal frequency response has the shortest distance from the desired response among all candidates in $\mathcal{F}^2$.

2.5 Performance Comparison with Windowing Methods

Windowing methods construct the filter response $\mathbf{h}$ by multiplying the desired response with a frequency-domain window in the interval of $[-\omega_s/2, \omega_s/2]$, transforming back to a time domain sequence using IDTFT, and then applying a time-domain window in the interval of $[-M, M]$ to this sequence. In matrix notations, this is equivalent to multiplying three matrices to the desired response sequentially, a diagonal quasi-matrix, the IDTFT quasi-matrix, and a diagonal $(2M+1) \times (2M+1)$ matrix. Owing to this structural constraint, this approach cannot outperform the optimal method proposed in this paper, especially if the windows are chosen as being determined by only a few parameters.

For example, the simplest windowing method is to truncate the desired response using the rectangular frequency windows $[-\omega_c, \omega_c]$, transform back to a time domain sequence using IDTFT, and truncate this sequence with the rectangular time-domain window $[-M, M]$.

The resulting filter can be expressed as

$$\mathbf{h}_{\text{rec}} = \frac{1}{\omega_s}\mathbf{F}^H\mathbf{H}_D \; . \tag{2.44}$$

The frequency response of this filter is given by

$$\mathbf{H}_{\text{rec}}(e^{j\omega T_s}) = \frac{1}{\omega_s}\mathbf{F}\mathbf{F}^H\mathbf{H}_D = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H\mathbf{H}_D(\omega) \; , \tag{2.45}$$

which is a projection of $\mathbf{H}_D$, but not an orthogonal projection onto the subspace $\mathcal{F}^2$. Herein, the coordinates obtained from $\mathbf{U}^H\mathbf{H}_D$ are not immediately used to construct the optimal filter, but are scaled linearly. To quantify the difference between the IDTFT approach and the least square approach, we write their errors as follows:

$$E_{\text{rec}} = \frac{1}{2\omega_c}\|(\mathbf{I} - \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H)\mathbf{H}_D(\omega)\|_2^2 \; , \tag{2.46}$$

$$E_{\text{LS}} = \frac{1}{2\omega_c}\|(\mathbf{I} - \mathbf{U}\mathbf{U}^H)\mathbf{H}_D(\omega_c)\|_2^2 \; . \tag{2.47}$$

Note that $\mathbf{U}^H\mathbf{U} = \mathbf{I}$ but $\mathbf{U}\mathbf{U}^H \neq \mathbf{I}$. The difference between these two errors $\Delta E = E_{\text{rec}} - E_{\text{LS}}$ is given as follows:

$$\begin{aligned}\Delta E &= \frac{1}{2\omega_c}\mathbf{H}_D^H(\omega)\mathbf{U}(\mathbf{I} - \mathbf{\Lambda})^2\mathbf{U}^H\mathbf{H}_D(\omega) \\ &= \frac{1}{2\omega_c}\sum_{k=0}^{2M}|H_k'|^2(1-\lambda_k)^2 > 0 \; .\end{aligned} \tag{2.48}$$

Therefore, the optimal filter is always better than the IDTFT method. The relative error difference is highly dependent on the form of $\mathbf{H}_D^H(\omega)$, and it happens that for the desired dispersion and dispersion slop frequency response studied in this paper, E_{rec} is much larger than E_{LS}.

2.6 Summary

We have considered the optimal design of an FIR filter as the time-domain implementation for the dispersion and dispersion slope characteristics. The objective was to minimize the integral of the squared error between the FIR response and the desired response over

the band of interest, which can be computed based on adaptive integration techniques such as Gauss-Kronrod quadrature. This reduces the error floor to the order of 10^{-15} without adding computational complexity. This least square (LS) problem was solved in two approaches: the normal equation approach gives an explicit solution; the singular value decomposition (SVD) approach provides geometrical insights, based on the theory of DPSS. The introduction of DPSS also reveals several numerical problems that will be discussed in the next chapter.

Chapter 3

Numerical Issues and Modified Filters

The optimal filter has been derived based on the normal equation approach and the SVD approach; however, the theory of DPSS reveals that in certain instances, the normal equation could be ill-conditioned. Hereafter, several techniques are introduced to mitigate various numerical problems. We first add a regularization term to the objective function to provide robustness and introduce the fast MLD algorithm. Then, to suppress singular behaviors in the frequency domain such as overshoots, a maximum magnitude constraint is enforced on the frequency response, which can be formulated into a standard QCQP problem.

3.1 The Condition Number of A

The normal equation (2.17) is well-suited to solve the optimal filter response $\mathbf{h}_o$. Theoretically, this equation can be solved trivially by matrix inverse; computationally, however, there can be numerical problems. This system of linear equations is said to be ill-conditioned if a small perturbation in $\mathbf{A}$ and (or) $\mathbf{b}$ can lead to a large change in the solution $\mathbf{h}_o$. Since $\mathbf{b}$ is estimated using numerical integration, it cannot avoid the errors caused by the adaptive quadrature methods. Although $\mathbf{A}$ has an explicit expression, computing its entries is still subject to roundoff errors. These small errors could be amplified significantly in the solution if the problem is ill-conditioned.

This depends on the properties of $\mathbf{A}$, or more specifically, its condition number. which is defined as the ratio of the maximal to minimal singular value. In this special case, since

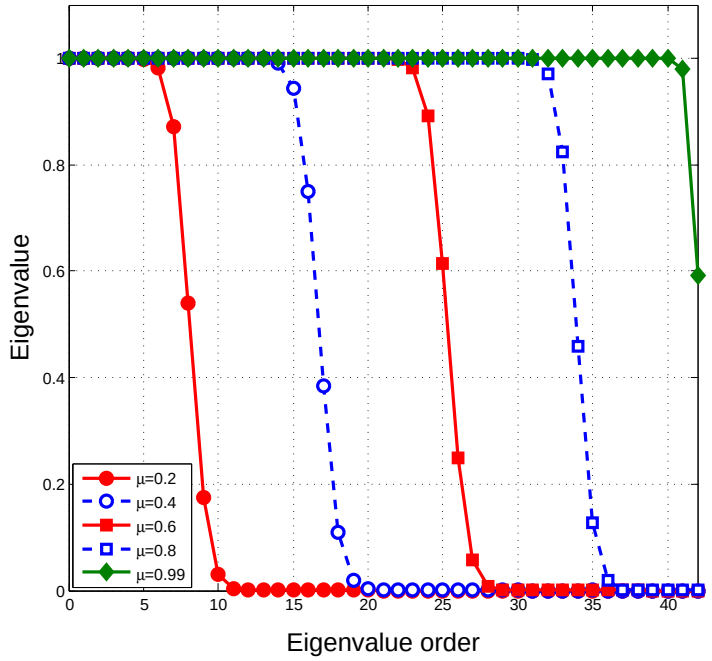

Fig. 3.1 The eigenvalues of $\mu\mathbf{A}$: λ_k, $k = 0, 1, \cdots, 42$. The time length is 65.625ps and effective bandwidth ratio $\mu = 0.2, 0.4, 0.6, 0.8, 0.99$.

$\mathbf{A}$ is Hermitian positive definite, its singular values are the same as its eigenvalues and its condition number is given by

$$\kappa(\mathbf{A}) = \frac{\lambda_{\max}(\mathbf{A})}{\lambda_{\min}(\mathbf{A})} = \frac{\lambda_0}{\lambda_{2M}}\ , \tag{3.1}$$

where μ has been canceled in this ratio. If $\kappa(\mathbf{A})$ is very large, the problem in (2.17) is ill-conditioned. Under this scenario, the small numerical error in $\mathbf{b}$ can be amplified by inversion of matrix $\mathbf{A}$, leading to large error in the solution.

The theory of DPSSs and DPSWFs states that λ_k are all distinct, real and positive, with some of them clustered near 1, and the others near 0, except very few of them takes intermediate values. This is illustrated in Fig. 3.1 for $2M + 1 = 43$. The results for different effective bandwidth ratios $\mu = 0.2, 0.4, 0.6, 0.8, 0.99$ are included. The number of eigenvalues near 1 is around $(2M + 1)\mu$. Except when μ is in close proximity with 1, the condition number tends to be significantly large, rending the problem ill-conditioned.

The contours of the logarithmic value of the condition number, i.e., $\varepsilon = \log_{10} \kappa(\mathbf{A})$,

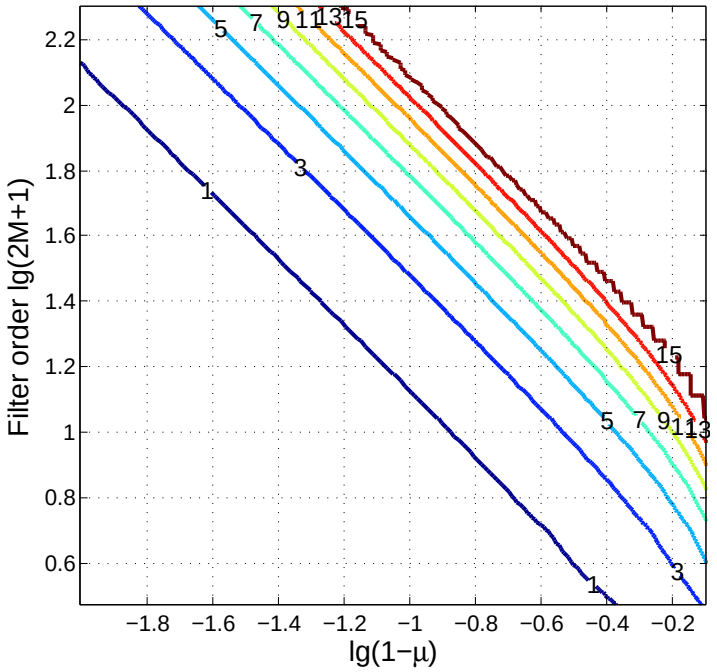

Fig. 3.2 Contours of the logarithmic value of the condition number, $\log_{10}\kappa(\mathbf{A})$, with filter order $2M+1$ and effective bandwidth ratio μ.

versus $\log_{10}(1-\mu)$ and $\log_{10}(2M+1)$, is shown in Fig. 3.2. The effective bandwidth ratio μ ranges from 0.2 to 0.99 and the filter order increases from 3 to 201. The logarithmic scale is used such that these contours are close to straight lines. In this context, the whole region is divided into two parts, namely, the well-conditioned region and ill-conditioned region. For instance, if a matrix whose condition number is higher than 10^{ε} is regarded as ill-conditioned, then all μ and $2M+1$ satisfying

$$(2M+1)(1-\mu) \geq 0.51\varepsilon + 0.72 \tag{3.2}$$

would approximately enclose an ill-conditioned region, i.e., the upper right triangular region. The borderline between these two regions is dependent on the numerical precision used: a double-precision machine can tolerate higher condition numbers than a single-precision machine, hence the ill-conditioned region would retreat towards the upper right corner. Unfortunately, our problem does not usually fall in the well-conditioned area and the algorithm needs to be modified in this scenario.

Next, we analyze the effect of an ill-conditioned $\mathbf{A}$ on the squared error. For simplicity, we assume that the exact $\mathbf{A}$ is used, and define the error of $\mathbf{b}$ as $\Delta\mathbf{b}$. The optimal solution is $\mathbf{h}_o' = \mathbf{A}^{-1}(\mathbf{b} + \Delta\mathbf{b})$ and the error function based on (2.5) is given by:

$$E_{LS}(\mathbf{h}_o') = 1 - \mathbf{b}^H\mathbf{A}^{-1}\mathbf{b} + \Delta\mathbf{b}^H\mathbf{A}^{-1}\Delta\mathbf{b} \; . \tag{3.3}$$

In this expression, the first two terms represent the error due to the accurate $\mathbf{h}_o$, and the last term is due to the error term $\Delta\mathbf{b}$. The eigenvalue decomposition of $\mathbf{A}$ leads to

$$\Delta\mathbf{b}^H\mathbf{A}^{-1}\Delta\mathbf{b} = \mu\Delta\mathbf{b}^H\mathbf{V}\mathbf{\Lambda}^{-1}\mathbf{V}^H\Delta\mathbf{b} = \sum_{k=0}^{2M}\frac{\mu|\mathbf{v}_k^H\Delta\mathbf{b}|^2}{\lambda_k} \; . \tag{3.4}$$

Herein, we model the error vector as a zero mean circularly symmetric complex Gaussian random vector with covariance matrix $\mathrm{E}\left\{\Delta\mathbf{b}\Delta\mathbf{b}^H\right\} = 10^{-\zeta}\mathbf{I}$. It is easy to verify that $\mathrm{E}\{|\mathbf{v}_k^H\Delta\mathbf{b}|^2\} = 10^{-\zeta}$. Because the eigenvalues can be close to 0, the summands could be large. For example, the error of the vector $\mathbf{b}$ obtained from IFFT or Riemann sums employing 10^6 points is typically on the order of 10^{-6}. In this case, a small eigenvalue, assuming 10^{-12}, could induce an error term of order 1. Furthermore, this error might be easily overlooked if one does not realize the inherent ill-conditioned property of $\mathbf{A}$ and uses (2.18) to compute the error instead of (2.5).

From the above analysis, we conclude that the ill-conditioned property of $\mathbf{A}$ is critical to this problem and we introduce a regularization-based approach to solve this issue.

3.2 Regularization

The problem that comes with the ill-conditioned $\mathbf{A}$ is that it can over-amplify the solution $\mathbf{A}^{-1}\mathbf{b}$. To mitigate this effect, we add a regularization term to the original LS error function:

$$E_{LS}' = E_{LS} + \nu\|\mathbf{h}\|^2 \; . \tag{3.5}$$

This term lowers the norm of $\mathbf{h}_o$ to some extent as controlled by the weighting parameter ν. Taking the derivative with respect to $\mathbf{h}^*$, the regularized optimal solution satisfies

$$(\mathbf{A} + \nu\mathbf{I})\mathbf{h}_o = \mathbf{b} \; . \tag{3.6}$$

The regularization coefficient is chosen such that the condition number is smaller than 10^{ε}, which is approximately equivalent to $\nu = 10^{-\varepsilon}$.

Next, we analyze the error of the solution $\mathbf{h}_o$ due to the numerical error $\Delta\mathbf{b}$. When using a perturbed $\mathbf{b} + \Delta\mathbf{b}$, the error expression becomes

$$\begin{aligned} E_{LS}(\mathbf{h}_o) = 1 &- \mathbf{b}^H(\mathbf{A}+\nu\mathbf{I})^{-2}(\mathbf{A}+2\nu\mathbf{I})\mathbf{b} \\ &- 2\Re\{\nu\Delta\mathbf{b}^H(\mathbf{A}+\nu\mathbf{I})^{-2}\mathbf{b}\} \\ &+ \Delta\mathbf{b}^H(\mathbf{A}+\nu\mathbf{I})^{-1}\mathbf{A}(\mathbf{A}+\nu\mathbf{I})^{-1}\Delta\mathbf{b}\ . \end{aligned} \tag{3.7}$$

where the first two terms are the error terms due to the limited filter order and the regularization parameter ν, and the last two terms represent the effects of $\Delta\mathbf{b}$. The third term is a zero-mean random variable whose variance satisfies

$$\begin{aligned} &\mathrm{E}\big\{\big(2\Re(\nu\Delta\mathbf{b}^H(\mathbf{A}+\nu\mathbf{I})^{-2}\mathbf{b}\big)^2\big\} \le \mathrm{E}\{|2\nu\Delta\mathbf{b}^H(\mathbf{A}+\nu\mathbf{I})^{-2}\mathbf{b}|^2\} \\ &\le 4\nu^2\|\mathbf{b}\|_2^2\|(\mathbf{A}+\nu\mathbf{I})^{-2}\|_2^2\mathrm{E}\{\|\Delta\mathbf{b}\|_2^2\} \\ &\le \frac{4\nu^2}{\mu}(\frac{\lambda_{\min}}{\mu}+\nu)^{-4}(2M+1)10^{-\zeta}\ , \end{aligned} \tag{3.8}$$

where we have used $\Re(x) \le |x|$, the Cauchy-Schwarz inequality $|\mathbf{x}^H\mathbf{y}|^2 \le \|\mathbf{x}\|_2^2\|\mathbf{y}\|_2^2$, $\|\mathbf{G}\mathbf{x}\|_2 \le \|\mathbf{G}\|_2\|\mathbf{x}\|_2$, and the Parseval's theorem $\|\mathbf{b}\|_2^2 \le 1/\mu$. In general, we let the regularization coefficient ν satisfy $\nu \gg \lambda_{\min}$, then the above expression is approximately equal to $4\nu^{-2}(2M+1)10^{-\zeta}/\mu$. The numerical error introduced by numerical integration techniques can be as low as 10^{-14} ($\zeta = 28$). In this case, if we let $\nu = 10^{-8}$, the variance of the third term in (3.7) is still upper bounded by a term on the order of 10^{-12}. In addition, the above inequalities will typically provide very loose upper bounds and even if ζ is smaller, the third term is still controllable. Similarly, the fourth term satisfies

$$\begin{aligned} &\mathrm{E}\{2\omega_c\Delta\mathbf{b}^H\mathbf{V}(\mathbf{\Sigma}^2+2\nu\omega_c\mathbf{I})^{-1}\mathbf{\Sigma}^2(\mathbf{\Sigma}^2+2\nu\omega_c\mathbf{I})^{-1}\mathbf{V}^H\Delta\mathbf{b}\} \\ &= \sum_{k=0}^{2M}\frac{\mathrm{E}\{|\mathbf{v}_k^H\Delta\mathbf{b}|^2\}}{\frac{\lambda_k}{\mu}+2\nu+\frac{\mu\nu^2}{\lambda_k}} \le \frac{1}{4\nu}\sum_{k=0}^{2M}\mathrm{E}\{|\mathbf{v}_k^H\Delta\mathbf{b}|^2\} \\ &= \frac{(2M+1)10^{-\zeta}}{4\nu}\ . \end{aligned} \tag{3.9}$$

The above inequality is not tight because the equality only holds for $\lambda_k = \nu$. If $\nu = 10^{-6}$,

$\zeta = 28$ and $2M + 1 = 101$, this second-order error can still be controllable (on the order of 10^{-20}). In summary, the effect of $\Delta\mathbf{b}$ has been efficiently suppressed by adding the regularization term and therefore will not be considered in later discussions.

3.3 Fast Implementation: Modified Levinson-Durbin (MLD)

So far, by introducing the regularization term, we have successfully turned the original least square problem into the problem of solving a well-conditioned system of linear equations, i.e., (3.6). Matrix $(\mathbf{A}+\nu\mathbf{I})$ is a symmetric positive matrix of dimension $(2M+1)\times(2M+1)$. More importantly, it is also a Toeplitz matrix where each descending diagonal from left to right is constant, that is, $()_{m,n}$ only depends on $m-n$. Henceforth, $(\mathbf{A}+\nu\mathbf{I})$ is uniquely determined by the $2M+1$ entries in the first column (or row). This nice Toeplitz structure enables a recursive approach of solving the linear equation in (3.6), which is significantly faster than Gaussian elimination.

The general Levinson-Durbin algorithm was proposed by N. Levinson in 1947, and improved by J. Durbin in 1960[38]. Its underlying principle is to start from solving a trivial one-dimensional matrix equation and to obtain the solution of a $(k+1)$-order Toeplitz matrix equation recursively based on the solution of a k-order system. Therefore, this algorithm can generate all the lower-order FIR filters when solving an n-dimensional equation, with a total computational complexity of $O(n^2)$.

Our model has additional structural characteristics: the filter order is constrained to be an odd number. Based on the same principle, we propose a modified version which explicitly uses these additional structures. This algorithm increases the filter order by two per update, one from above and the other from below, until it finds the lowest order needed to satisfy the requirement on the fitting error, with a total complexity of $O\big((2M+1)^2\big)$. In contrast, the searching process for Gaussian elimination requires a total complexity of $O((2M+1)^4)$, with the kth order alone consuming $O(k^3)$ operations [12]. The underlying principle of Levinson-Durbin algorithm is compute a higher-order Toeplitz problem recursively, based on the solution of a lower-order problem. It explicitly solves a set of problems from a trivial one-order one, until up to the original order. The normal equation in (3.6) has this nice Toeplitiz structure but $\mathbf{h}_o$ has only an odd number of entries, which necessitates our derivation of the following modified algorithm. We first introduce a few matrices and

vectors before we derive the algorithm. The permutation matrix is

$$\mathbf{P} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & \iddots & 0 \\ 1 & 0 & 0 \end{bmatrix} .$$

The $(2k+1)$-order reduced form of $\mathbf{A}$ is

$$\mathbf{A}_{2k} = \begin{bmatrix} r(0) & r(1) & \cdots & r(2k) \\ r(1) & \ddots & \ddots & r(2k-1) \\ \vdots & \ddots & \ddots & \vdots \\ r(2k) & \cdots & r(1) & r(0) \end{bmatrix} .$$

We also define the following column vectors:

$$\begin{aligned} \mathbf{r}_{2k} &= [r(1), r(2), \cdots, r(2k+1)]^T , \\ \mathbf{h}_{2k} &= [h_{-k}, h_{-k+1}, \cdots, h_{k-1}, h_k]^T , \\ \mathbf{b}_{2k} &= [b_{-k}, b_{-k+1}, \cdots, b_{k-1}, b_k]^T , \end{aligned}$$

and the Yule-Walker equation is

$$\mathbf{A}_{2k}\mathbf{y}_{2k} = -\mathbf{r}_{2k} \tag{3.10}$$

where $\mathbf{y}_{2k} \in \mathbb{C}^{(2k+1)\times 1}$ will plays an important role in this iterative updating algorithm.

Our goal is to solve the linear equation $\mathbf{A}_{2M}\mathbf{h}_{2M} = \mathbf{b}_{2M}$ recursively. The zero order equation $\mathbf{A}_0 h_0 = b_0$ is a scalar equation and can be trivially solved. Next, we show how to obtain the solution of $\mathbf{A}_{2k+2}\mathbf{h}_{2k+2} = \mathbf{b}_{2k+2}$ from that of a lower-order system $\mathbf{A}_{2k}\mathbf{h}_{2k} = \mathbf{b}_{2k}$ and that of its accompanied Yule-Walker equation in (3.10). In each update, the filter order increases from $2k+1$ to $2k+3$ and two coefficients b_{-k-1} and b_{k+1} are added to the right-hand side, one from above and the other from below. At the same time, $\mathbf{A}_{2k}$ is augmented in four directions to form $\mathbf{A}_{2k+1}$, upward, leftward, downward, and rightward.

Our derivation follows two steps. Firstly, we apply an one-order update to $\mathbf{A}_{2k}$ from the left and the above, leading to a new equation $\mathbf{A}_{2k+1}\mathbf{h}_{2k+1} = \mathbf{b}_{2k+1}$. It can be written

in a block form as

$$\begin{bmatrix} r(0) & \mathbf{r}_{2k}^T \\ \mathbf{r}_{2k} & \mathbf{A}_{2k} \end{bmatrix} \begin{bmatrix} \xi_1 \\ \mathbf{\Gamma}_1 \end{bmatrix} = \begin{bmatrix} b_{-k-1} \\ \mathbf{b}_{2k} \end{bmatrix} .$$

The second row, $\xi_1 \mathbf{r}_{2k} + \mathbf{A}_{2k}\mathbf{\Gamma}_1 = \mathbf{b}_{2k}$, leads to

$$\mathbf{\Gamma}_1 = \mathbf{A}_{2k}^{-1}(\mathbf{b}_{2k} - \xi_1 \mathbf{r}_{2k}) = \mathbf{h}_{2k} + \xi_1 \mathbf{y}_{2k} . \tag{3.11}$$

Here we have used $\mathbf{A}_{2k}\mathbf{y}_{2k} = -\mathbf{r}_{2k}$ and $\mathbf{A}_{2k}\mathbf{h}_{2k} = \mathbf{b}_{2k}$. Substituting back into the first row, we have

$$b_{-k-1} = r(0)\xi_1 + \mathbf{r}_{2k}^T\mathbf{\Gamma}_1 = \xi_1\big(r(0) + \mathbf{r}_{2k}^T\mathbf{y}_{2k}\big) + \mathbf{r}_{2k}^T\mathbf{h}_{2k} ,$$

which results in

$$\xi_1 = (b_{-k-1} - \mathbf{r}_{2k}^T\mathbf{h}_{2k})/(r(0) + \mathbf{r}_{2k}^T\mathbf{y}_{2k}) , \tag{3.12}$$

Therefore, this $(2k+2)$-order system of linear equations can be solved iteratively based on the solutions of the $(2k+1)$-order equations, $\mathbf{A}_{2k}\mathbf{y}_{2k} = -\mathbf{r}_{2k}$ and $\mathbf{A}_{2k}\mathbf{y}_{2k} = -\mathbf{r}_{2k}$.

After obtaining $\mathbf{h}_{2k+1}$, the second step is to solve the equation $\mathbf{A}_{2k+2}\mathbf{h}_{2k+2} = \mathbf{b}_{2k+2}$ using a similar block form

$$\begin{bmatrix} \mathbf{A}_{2k+1} & \mathbf{P}\mathbf{r}_{2k+1} \\ \mathbf{r}_{2k+1}^T\mathbf{P} & r(0) \end{bmatrix} \begin{bmatrix} \mathbf{\Gamma}_2 \\ \xi_2 \end{bmatrix} = \begin{bmatrix} \mathbf{b}_{2k+1} \\ b_{k+1} \end{bmatrix} .$$

Similarly, it follows that

$$\mathbf{\Gamma}_2 = \mathbf{h}_{2k+1} + \xi_2 \mathbf{P}\mathbf{y}_{2k+1} , \tag{3.13}$$

$$\xi_2 = (b_{k+1} - \mathbf{r}_{2k+1}^T\mathbf{P}\mathbf{h}_{2k+1})/(r(0) + \mathbf{r}_{2k+1}^T\mathbf{y}_{2k+1}) . \tag{3.14}$$

where we have used $\mathbf{A}_{2k+1}\mathbf{y}_{2k+1} = -\mathbf{r}_{2k+1}$, $\mathbf{A}_{2k+1}\mathbf{h}_{2k+1} = \mathbf{b}_{2k+1}$ and $\mathbf{A}_{2k+1}^{-1}\mathbf{P} = \mathbf{P}\mathbf{A}_{2k+1}^{-1}$ (Toeplitz).

Therefore, following the above two steps, the solution is updated using only matrix-vector multiplications and scalar operations, without computing inverse matrices and matrix products. Note that the Yuler-Walker equation can be solved using the standard Levinson-Durbin algorithm. The total complexity of this algorithm iterating from $k = 1$ to $k = M$, is only on the order of $O(M^2)$.

3.4 QCQP-based Optimal Design

Although the regularized least square approach fixes the ill-conditioned issue and benefits from the highly efficient MLD algorithm, it considers only the error in $[-\omega_c, \omega_c]$ without paying attention to its behaviors outside this frequency range. However, the overshooting phenomenon of the frequency response, if exists in the uninterested band, may deteriorate fiber simulations in an intricate manner. Despite the spectrum of the input signals, even WDM signals, cannot fully occupy the band $[-\omega_c, \omega_c]$, the nonlinearity operations can cause energy leakage outside this range, usually in a very small portion. This energy can be amplified by the overshoots of the frequency response and then moved back to the interested band by nonlinearity again. For short distance, this is not a severe issue; but when the simulated fiber link is hundreds or thousands of kilometers long, the linear FIR filter and nonlinear operation are repeated for hundreds or thousands of times, which would renders the simulation results unreliable.

Here, we analyze whether the least square solution without (or with) regularization is affected by this problem. The frequency response $H_o(e^{j\omega T_s})$ in (2.42) is a linear combination of basis responses, $U_k(\omega)$, each one of which has been normalized within $[-\omega_c, \omega_c]$. Based on the property of DPSWF, the portion of energy within the band of interest relative to the total energy is λ_k for $U_k(\omega)$. Therefore, the out-of-band energy of $U_k(\omega)$ for small λ_k is large, which would contribute to overshoots unless the coefficient $\mathbf{U}_k^H \mathbf{H}_D$ is negligibly small. In parallel, the impulse response of $\mathbf{h}_o$ in (2.41) is a linear combination of vectors, $\mathbf{v}_k/\sqrt{\omega_s \lambda_k}$. If λ_k is very small, the kth term would contribute much to a very large $\|\mathbf{h}_o\|_2$ unless the corresponding h'_k is also very small. Based on Parseval's theorem, this large energy can only be squeezed into the uninterested band because the spectrum is already well-fitted in the band of interest. Admittedly, the regularization term can mitigate the overshooting effect to some extent if a higher order filter is employed, but sacrificing more filter order is far from a satisfactory solution because this would instead increase computational complexity of time-domain convolution.

To suppress overshoot efficiently without over-sacrificing filter order, we impose additional constraints on the amplitudes of the frequency response. The constrained least

square problem with regularization is given by

$$\text{minimize} \quad 1 - \mathbf{b}^H\mathbf{h} - \mathbf{h}^H\mathbf{b} + \mathbf{h}^H\mathbf{A}\mathbf{h} + \nu\|\mathbf{h}\|^2 \,, \tag{3.15a}$$

$$\text{subject to} \quad |H(e^{j\omega})| \leq 1 + \varepsilon, \quad \omega_c \leq |\omega| \leq \omega_s/2 \,, \quad . \tag{3.15b}$$

where $\varepsilon \geq 0$ is a very small number such as 10^{-5}. To simplify this problem a little bit, we replace the constraints imposed in the continuous frequency intervals with m constraints at m discrete frequency points in $[-\omega_s/2, -\omega_c]$ and $[\omega_c, \omega_s/2]$. This problem can be formulated into a standard *quadratically constrained quadratic programing* (QCQP) problem [39], that is,

$$\text{minimize} \quad \mathbf{h}^H\mathbf{P}_0\mathbf{h} + 2\Re(\mathbf{q}_0^H\mathbf{h}) + \mathbf{r}_0 \,, \tag{3.16a}$$

$$\text{subject to} \quad \mathbf{h}^H\mathbf{P}_k\mathbf{h} + 2\Re(\mathbf{q}_k^H\mathbf{h}) + \mathbf{r}_k \leq 0 \,, \quad \forall\, k = 1, \cdots, m \,. \tag{3.16b}$$

The parametric matrices and vectors are defined as

$$\mathbf{P}_0 = \mathbf{A} + \nu\mathbf{I} \,, \tag{3.17a}$$

$$\mathbf{q}_0 = -\mathbf{b} \,, \tag{3.17b}$$

$$\mathbf{r}_0 = 1 \,, \tag{3.17c}$$

$$\mathbf{P}_k = \mathbf{a}(\omega_k)\mathbf{a}^H(\omega_k) \,, \tag{3.17d}$$

$$\mathbf{q}_k = 0 \,, \tag{3.17e}$$

$$\mathbf{r}_k = -(1+\varepsilon)^2, \forall\, k = 1, \cdots, m \,. \tag{3.17f}$$

Since $\mathbf{P}_k$ is positive semidefinite for all k, this optimization problem is convex and can be readily solved using state-of-the-art interior point methods[26].

The frequency responses of the unconstrained regularized optimal filter in (3.6) and the QCQP-based filter are compared in Fig. 3.3. The single mode fiber (SMF) with $D = 17\text{ps}/(\text{nm}\cdot\text{km})$ and $S = 0.08\text{ps}/(\text{nm}^2\cdot\text{km})$ is used for simulation. The split-step method takes a step length of $\Delta z = 2\text{km}$. The sampling frequency is $\omega_s = 3.75 \times 10^{12}$ rad/s and the effective bandwidth ratio is taken as $\mu = 0.8$. The regularization parameter is set to $\nu = 10^{-6}$. One can observe that all these three filters match the desired response in the band

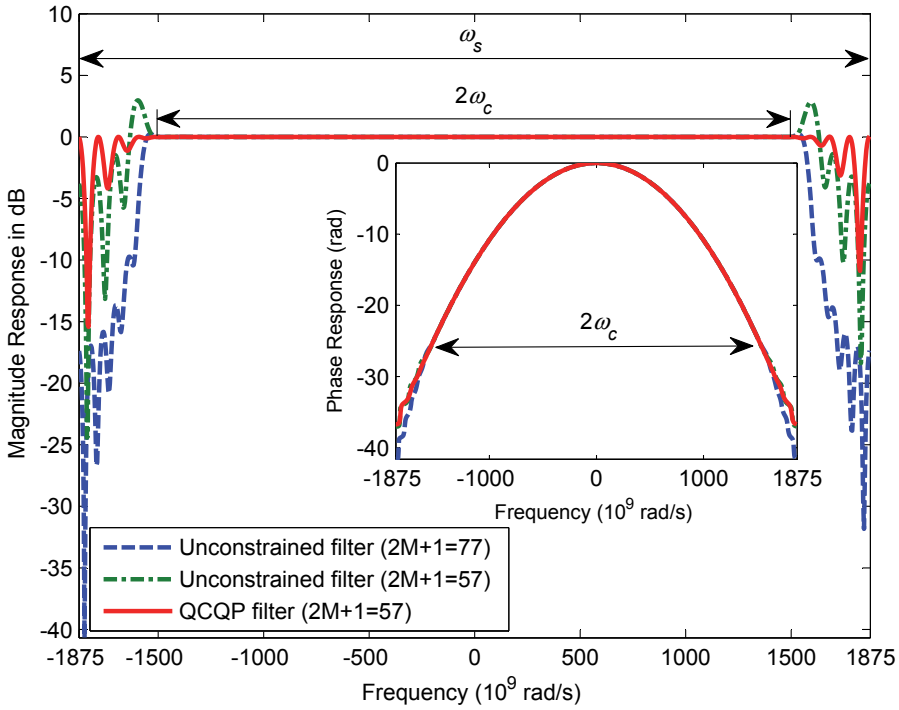

Fig. 3.3 Comparison of frequency responses for three different filters: unconstrained optimization with filter order $2M+1=77$, unconstrained optimization with $2M+1=57$ and QCQP-based filter with $2M+1=57$.

of interest, with their squared error in $[-\omega_c, \omega_c]$ satisfying $E_{LS} < 10^{-6}$. However, the 57-order unconstrained filter produces a 2.5dB overshoot in the outer band. By increasing the number of order to 77, this overshoot is sufficient suppressed. In contrast, the QCQP-based optimal filter successfully controls the overshooting behavior. In summary, both regularized LS and QCQP-based filter "squeeze" the strong ripples out of $[-\omega_c, \omega_c]$ by sacrificing the uninterested band: the regularized LS filter does not control singular behaviors, whereas QCQP-based filter reshapes the frequency response outside $[-\omega_c, \omega_c]$ by "pressing" its head below the sea level of 0dB. The reason for this "partial-band" formulation is that if $\mu \approx 1$, the squared error can be reduced by increasing the filter order, but the ripples due to Gibbs effect cannot be suppressed. In fact, when $\mu = 1$, the column vectors of $\mathbf{F}$ are already orthogonal, therefore its SVD satisfies $\mathbf{V} = \mathbf{I}$, $\mathbf{\Sigma} = \sqrt{\omega_s}\mathbf{I}$. In this scenario, the LS filter is actually the IDTFT of the desired response.

Note that there is a trade-off in complexity between the unconstrained regularized LS and the QCQP formulation if overshoot control is required. Thanks to the MLD algorithm,

the former filter is easier to extract but has a larger filter order; the latter filter is shorter but harder to extract. Therefore, the former design is more appropriate when the linear effects are changing constantly so that the filter needs to be regenerated every few steps, such as in the time-domain implementation of both chromatic dispersion and polarization mode dispersion (PMD)[40]. In addition, this also enables numerical simulations which use variable step size[5]. The latter design is more suitable if the filter does not change and only need to be computed once or infrequently.

To see the influence of overshot in numerical accuracy, we employ three different filters in time-domain approach. SMF described above is used for simulation. The propagation distance is 450km, thus involved 450 time-domain convolutions. The results for three different filters, 57-order unconstrained filter, 77-order unconstrained filter and 57-order QCQP-based filter are presented in Fig. 3.4. In accordance with previous analysis, because the strong overshot existing in 57-order unconstrained filter, the corresponding simulation results are unreliable: the stable pulse shape becomes several rambling points. Adding the filter order up to 77 and using 57-order QCQP-based filter can both overcome this shortcoming and obtain the stable pulse shape, which is consistent with the results from SSFM(FFT).

Note that there is a trade-off in complexity between the unconstrained regularized LS and the QCQP formulation if overshoot control is required. Thanks to the MLD algorithm, the former filter is easier to extract but has a larger filter order; the latter filter is shorter but harder to extract. Therefore, the former design is more appropriate when the linear effects are changing constantly so that the filter needs to be frequenntly regenerated every step, such as in the time-domain implementation of both chromatic dispersion and polarization mode dispersion (PMD)[40]. In addition, this also enables variable step size, which is preferable in numerical simulations[5]. The latter design is more suitable if the filter does not change and only need to be computed once.

3.5 The Order of Optimal Filter

In a nutshell, the purpose of designing the FIR filter is to fit the desired frequency response to a given precision, with a small filter order. This can significantly reduce the computational overhead in computing the convolution of the input sequence and the FIR filter. Hereafter, we investigate the relationship between the required filter order and system

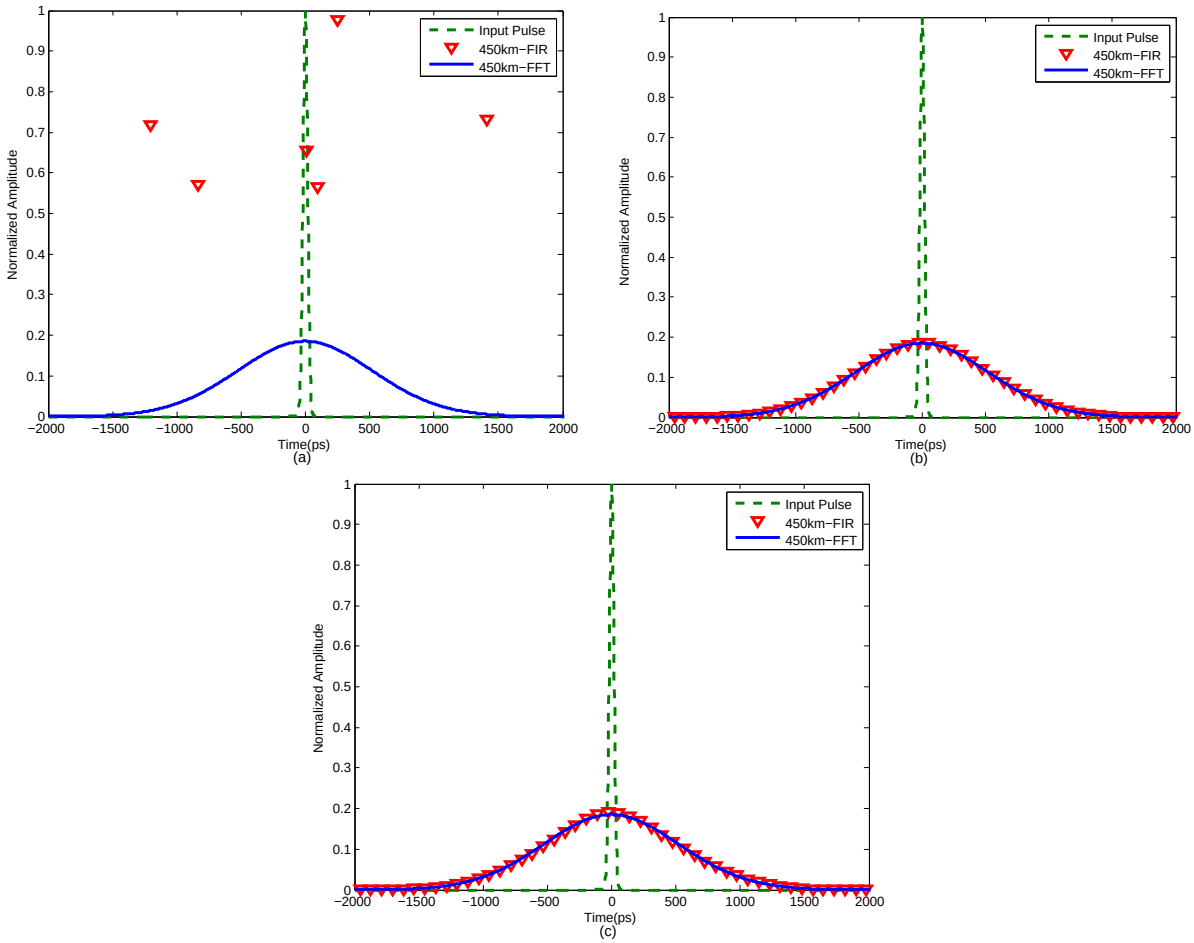

Fig. 3.4 Time-domain split-step simulation of Gaussian pulse propagation based on: (a) 57-order unconstrained filter; (b) 77-order unconstrained filter; (c) 57-order QCQP-based filter.

parameters both theoretically and numerically.

3.5.1 Theoretical Analysis Based on Group Delay

Since the optimal filter $\mathbf{h}_o = (\mathbf{A} + \nu\mathbf{I})^{-1}\mathbf{b}$ is a linear transformation of the vector $\mathbf{b}$, the ability of $\mathbf{h}_o$ to resemble the desired response $H_D(j\omega)$ in the band of interest $[-\omega_c, \omega_c]$ is limited by how much information is preserved in the $(2M+1)$-dimensional vector $\mathbf{b}$. A careful examination shows that b_k is actually equal to the IDTFT of a windowed version

of the desired frequency response, except by a scale factor of $1/\mu$. That is to say,

$$b_k = \frac{1}{\mu}\,\frac{1}{\omega_s}\int_{-\frac{\omega_s}{2}}^{\frac{\omega_s}{2}} H_D(j\omega)W(j\omega)e^{jk\omega T_s}\mathrm{d}\omega\ , \tag{3.18}$$

for all $k \in \mathbb{Z}$, where the frequency-domain window is defined as

$$W(j\omega) = \begin{cases} 1, & \text{if } |\omega| \le \omega_c\ ; \\ 0, & \text{if } |\omega_c| < |\omega| \le \omega_s/2\ . \end{cases} \tag{3.19}$$

Henceforth, $\mathbf{b} = [b_{-M}, \cdots, b_M]^T$ is actually a truncated version of the time-domain sequence corresponding to the desired response truncated by the frequency-domain rectangular window $W(j\omega)$. Some information in the infinitely long sequence $\{b_k\}$ is lost due to this truncation, and the smaller the filter order, the more information loss is incurred.

Therefore, the required filter order is closely related to the effective time duration of the sequence $\{b_k\}$, which can be viewed as the range of indexes for which the value of $|b_k|$ is not trivially small. In general, the time-bandwidth product of a simple sequence, such as a rectangular frequency window, is a constant or on the same order, say 1. As the bandwidth becomes smaller, the pulse sequence becomes wider. However, this is not true for the truncated response

$$H_D(j\omega)W(j\omega) = \begin{cases} \exp\left[j\left(\frac{\beta_2\omega^2}{2} - \frac{\beta_3\omega^3}{6}\right)\frac{\Delta z}{2}\right] & \text{if } |\omega| \le \omega_c, \\ 0 & \text{otherwise.} \end{cases}$$

The phase response is a nonlinear function of the frequency ω and this nonlinear modulation implies that the time-bandwidth product might not be a constant. Indeed, as we show later, when the bandwidth ω_c increases, the sequence $\{b_k\}$ does not becomes narrower, but wider instead.

The concept of *group delay* plays the most important role in the following analysis, which is defined as

$$\tau_g = -\frac{\mathrm{d}}{\mathrm{d}\omega}\angle H_D(j\omega) = -\frac{\beta_2\omega\Delta z}{2} + \frac{\beta_3\omega^2}{4}\Delta z\ . \tag{3.20}$$

For simplicity, we considers the most practical scenario when $|\beta_2| \gg |\beta_3|$ and the second

term in (3.20) is dominated by the first term[3]. We hence drop the second term and the extension of the resulting conclusions to other cases is straightforward. The lower bound for the required filter order can be established by the following arguments on $H_D(j\omega)$:

1) We extend the desired response to the whole frequency range, that is,

$$H_{\text{ext}}(j\omega) = \exp\left(j\frac{\beta_2\omega^2\Delta z}{4}\right) . \tag{3.21}$$

This is the Fourier transform of the following continuous-time unit impulse response[41]

$$h_{\text{ext}}(t) = \sqrt{\frac{j}{\pi\beta_2\Delta z}}\exp\left(-\frac{jt^2}{\beta_2\Delta z}\right) . \tag{3.22}$$

The group delay at the frequency ω_0 is

$$\tau_g(\omega_0) = -\frac{\omega_0\beta_2\Delta z}{2} . \tag{3.23}$$

The instantaneous frequency of $h_{\text{ext}}(t)$ at $t = \tau_g(\omega_0)$, defined as the derivative of the phase with respect to time t, is equal to

$$\omega_0' = \frac{\mathrm{d}\angle h_{\text{ext}}(t)}{\mathrm{d}t}|_{t=\tau_g(\omega_0)} = -\frac{2\tau_g(\omega_0)}{\beta_2\Delta z} = \omega_0 . \tag{3.24}$$

The instantaneous frequency at the group delay of ω_0 is ω_0 itself, which means that there is a one-to-one correspondence between the time domain and the frequency domain.

2) Although the desired response is merely a truncated version of the extended response, it still inherits this one-to-one correspondence, though in an imperfect manner. In essence, $\{b_k\}$ is obtained by sampling the convolution between the sinc pulse corresponding to the rectangular frequency window (3.19), and the extended continuous-time response $h_{\text{ext}}(t)$ in (3.22). This implies that the frequency component of ω_0 is somewhat expanded around its group delay, according to the pattern defined by the sinc pulse. Normally, we hope to fit the desired response within a wide bandwidth $2\omega_c$, and thereby the sinc pulse is pretty narrow. Therefore, the range of group delay for frequencies between $-\omega_c$ and ω_c approximately spans almost all the nontrivial samples of $\{b_k\}$. From (3.23), with ω ranging from $-\omega_c$ to ω_c, τ_g is constrained in the time interval of $[-|\beta_2|\omega_c\Delta z/2, -|\beta_2|\omega_c\Delta z/2]$. With a sampling

period of T_s, to preserve most energy of $\{b_k\}$, the filter order $2M+1$ have to satisfy

$$2M+1 \geq \frac{|\beta_2|\omega_c \Delta z}{T_s} = \frac{|\beta_2|\omega_c^2 \Delta z}{\pi\mu} = \frac{4\phi_{\max}}{\pi\mu} = L_{\min} \ , \tag{3.25}$$

where we define $\phi_{\max} = |\beta_2|\omega_c^2 \Delta z/4$ as the maximum phase shift due to second-order dispersion.

3) The lower bound for the filter order, $L_{\min}$, is proportional to the maximum phase shift $\phi_{\max}$, and inversely proportional to μ. The dependence of $L_{\min}$ on the four parameters, namely, ω_c, ω_s, β_2 and Δz, boils down to a linear relationship involving only two independent quantities. On the one hand, $\phi_{\max}$ is related to the number of oscillations experienced by the real and imaginary parts of the desired response; on the other hand, μ represents the degree to which this oscillatory response is expanded in the fundamental period of the spectrum $[-\omega_s/2, \omega_s/2]$.

3.5.2 Numerical Experiments

In the following, we study the required filter order for a given error tolerance numerically. Since the theoretical lower bound $L_{\min}$ was derived based on an approximate time-frequency mapping, the required filter order would be different from but highly related to $L_{\min}$. In fact, it is approximately a linear function of the theoretical lower bound. Since $L_{\min}$ depends on both the maximum phase shift $\phi_{\max}$ and the effective bandwidth ratio μ, we fix one of them and change the other one in the following numerical experiments.

Firstly, we fix $\mu = 0.75$ and then study the relation between the required filter order and $\phi_{\max}$. There are two ways to adjust $\phi_{\max}$, by altering either Δz or ω_c. Here, we first let $\omega_c = 1.5 \times 10^{12}$ rad/s and change Δz between 0.2km and 16.2km. Independently, we let $\Delta z = 1.8$km and change ω_c between $\omega_c = 0.5 \times 10^{12}$ rad/s and $\omega_c = 4.5 \times 10^{12}$ rad/s. The simulation is also based on the same SMF used in Fig. 3.3. The minimum filter order required to achieve the error level of 10^{-4}, 10^{-6} or 10^{-8} is plotted versus $\phi_{\max}$ in Fig. 3.5, respectively. We can observe that when μ is fixed, the filter order required for a specific fitting error is uniquely determined by the single parameter $\phi_{\max}$. Once it is fixed, the filter order would not vary even if ω_s, ω_c, Δz may change, which is consistent to our previous analysis.[1] More importantly, for a specified error tolerance, the required filter

[1] In fact, even if β_2 changes, the filter order does not change as long as $\phi_{\max}$ and μ are fixed.

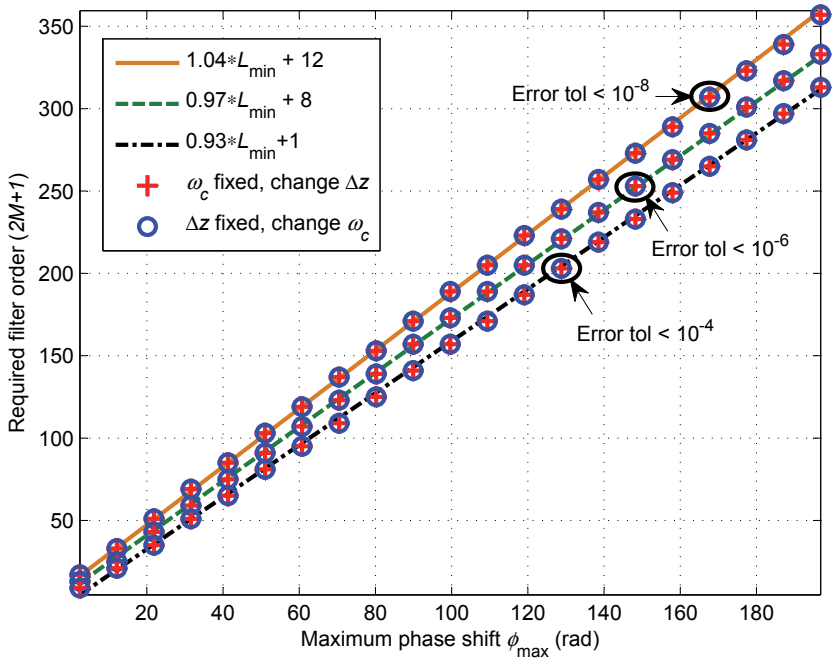

Fig. 3.5 The required minimum filter order versus maximum phase shift $\phi_{\max}$ for different error tolerances. $\phi_{\max}$ can be changed by either adjusting ω_c or Δz (with the other one fixed). Straight lines are used to fit the numerical results. Parameter settings: $\mu = 0.75$, $D = 17\text{ps/(nm·km)}$ and $S = 0$.

order $(2M+1)$ can be fitted by a corresponding linear function of $\phi_{\max}$ and thus also $L_{\min}$. Because our derivation of (3.25) takes a conceptual approach, $(2M+1)$ is not equal to $L_{\min}$, and the slope of the straight lines are not necessarily one and the y-intercepts are not necessarily zero. Secondly, we fix $\phi_{\max} = 21.87\text{rads}$ and investigate the relation between $2M+1$ and the effective bandwidth ratio μ for the same SMF. As illustrated by Fig. 3.6, $(2M+1)$ increases almost linearly with $1/\mu$, which can also be fitted by linear functions of $L_{\min}$. We emphasize that $\beta_3 = 0$ in the above simulations and even if $\beta_3 \neq 0$, our numerical experiments verified that the required filter order hardly changes because the second-order dispersion usually dominates over third-order terms[3].

So far, we have verified that the minimum filter order for a given error tolerance is independently determined by $\phi_{\max}$ and μ. Their linear relationship as seen from numerical experiments is consistent with the conclusion in (3.25) drawn from our theoretical analysis. These expressions have both theoretical and practical meanings. To begin with, they provide theoretical insights into the desired response and its time-domain counterpart by

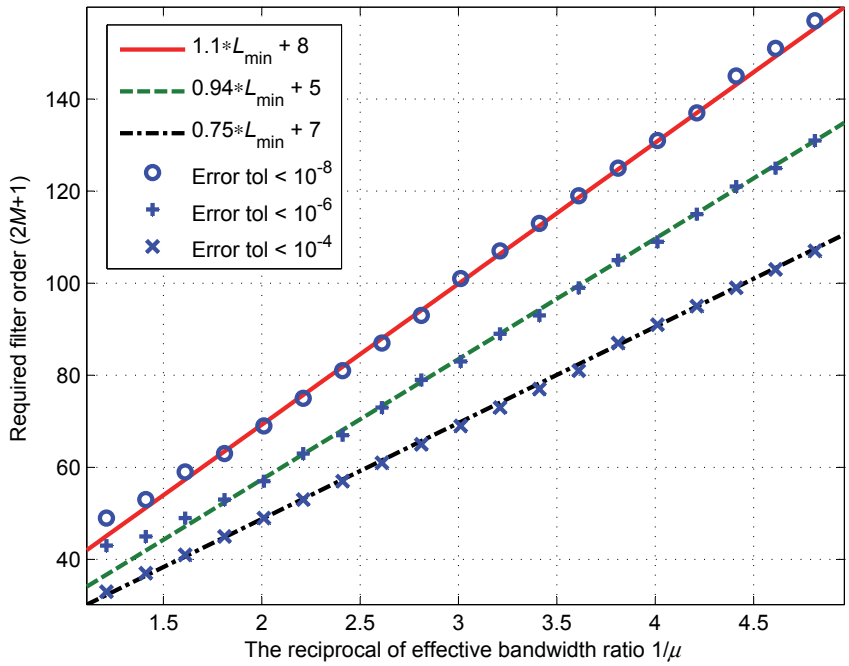

Fig. 3.6 The required minimum filter order versus the reciprocal of effective bandwidth $1/\mu$ for different error tolerances. Here, μ is changed by either adjusting ω_s, or ω_c (but with $\phi_{\max}$ fixed). Straight lines are used to fit the numerical results. Parameter settings: $\mu = 0.2 \sim 0.9$, $D = 17$ps/(nm·km), $S = 0$, $\omega_c = 1.5 \times 10^{12}$ rad/s, $\Delta z = 1.8$ km.

establishing a one-to-one correspondence between group delay and instantaneous frequency. Another important role is that they provide an initial guess of the minimum filter order needed for a given error tolerance. Henceforth, one may start searching from this initial filter order instead of all the way from order 1. Lastly but not the least, since the filter order is almost proportional to the step size Δz (via $\phi_{\max}$), the computational complexity of the convolution increases linearly as Δz increases. At the same time, the number of steps decreases in an inversely proportional manner. Consequently, the overall complexity to implement the direct convolutions does not change considerably. This saves the task of choosing step size because it only has to satisfy the requirement due to the "split-step" itself.

To better understand the error behaviors, we plot the squared error versus $2M + 1$ in Fig. 3.7. Three curves are for the LS solutions with the regularization parameter $\nu = 10^{-4}$, 10^{-6}, 10^{-8}, respectively. The fourth curve is for the LS filter without regularization and the

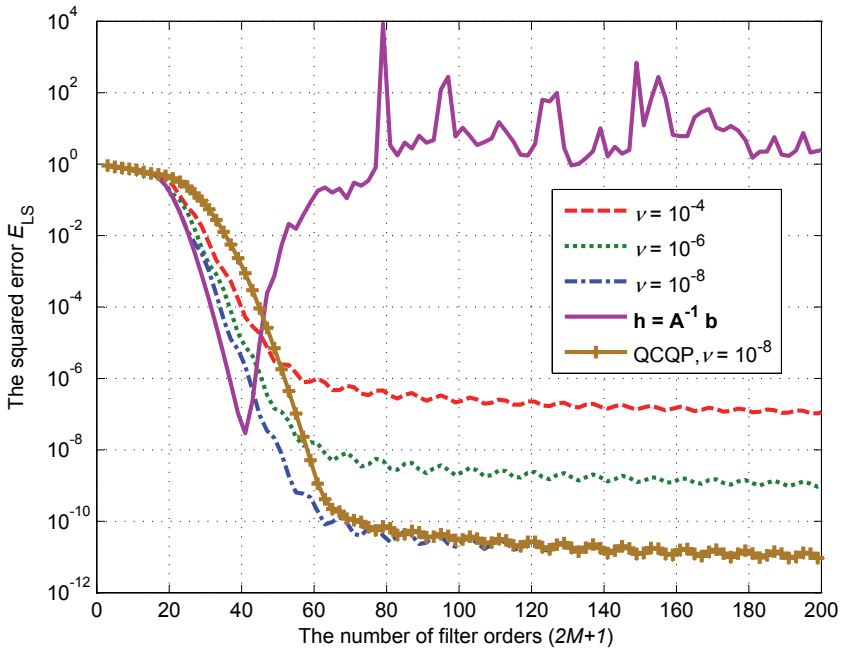

Fig. 3.7 The error behaviors of different filters as the filter order increases. The parameter used for simulation are: $D = 17\text{ps/(nm·km)}$, $S = 0.08\text{ps/(nm}^2\text{·km)}$, $\omega_c = 1.5 \times 10^{12}$ rad/s; and Δz is fixed as 1.8 km.

fifth curve is for the QCQP-based filter with $\nu = 10^{-8}$. The following parameters are used: $D = 17\text{ps/(nm·km)}$, $S = 0.08\text{ps/(nm}^2 \cdot \text{km)}$, $\omega_c = 1.5 \times 10^{12}$ rad/s, and $\Delta z = 1.8\text{km}$. The following conclusions can be made: firstly, the original LS solution without regularization, i.e., $\mathbf{A}\mathbf{h}_o = \mathbf{b}$, suffers appreciably from the ill-conditioned $\mathbf{A}$. After the filter order exceeds a threshold, the error becomes uncontrollable. Secondly, this ill-conditioned problem is successfully mitigated by regularization; however, this introduces an error floor to the resulting solution and the larger ν is, the higher this error floor rises up. Thirdly, the QCQP-based filter wastes some degrees of freedom in suppressing overshoots, and hence requires 4 to 8 orders more than the regularized LS filter to achieve the same error performance. In addition, the oscillatory behaviors of these curves can be explained by the oscillatory property of the desired response. Lastly, the fitting error due to small perturbations in $\mathbf{b}$ which comes from the Gaussian quadrature algorithm is not included in this figure because we have no knowledge of the exact $\mathbf{b}$. As analyzed before, this error becomes more unpredictable when ν approaches zero. Therefore, there is a fundamental trade-off in selecting ν to balance these two errors. In general, ν from 10^{-6} to 10^{-10} is a good candidate.

The filter orders of the proposed design are also compared with those in recent works. The following parameters are used: $2\omega_c = 3 \times 10^{12}$rad/s and $\omega_s = 4 \times 10^{12}$rad/s, $\beta_2 = -21.6\text{ps}^2/\text{km}$, $\beta_3 = 0.117\text{ps}^3/\text{km}$. For the step size of $[0.5, 0.7, 1, 2, 3, 4]$ (km), the corresponding smallest filter orders to satisfy $E_{LS} \leq 10^{-6}$ are $2M + 1 = [23, 27, 35, 57, 79, 101]$ based on the QCQP approach. This reduces the filter order by 1/3 to 1/2 when compared with the recent work in [13] ($2M + 1 = [47, 57, 69, 83, 111, 147]$). Because the optimization approach imposes no structural constraint, it is capable of exploiting more degrees of freedom.

3.6 Summary

Based on the theory of DPSS, we have revealed that in certain instances, the normal equation could be ill-conditioned. Henceforth, the solution might be sensitive to numerical errors and could also generate overshoots outside the band of interest, which would be amplified and transformed back into the band of interest by the nonlinear operations. This will generate unreliable results after propagating long distances. If the problem is ill-conditioned, we added a regularization term to the objective function to provide robustness. The resulting filter can also suppress overshoots by increasing its length; however, we improved it by imposing a maximum magnitude constraint on the frequency response to control overshoots more efficiently. The resulting quadratically constrained quadratic programming (QCQP) problem can be readily solved by state-of-the-art interior-point methods.

For a given error tolerance, we established the relationship between the required filter order and several parameters both theoretically and numerically. Based on the one-to-one correspondence between group delay and instantaneous frequency, we derived a tight lower bound of the filter order as a linear function of the step size, whose validity is also verified by numerical experiments. This can simplify the task of choosing the step size from the perspective of reducing computational complexity.

The proposed optimal filters reduce the total computational complexity, both when extracting the filter and implementing linear convolutions. On the one hand, the unconstrained regularized LS filter is the solution of a Toeplitz system. This enables a fast modified Levinson-Durbin algorithm with the complexity of $O(n^2)$. The QCQP-based filter can be computed with efficient interior-point methods. On the other hand, the computational complexity of linear convolutions depends exclusively on the filter length. Numerical

simulations show that the QCQP-based filter saves at least 1/3 of the total filter order when compared with most recent work. Moreover, there is a complexity trade-off between the unconstrained regularized LS filter and QCQP-based filter if the overshoot control is required: the former is easier to extract but the latter is shorter. The choice depends on whether the filter needs to be regenerated frequently or not.

Chapter 4

Time-Domain Simulation of Pulse Propagation in Optical Fiber

The proposed filters can easily fit the desired response with a square error lower than 10^{-8}. It remains unverified whether this filter outputs a signal similar to that produced by the standard SSFM, which is the main purpose of this chapter. Besides, we also quantify the computation complexity of the time-domain approaches and compare it with previous methods. To reduce both flops and memory usage, we suggest an overlap-type scheme that further improves the computational efficiency.

4.1 Choosing the Step Size

To begin with, we discuss the issue of choosing the step size for numerical simulations of signal propagation in optical fiber, which is in essence a comprise between accuracy and complexity. Different methods for choosing step size have been proposed. *Nonlinear Phase-Rotation Method* chooses the step size so as to keep the phase rotation caused by nonlinearities within a pre-given level. This method is effective for soliton systems rather than WDM systems. *Logarithmic Step-size Distribution* chooses the step size according to a logarithm distribution in order to suppress the spurious FWM[42]. *Walk-Off Method* automatically adjusts step sizes so that they are inversely proportional to the largest group velocity difference between channels. *Local-error Method* selects the step size by evaluating the relative local error of each single step, thus the error level is bounded and under control[5]. *Constant Step Method* keeps the step size constant along the transmission. The

smaller the step size is, the more accurate the simulation results become. However, a small step size will introduce high computational complexity. For simplicity, we confine our discussions in this chapter to the scenario of constant step size.

For single-channel simulations, the constant step size is determined by dispersion length and nonlinear length. Assuming the width of a pulse is T_0 and its power is P, the second-order dispersion length, third-order dispersion length and nonlinear length are defined as [3]

$$L_{D2} = \frac{T_0^2}{|\beta_2|} , \tag{4.1}$$

$$L_{D3} = \frac{T_0^3}{|\beta_3|} , \tag{4.2}$$

$$L_{nl} = \frac{1}{\gamma P} . \tag{4.3}$$

To maintain sufficient accuracy, the step size should be chosen to be smaller than any of the above three values. For WDM simulations, the step size should be smaller than nonlinear length, walk-off length and FWM length that are defined as follows [29],

$$L_{NL} = \frac{1}{\gamma P_T \frac{2N-1}{N}} , \tag{4.4}$$

$$L_{WO} = \frac{1}{2\pi(N-1)|\beta_2|\Delta f R} , \tag{4.5}$$

$$L_{FWM} = \frac{1}{\pi^2|\beta_2|(N-1)^2\Delta f} , \tag{4.6}$$

where P_T is the total transmission power, N is the number of channels, Δf is the channel spacing and R is the symbol rate.

The computational complexity of a time-domain split-step method depends on the length of the FIR filter and the total number of steps. For our proposed filters, the filter length is a linear function of the step size. As the step size increases, the filter length increases but the total number of steps decreases. Henceforth, the computational complexity does not change very much and then the step size can be simply chosen to be smaller than above three lengths.

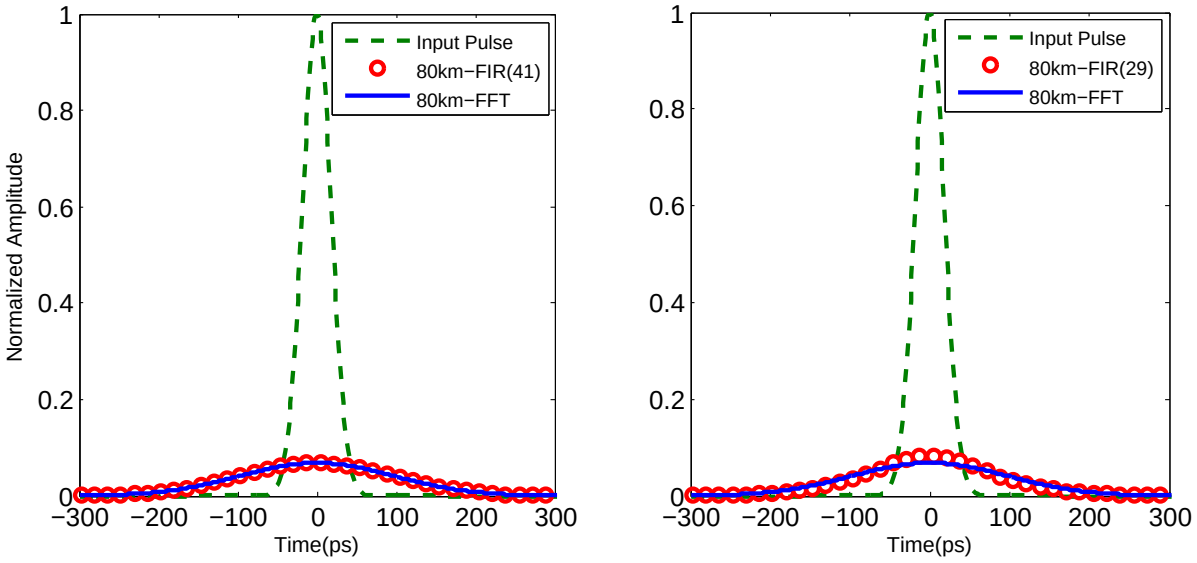

Fig. 4.1 The impact of the squared error on the pulse propagation.

4.2 Numerical Validations

4.2.1 The Impact of the Squared Error

In order to illustrate the impact of the squared error on pulse propagation, we compare the output signals of the time-domain split-step methods based on two QCQP-based filters with different lengths, 29 and 41. The squared error for these two filters is on the level of 10^{-5} and 10^{-8}, respectively. The step size Δz is chosen as 1km. The frequency responses of these filters match the desired dispersion characteristics within a guaranteed bandwidth of $2\omega_c = 3 \times 10^{12}$rad/s with the effective bandwidth ratio $\mu = 0.8$. An input Gaussian pulse with FWHM 30ps and peak power 1mW is passed through a single channel (at 1550nm). The simulation is based on SMF with the following parameters: $D = 17$ps/(nm·km), $S = 0.08$ps/(nm^2·km), $\alpha = 0.2$dB/km and $\gamma = 2\text{W}^{-1}$/km.

Fig. 4.1 compares the output signals generated by the time-domain split-step methods based on these two FIR filters. The 41-order filter generates a output pulse whose shape is almost the same as that produced by SSFM, whereas the 29-order filter leads to noticeable difference as seen in the right-sided plot. Therefore, higher squared error between the FIR response and the desired response can induce higher level of output mismatch. This is even worsened as the propagation distance increases. Henceforth, a squared error low enough is

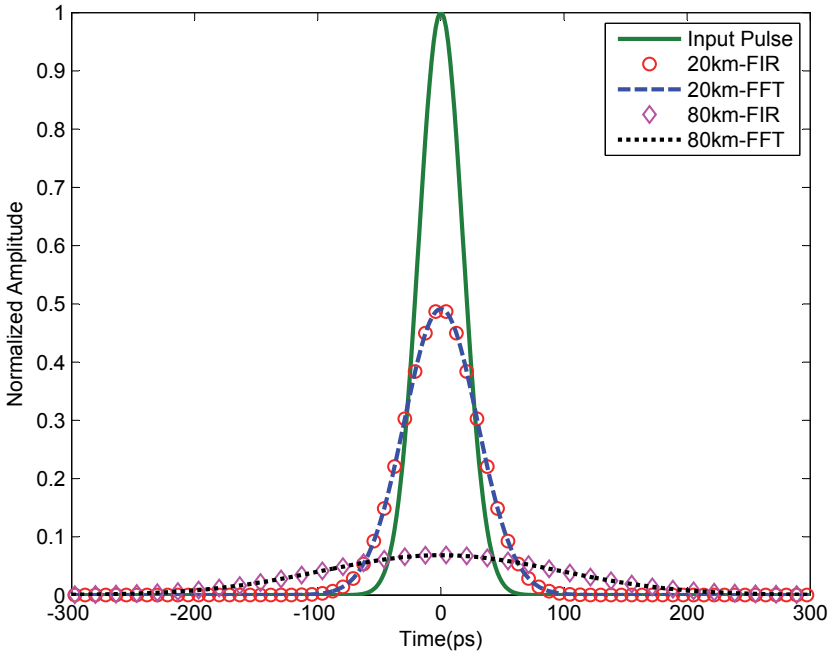

Fig. 4.2 The output signal of SMF with the Gaussian pulse of peak power 1mW as input (with both second-order and third-order dispersion).

indispensable to guarantee the reliability of the simulation results.

4.2.2 Single Channel

We pass an input Gaussian pulse through a single mode fiber (SMF) and compare the output pulses generated by split-step methods based on the FFT and the proposed QCQP-based filter. The parameters are set as follows: the input pulse has a full width at half maximum (FWHM) of 30ps, and a peak power of 1mW; we use a standard SMF (1550nm) with $D = 17\text{ps/(nm·km)}$, $S = 0.08\text{ps/(nm}^2\text{·km)}$, $\alpha = 0.2\text{dB/km}$ and $\gamma = 2\text{W}^{-1}\text{/km}$; the step size $\Delta z = 1\text{km}$, $\omega_c = 1.5 \times 10^{12}\text{rad/s}$, and the optimal FIR filter has an order of 47 with a square error of $E_{LS} \leq 10^{-8}$; the effective bandwidth $\mu = 0.8$ and the regularization parameter $\nu = 10^{-6}$. The comparisons between the SSFM (implemented using FFT) and the time-domain approach based on the proposed FIR filter are shown in Fig. 4.2. The pulse shapes are almost identical for both methods after propagating in the fiber of length either 20km or 80km.

In order to observe stronger nonlinear effects, we increase the power to 100mW for

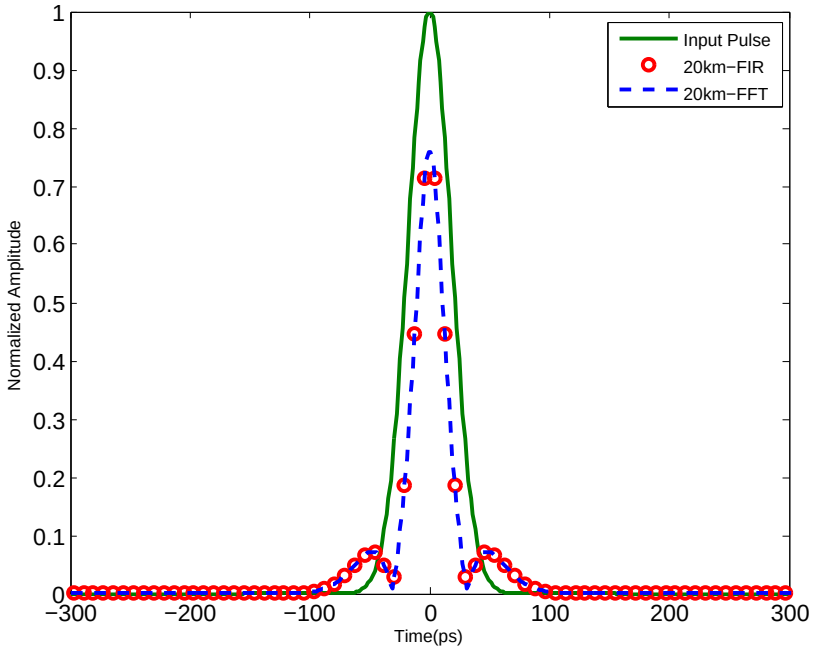

Fig. 4.3 The output signal of SMF with the Gaussian pulse as input (with both second-order and third-order dispersion) with peak power 100mW.

the input pulse and pass it through the same SMF. The output pulses after 20km of propagation, generated by the SSFM and the time-domain approach, are illustrated in Fig. 4.3. In this scenario of high input power, the pulse is not only broadened in time, but also experiences changes in its shape. This is because when the input power increases to 100mW, nonlinearities become the dominating effects during the pulse propagation. The output signals generated by the SSFM and the time-domain method are consistent with each other. In summary, the input power in fiber-optic communication systems cannot be too high so as to reduce the signal shape distortion during the transmission; it cannot be too low, either, to ensure a high enough SNR at the receiver after one span of propagation (usually 80km).

Although second-order dispersion usually dominates over third-order dispersion in most cases of practical interests, the third-order dispersion plays an important role at the zero-dispersion wavelength. Hence, we also verify the case when the pulse propagates at zero-dispersion wavelength and thereby only the third-order dispersion is considered. The parameters are set as follows: we choose two different fibers with $S = 0.08\text{ps}/(\text{nm}^2\cdot\text{km})$ and $S = -0.08\text{ps}/(\text{nm}^2\cdot\text{km})$; the step size is chosen as $\Delta z = 200\text{km}$; the filter order is still 47

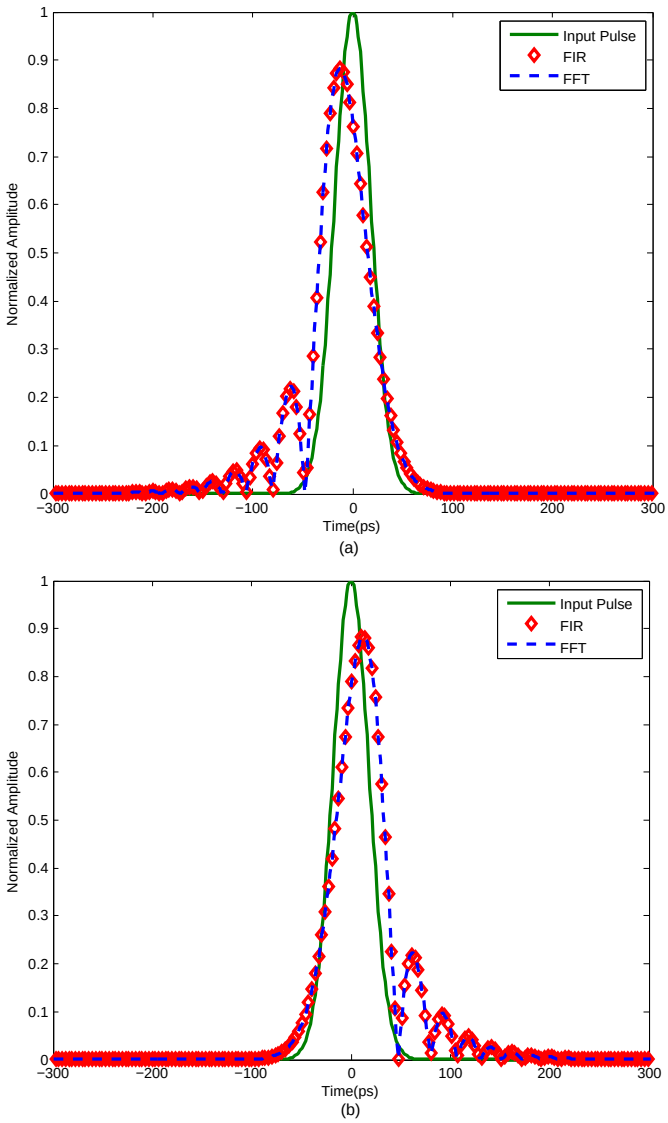

Fig. 4.4 The output signal of fiber after propagating 100000km based on SSFM and split-step FIR filtering approach: (a) with negative third-order dispersion only and (b) with positive third-order dispersion only.

with $E_{LS} \leq 10^{-8}$ over the bandwidth of interest; other parameters remain the same. As observed in Fig. 4.4, the results based on the SSFM and the time-domain approach are still almost identical, which verifies the validity of the proposed filters for third-order dispersion. Moreover, third-order dispersion causes asymmetrical distortion of the signals, i.e., oscillations appear at the leading edge of the pulse when S is negative and at the trailing

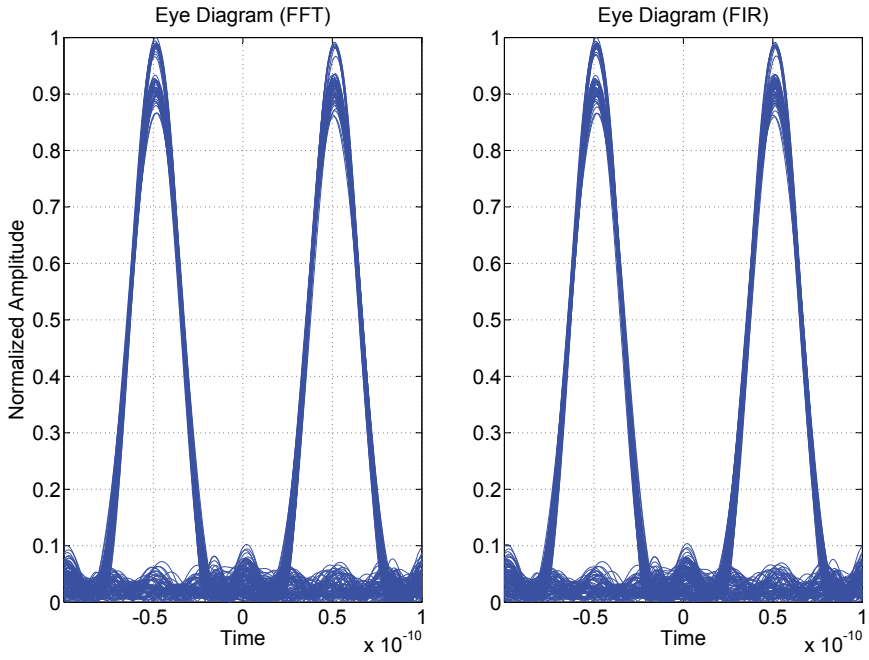

Fig. 4.5 Eye-diagrams of the output signals after propagating 1500km.

edge of the pulse when S is positive. The oscillations are deep and the pulse amplitude approaches zero between successive oscillations.

Up to here, single-channel experiments have shown that the time-domain split-step methods based on our proposed filters can be used in practice. In the next section, we will extend our discussion to WDM channels.

4.2.3 WDM Systems

When simulating a WDM system, different channels can be treated together as a single electrical field that incorporates all the nonlinear effects. Let N be the total number of channels and the envelop of the input signal at the mth channel be A_m, $m = 1, 2, \cdots, N$. Thus, the total field can be represented as $A = \sum_{m=1}^{N} A_m \exp(jm\Delta\omega t)$, where $\Delta\omega$ is the channel spacing in rad/s. This total field is viewed as the electrical field at the input of the fiber channel and its propagation in the fiber channel is still described by the NLSE. This is called "total-field" simulation of a WDM system. Although a variety of other algorithms have also been proposed, we restrict our discussion to the most popular "total-field" approach.

Table 4.1 Specifications of a 16 × 10Gb/s WDM System

Parameters for the WDM system		
Channel numbers	16	
Reference wavelength	1550nm	
Bit rate	10Gbps	
Chanel spacing	50GHz	
Pulse shape	Gaussian with FWMH 20ps	
Sampling frequency	8.0425×10^{12}rad/s	
Number of bit	2^7	
Number of span	15	
Fiber parameters	SMF	DCF
Second-order dispersion (D)	17ps/(nm·km)	−68ps/(nm·km)
Third-order dispersion (S)	0.08ps/(nm^2·km)	−0.08ps/(nm^2·km)
Loss	0.2dB/km	0.6dB/km
Nonlinear coefficient	2 W^{-1}·km^{-1}	2 W^{-1}·km^{-1}
Length per span	80km	20km

The WDM system under simulation is summarized in Table 4.1. The input signal is an OOK-modulated Gaussian pulse train. The system consists of 15 transmission spans and each span has two stages. The first stage is an 80km standard SMF with the same parameters as those in the single-channel simulation. The second stage is a 20km dispersion-compensation fiber (DCF). After each stage, the signal is amplified by EDFA with a gain of $G = 14$dB to compensate for the transmission loss. For simplicity, the effects of amplified spontaneous emission noise are not taken into consideration. The reference wavelength is 1550nm, the operating wavelength ranges from 1547.2nm to 1553.2nm, and the wavelength spacing is 0.4nm. In the simulation, we set $\omega_s = 8.0425\times10^{12}$rad/s, $\mu = 0.8$, and $\nu = 10^{-10}$. The step size is chosen as 0.2km. The extracted QCQP-based FIR filters for SMF and DCF links are respectively of length 55 and 119, with their errors satisfying $E_{LS} \leq 10^{-12}$. For the input power of 1mW, Fig. 4.5 shows the eye-diagrams of the output signals. Again, the outputs from the SSFM and the time-domain method are almost the same. As long as the squared error is low enough, the accuracy of split-step time-domain method will maintain

after thousands of kilometers transmission. It is worth mentioning that the unconstrained regularized LS filter with a higher filter order also generates similar outputs as verified by simulations not shown here.

4.3 Computational Complexity

Computational complexity of different split-step methods are compared in this subsection. For simplicity, we only consider the maximum order and do not differentiate terms like $3n^3$ and $5n^3$. Detailed analysis with the coefficients depends on specific algorithms and hardware architecture, which varies on a case-by-case basis. Before analyzing the computational complexity, we first introduce the overlap-add and overlap-save method that can compute the linear convolution more efficiently[43]. The introduction of overlap-type convolution techniques further reduces the overall complexity of the time-domain split-step approaches.

4.3.1 Overlap-Add and Overlap-Save Method

Overlap-add and overlap-save are efficient ways to calculate the linear convolution between a long signal sequence $A(z,n)$ and a short FIR filter $h_D(n)$. For the signal of length P and FIR filter of length M', the circular convolution and linear convolution are identical if the length of circular convolution is $P+M'-1$. Therefore, if we augment both the signal and the FIR filter to length $P+M'-1$ with zero samples, the linear convolution can be computed using DFT. If the P is very large, using DFT for a large number of points is computationally expensive.

Block convolution can be used to solve this problem. It segments the long input sequence into small sections, and then each section is filtered by the FIR filter, and finally the filtered sections are combined together in an appropriate way. The filtering processes are computed based on DFT. For the overlap-add method, the signal sequence $A(z,n)$ is segmented into N sections of length L

$$A(z,n) = \sum_{i=0}^{N-1} A_i(z, n-iL) \ , \tag{4.7}$$

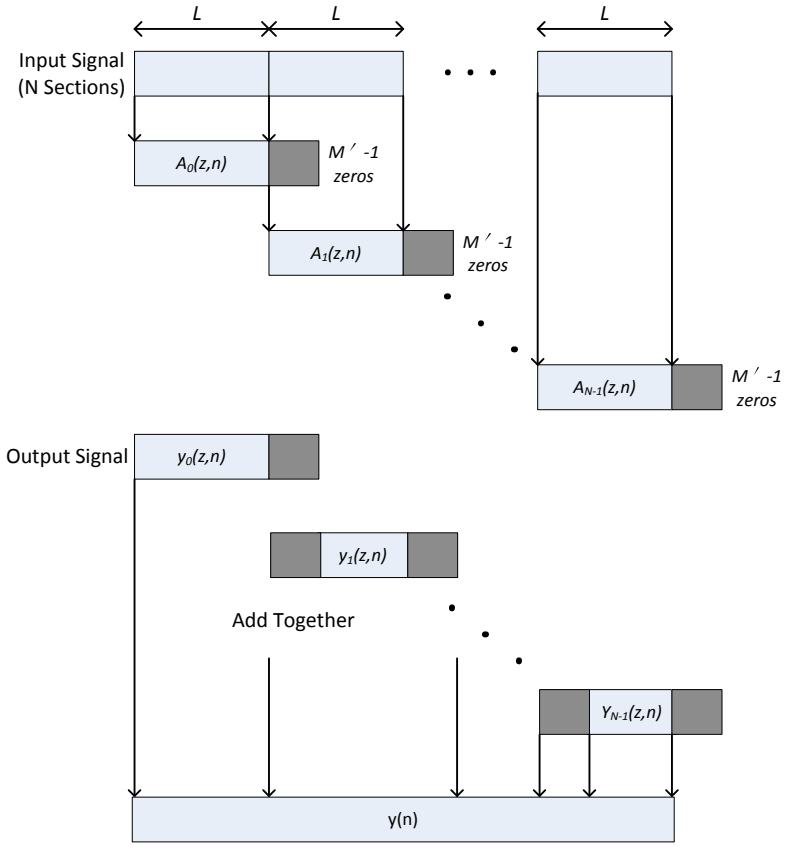

Fig. 4.6 Overlap-add method.

where

$$A_i(z,n) = \begin{cases} A(z, n+iL), & \text{if } 0 \le n \le L-1, \\ 0, & \text{otherwise} \end{cases}$$

Fig. 4.6 shows the diagram of the overlap-add method. The filtered output signal is given by

$$y(n) = \sum_{i=0}^{N-1} y_i(z, n-iL) \ , \tag{4.8}$$

where

$$y_i(z,n) = A_i(z,n) \otimes h_D(n) \ , \tag{4.9}$$

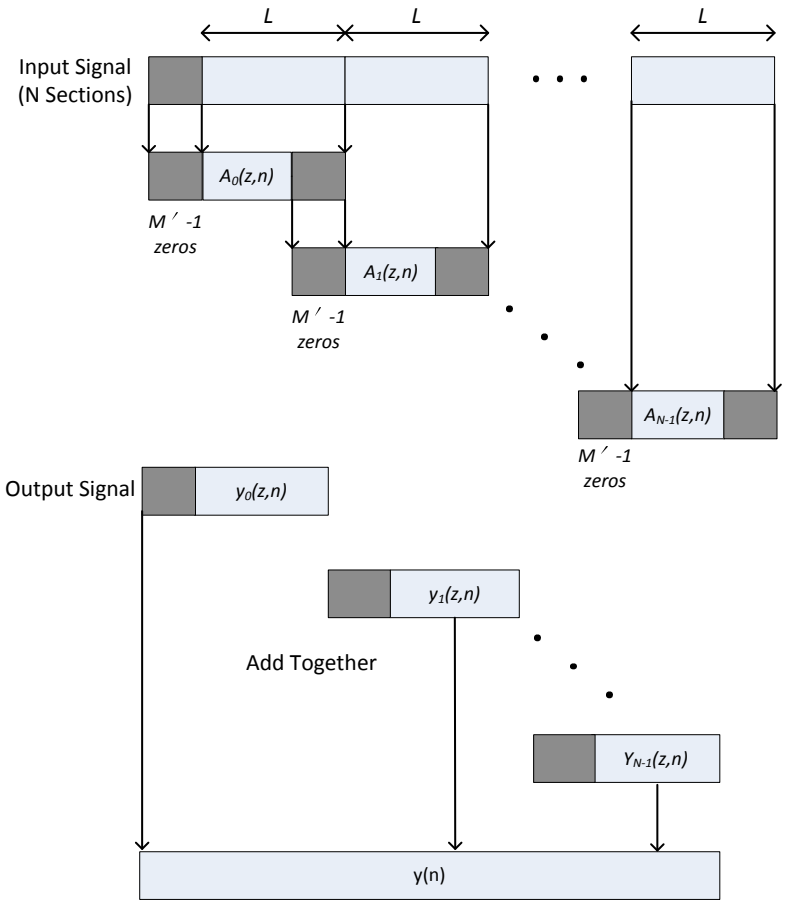

Fig. 4.7 Overlap-save method.

By zero padding the sequences $A_i(z,n)$ and $h_D(n)$ to be length K, where $K \geq P + M' - 1$, the linear convolution is equivalent to circular convolution. Thus the linear convolution can be computed using K-point DFT's. According to the filtering processing described in Fig. 4.6, the filtered sections will overlap by $M' - 1$ points, thus are combined appropriately in the final step.

Another efficient convolution procedure is overlap-save method. The input signal is divided still into N sections. Each section of length L and FIR filter of length M' are convoluted circularly. In the resulting sequences, the first $M' - 1$ are incorrect while the remaining L points are the same as those obtained from linear convolution. We define the

Table 4.2 Comparison of Computation Complexity

Frequency-domain	$O(P \log P)$
Time-domain: direct convolution	$O(PM')$
Time-domain: overlap-add	$O(P \log M')$

sections as

$$A_i(z,n) = A(z, n + i(L - M' + 1) - M' + 1), 0 \le n \le L - 1 \ .$$

Then after filtering based on L-point DFT's, each segment $y_i(z,n)$ are combined to realize the final output

$$y_i(n) = \sum_{i=0}^{N-1} y_{io}(z, n - i(L - M' + 1) + M' - 1) \ ,$$

where

$$y_{io}(z,n) = \begin{cases} y_i(z,n), & \text{if } M' - 1 \le n \le L - 1, \\ 0, & \text{otherwise} \end{cases}$$

This process is shown in Fig. 4.7.

In overlap-add and overlap-save methods, DFT is computed using FFT. Therefore, the linear convolution can be more efficiently implemented.

4.3.2 Linear Convolution with Low Complexity

The major difference between different split-step methods is how the linear parts are implemented. Herein, we consider both the frequency domain approach using FFT and the time-domain approach based on FIR filter. Specifically, for the latter approach, the convolution of a long input sequence and the FIR filter can either be computed directly, or obtained using FFT-based block convolution techniques such as overlap-add and overlap-save. These overlap-based techniques are different from SSFM and the block processing in [6] even though they all use FFT. The overlap-type methods divide the input signal into small blocks, compute the convolution of each block and the FIR filter using FFT, and then combine all blocks together. The standard SSFM computes FFT for the input sequence itself which is significantly longer. The block processing techniques mentioned in [6] also divide the input signal into small blocks, but then apply SSFM to each block before

recombining these blocks, and therefore still suffer from the time-aliasing problem.

The parameters are set as follows: the number of split steps is R, the length of the input signal sequence is P, and the filter order is $M' = 2M+1$. In typical optical fiber simulations, especially for WDM systems, $P \gg M'R$. The frequency-domain approach first applies a P-point FFT to the input sequence which requires $O(P \log P)$ operations. Multiplication in the frequency domain needs $O(P)$ operations (negligible) and the complexity of the IFFT is also $O(P \log P)$. The total complexity is still on the order of $O(P \log P)$. For the time-domain FIR filter approach, direct convolution requires $O(PM')$ flops. If overlap-add is used, we assume that the length of each block is M' and hence the total number of blocks is P/M'. Each FFT or IFFT requires approximately $(2M' - 1)\log(2M' - 1)$ operations so that the circular convolution is equal to the linear convolution. The cost of adding the samples from neighboring blocks is at most $O(P)$. It is easy to check that the total number of operations is on the order of $O(P \log M')$. As summarized in Table 4.2, the FIR filter approach with direct convolution is more efficient than SSFM only if $M' < \log P$, whereas overlap-based convolution prove to be much more efficient than SSFM: $O(P \log M')$ versus $O(P \log P)$. At the same time, all these time-domain methods require less memory than SSFM because they avoid large-point FFT's and IFFT's.

4.4 Summary

The single channel and wavelength-division multiplexing (WDM) simulations verified that the output signals generated by the proposed regularized LS filter and QCQP-based filter are almost the same as those by SSFM, even after propagating thousands of kilometers. In addition, we also introduced the overlap-add method that can reduce the computational complexity of the linear convolutions from $O(PM')$ to $O(P(\log M'))$, where P is the length of input signal and M' is the filter order.

Chapter 5

Time-Domain Backpropagation for Fiber Impairment Compensation

In this chapter, we will apply our proposed filter to the inverse process of pulse propagation, namely, digital backpropagation for fiber impairment compensation. We will first introduce the theory of digital backpropagation, then apply the algorithm based on the proposed filter to long-haul transmission systems. The simulation results and discussions are followed thereafter.

5.1 Theory of Digital Backpropagation

To achieve high transmission capacity, fiber impairment compensation plays an important role as dispersion and nonlinear related effects causes the degradation of performance [44, 45]. Recently, electrical dispersion compensation and electrical nonlinear compensation have gained great attentions, due to the maturity of coherent receiver which simultaneously enables the conversion of optical signals into electrical signals and preservation of the amplitude and phase information [46, 47, 48]. Electrical compensation and coherent receiver not only increase the fiber capacity limits significantly, but also are more reliable and cost-effective than previous all-optical signal processing [49]. Electrical dispersion compensation has been already well studied, however, nonlinear compensation remains a big challenge. The ultimate solution comes from digital backward propagation (backpropagation), which is the inverse process of pulse forward propagation in optical fiber. It takes a unified approach of treating linear and nonlinear effects, and thus electrical compensations for

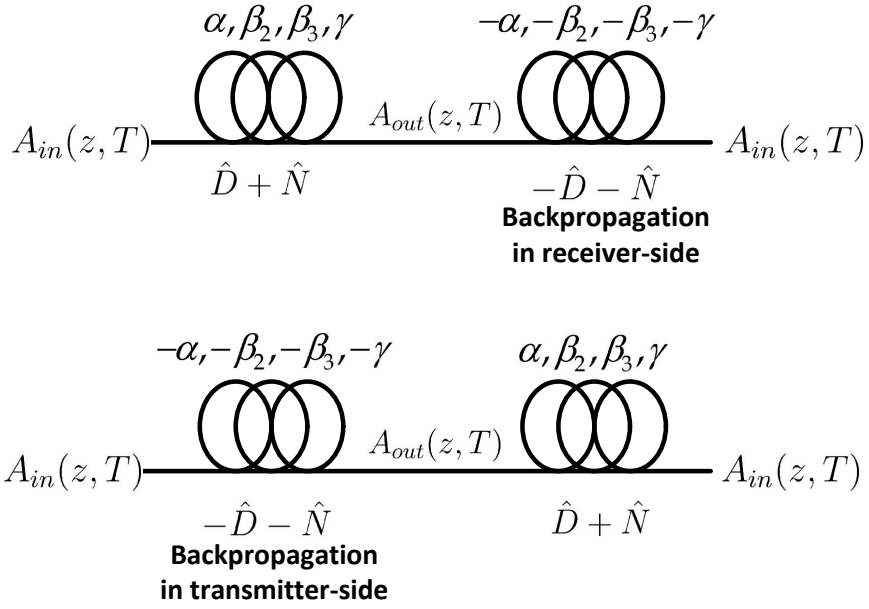

Fig. 5.1 Backpropagation implementation at the receiver-side and transmitter-side.

dispersions and nonlinearities are simultaneously realized.

In the absence of noise, the backpropagation compensation scheme recovers the transmitted signal from the received signal by virtue of inverse NLSE

$$\frac{\partial A(z,T)}{\partial z} = (-\hat{D} - \hat{N})A(z,T)\,, \tag{5.1}$$

where $\hat{D}$ and $\hat{N}$ are defined in (1.7). Comparing this equation with (1.6), it is evident that the idea of backpropagation is the operation of channel inverse. The channel inversion can be implemented at the transmitter side or the receiver side, or both, as illustrated in Fig. 5.1. Because NLSE is invertible, in the absence of noise, backpropagation successively compensates the distortion on the output signals caused by the fiber channel. However, implementing backpropagation at the transmitter-side is usually undesirable since it requires channel feedback.

Various numerical algorithms for solving the NLSE numerical such as S-SSM and A-SSM can be employed in digital backpropagation. As before, we restrict our discussion to S-SSM due to its high accuracy. Each step relies on one back-and-forth Fourier transform

and is given by

$$\hat{A}(z+\Delta z/2,\omega) = H_{-D}(\omega)\mathscr{F}[A(z,T)]\exp(\alpha\Delta z/4)\ , \tag{5.2}$$

$$\hat{A}_1(z+\Delta z/2,T) = \mathscr{F}^{-1}[\hat{A}(z+\Delta z/2,\omega)]\ , \tag{5.3}$$

$$\begin{aligned}\hat{A}_2(z+\Delta z/2,T) = {} & \exp(-j\Delta z\gamma|\hat{A}_1(z+\Delta z/2,T)|^2) \\ & \times \hat{A}_1(z+\Delta z/2,T)\ ,\end{aligned} \tag{5.4}$$

$$A(z+\Delta z,\omega) = H_{-D}(\omega)\mathscr{F}[\hat{A}_2(z+\Delta z/2,T)]\exp(\alpha\Delta z/4)\ , \tag{5.5}$$

where $H_{-D}(\omega)$ is the frequency response including dispersion related effects in fiber but with opposite parameters, read,

$$\begin{aligned}H_{-D}(\omega) = \exp\left[j\left(-\frac{\beta_2\omega^2}{2}+\frac{\beta_3\omega^3}{6}\right)\frac{\Delta z}{2}\right]\ , \\ -\omega_s/2 \le \omega < \omega_s/2\ .\end{aligned} \tag{5.6}$$

Different from fiber simulation which computes how the input signal propagates forward in the optical fiber and then predicts the signal at the output, digital backpropagation is used in real-time high data-rate systems to compensate for the distortions caused by dispersions and nonlinearities. Therefore, the algorithms should requires lower computational complexity and lower processing latency. Taking FFT/IFFT in the backpropagation algorithm is only suitable for off-line processing, and the back-and-forth Fourier transforms cause heavy computational load if the number of input samples is large. Therefore, unlike fiber simulations that can use both frequency-domain and time-domain methods, practical backpropagation algorithms can only be implemented in the time domain based on IIR or FIR digital filters. The FIR filter is superior to IIR for real-time implementations because there is no need to consider the stability issues that always accompany with the IIR filters, and FIR filters can be implemented efficiently by parallel processors such as FPGA and DSP chips.

The time-domain digital backpropagation is similar to the time-domain simulation of signal propagation in fibers. The general algorithm is given by replacing (5.2), (5.3) and

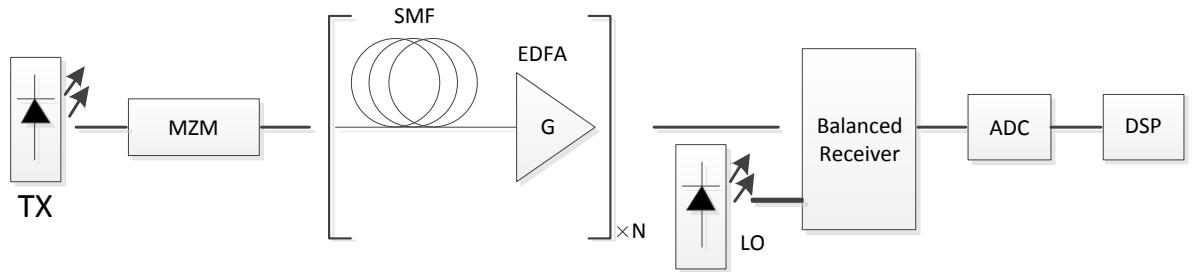

Fig. 5.2 Block diagram of a long-haul transmission system with inline amplification.

(5.5) with

$$\hat{A}_1(z+\Delta z/2,T) = h_{-D}(T) \otimes A(z,T)\exp(\alpha\Delta z/4)\ , \tag{5.7}$$

$$A(z+\Delta z,T) = h_{-D}(T) \otimes \hat{A}_2(z+\Delta z/2,T)\exp(\alpha\Delta z/4)\ , \tag{5.8}$$

where $h_{-D}(T)$ is a time-domain filter which has the same role as the frequency response in (5.6). Herein, the back-and-forth Fourier transform is avoided by introducing the time-domain digital filter that is well-fitted to the inverse of dispersion-related effects, which saves computational complexity and memory usage.

5.2 Performance of Time-Domain Backpropagation

5.2.1 Simulation Setup

In this section, we will apply the QCQP-based optimal filter to the time-domain digital backpropagation algorithm. The system under simulations is a polarization division multiplexed quadrature phase-shift keying (PDM-QPSK) system. It consists of DP-QPSK transmitter, transmission link, coherent receiver and DSP, according to Fig. 5.2. All processings except DSP are simulated using Optisystem 9.0 from Optiwave Systems Inc.. The simulation includes dispersion-related effects, SPM and polarization cross phase modulation. The PDM-QPSK signal is generated using the Mach-Zehnder modulators (MZM). After the signal propagates through the fiber channel, a coherent receiver is used to demodulate and decode. Specifically, the received signal is amplified, passed through a low-pass Gaussian filter and sampled by the ADC. The discrete-time signal is further processed by

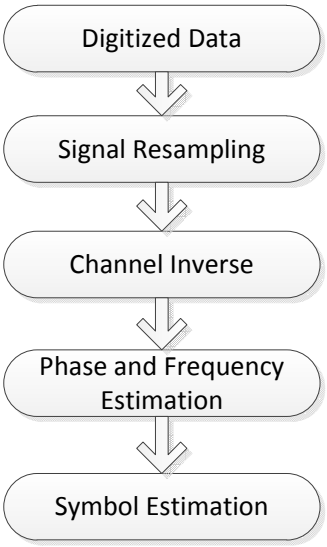

Fig. 5.3 DSP block.

the DSP module.

The detailed structure of the DSP module is illustrated in Fig. 5.3. The signal is firstly resampled to 2 samples per symbol, then the channel inverse for fiber impairment compensation is performed. The block of channel inverse filters the signal with the approximation of the inverse of the optical channel. In the simulation, when only dispersion is compensated, the dispersion operator D functions; when both dispersion and nonlinearity are compensated, we apply time-domain digital backpropagation algorithm in which dispersion filter is implemented based on an FIR filter. For the sake of low computational complexity, the FIR filter is extracted based on the QCQP algorithm. The time-domain digital backpropagation is performed with 1km per step and the resulting filter length is 45. In the beginning of each span, the effect of power gain in EDFA is removed, then followed by the channel inverse, whether compensates only dispersion or both dispersion and nonlinearity. The effect of propagation loss in fiber is also reversed. Afterwards, the Viterbi and Viterbi phase estimation algorithm compensates the phase and frequency mismatch, followed by symbol estimation. The DSP block is all processed in Matlab.

Various parameters are chosen as follows. The wavelength is 1550nm and the linewidth is 0.1MHz for the transmit laser. The baud rate of the DP-QPSK is 25Gbaud/s, giving a total throughput of 100 Gb/s. The transmission link includes a total of 15 spans. Each

span includes an 80km SMF and then an EDFA with the gain of 16dB. The transmission link is thus $15 \times 80 = 1200$km with dispersion $D = 17$ps/(nm·km), dispersion slop $S = 0.08$ps/(nm^2·km), attenuation $\alpha = 0.2$dB/km and nonlinear coefficient $\gamma = 1.3\text{W}^{-1}$/km. The noise figure of the EDFA is set as 5dB. At the coherent receiver, the wavelength of the local oscillator (LO) is 1549.99nm and the linedwidth is 1MHz, which means the frequency mismatch and laser phase noise are both taken into account.

5.2.2 Results and Discussions

We consider two processing methods used for the channel inverse part in DSP block: one compensates dispersion only and the other mitigates the effects of both dispersion and nonlinearities using backpropagation algorithm. Both X polarization and Y polarization are considered. Fig. 5.4 shows the eyediagrams of the output signals after being processed by these algorithms. The corresponding constellations are plotted in Fig. 5.5. When the backpropagation algorithm is used, the eyediagrams of the output signals generated by the digital backpropagation algorithm exhibit wider eye openings than the case when only dispersion is compensated. In turn, the constellation points are more concentrated for the digital backpropagation algorithm. They can be quantified in Q-factor later. The difference between the X polarization and the Y polarization is possibly due to the effect of polarization cross phase modulation. In a nutshell, the nonlinearities can degrade system performance and compensation techniques are necessary, especially for long-distance transmission systems. The digital backpropagation algorithm based on the proposed filters can compensate both dispersion and nonlinearities, and thus outperforms the algorithms that mitigate dispersion only.

The Q-factor is a quantitative performance measure of signal processing algorithms at receiver and is extracted from constellation[50]. We assume that the x and y axes are the decision thresholds for QPSK. The Q-factor is defined as

$$Q(dB) = 10\log_{10}\frac{\mu_x^2}{\sigma_x^2} = 10\log_{10}\frac{\mu_y^2}{\sigma_y^2} \tag{5.9}$$

where (μ_x, μ_y) is one arbitrary point of the four nominal constellation points of QPSK, and σ_x^2 or σ_y^2 is the variance of the points corresponding to the received signals (after processed by DSP) in the x or y direction. Geometrically, the Q-factor describes how

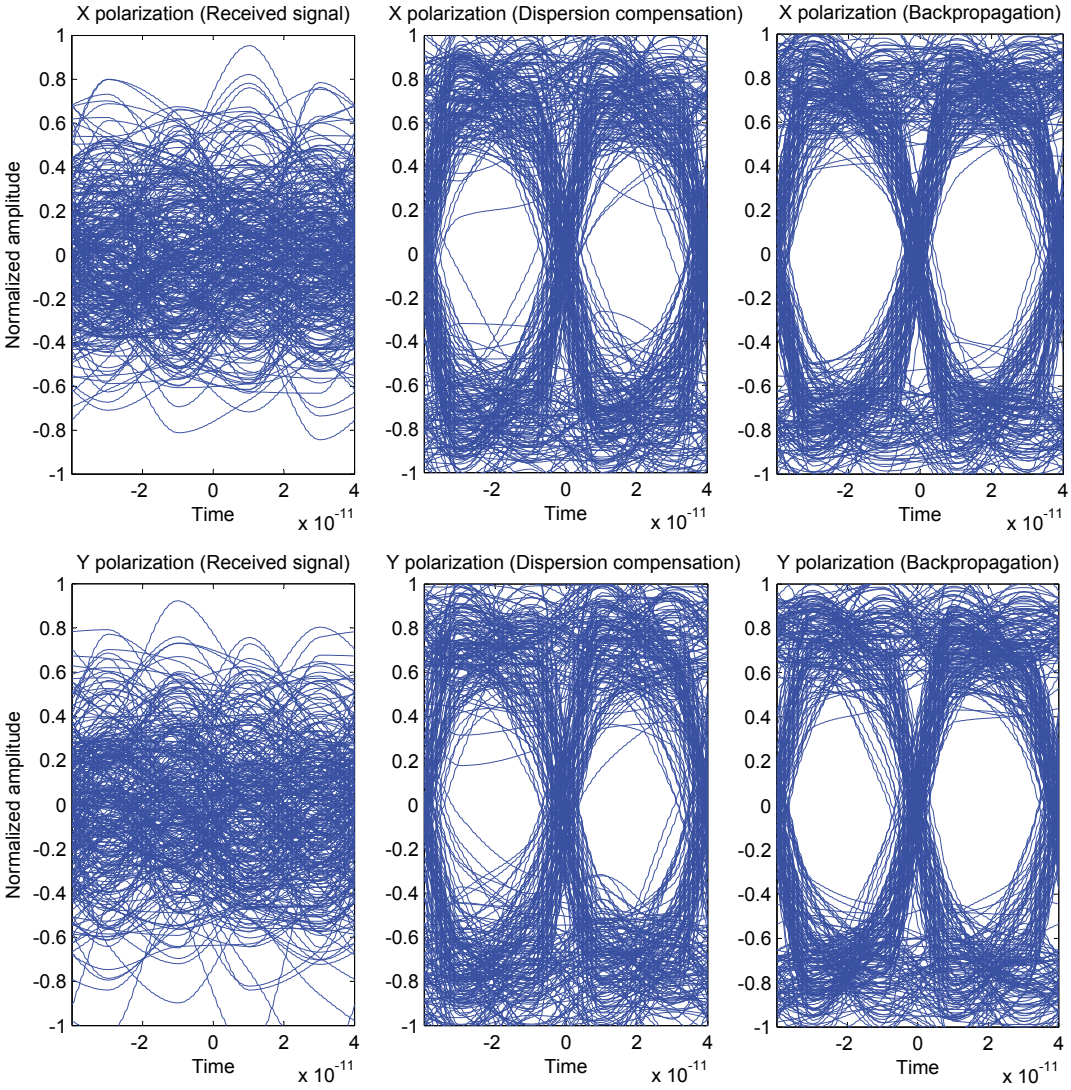

Fig. 5.4 Eye diagrams of signals before and after the DSP block.

the actually signals concentrate around the nominal constellation point. The scatterplots in Fig. 5.5 can be used to estimate the Q-factors. Only with dispersion compensation, the Q-factor is 11.7dB for the X polarization and 11.9dB for the Y polarization. With digital backpropagation, the Q-factor is 13dB for the X polarization and 12.9dB for the Y polarization. By employing the digital backpropagation algorithm to mitigate the effects of nonlinearities, we can obtain a Q-factor gain of 1dB.

The input power is an important factor in analyzing the Q-factor of the output sig-

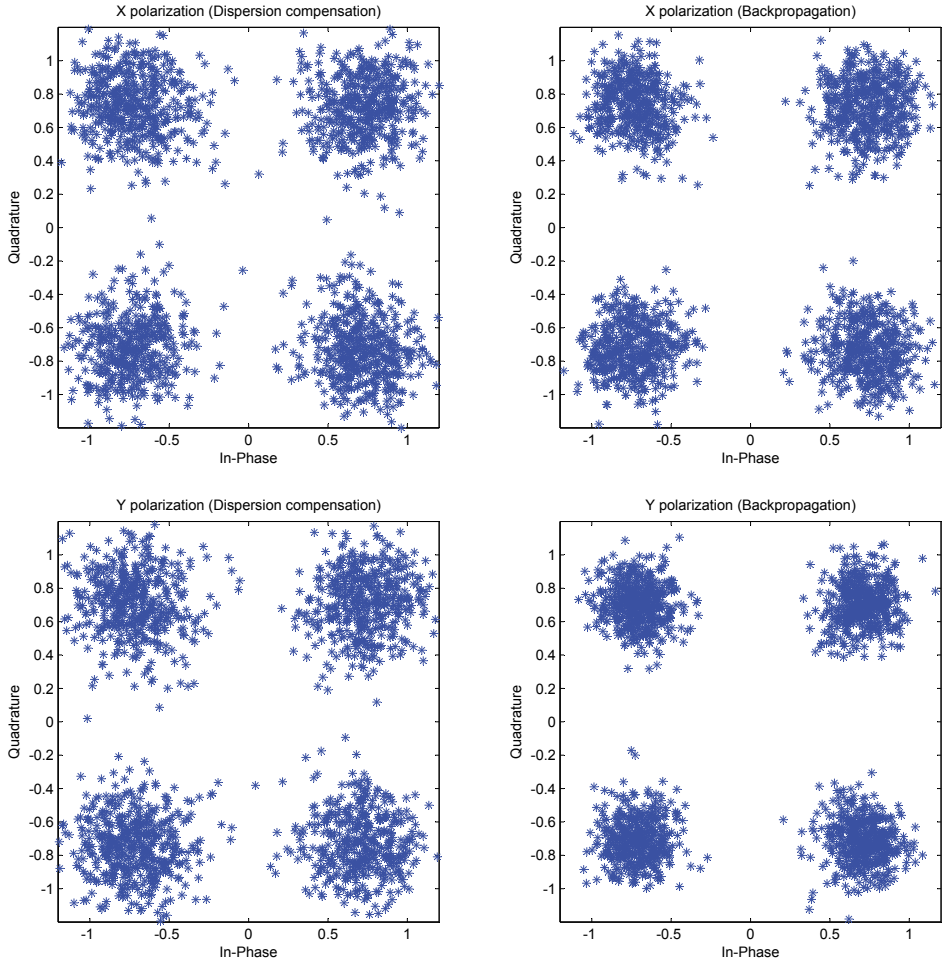

Fig. 5.5 Constellations of the signals after being processed by different DSP algorithms.

nals from the DSP module. Fig. 5.6 plots the Q-factor as a function of the input power ranging from −2dBm to 10dBm for both dispersion compensation and backpropagation. The Q-factors here are the average values of the X polarization and the Y polarization. The transmission distance is 1200km. As the input power increases, the Q-factor initially increases and then decreases after reaching a maximum value. When the input power is low, the nonlinearities are not strong and hence whether they are compensated or not does not affect system performance. The performances of the two algorithms are very close to each other. When the input power is high, the nonlinearities dominate over other effects and it becomes more necessary to compensate the nonlinearities. This can be seen from the

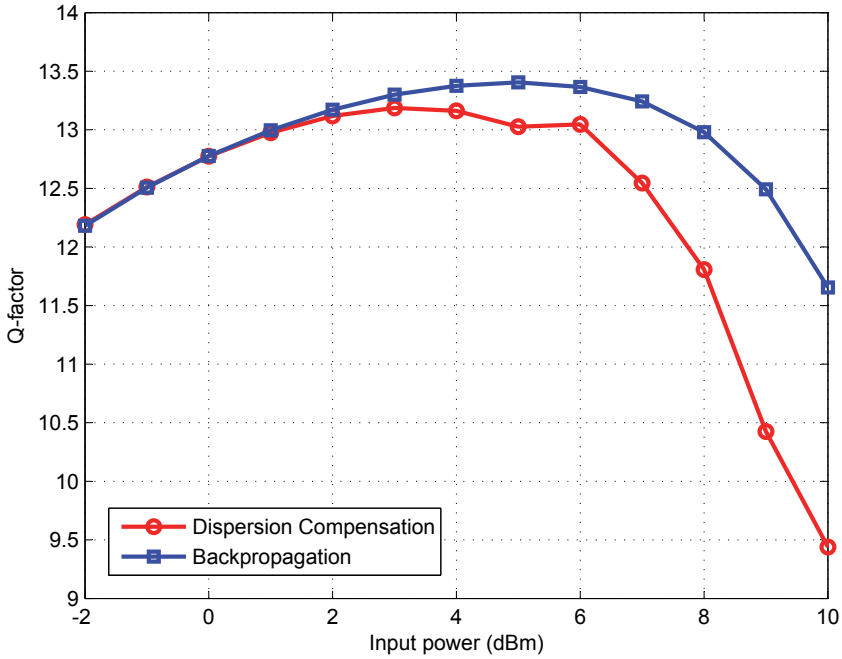

Fig. 5.6 The Q-factor versus the input power.

fact that the performance gap between digital backpropagation and dispersion compensation increases significantly as the input power increases. The maximum performance gain brought forth by compensating the nonlinearities can be as large as 2dBm. The optimal input power is 5dBm when the Q-factor attains the maximum value.

The transmission distance also plays an important role in the performance evaluation of long-haul transmission systems. As the transmission distance increase, the dispersion and nonlinearities accumulate and distort the signal waveform gradually. Furthermore, each additional span adds more ASE noise to the signal. In our simulation, the input power is 6dBm and Fig. 5.7 shows the Q-factor as a function of the propagation distance. Again, the Q-factors are the average values of the X polarization and the Y polarization. The Q-factor after dispersion compensation or digital backpropagation decreases approximately in a linear manner as the propagation distance increases. The performance gap between the two algorithms is becoming larger and larger as the propagation distance increases. Their performance are almost the same at the 160km because the nonlinearities are not evident for short distance of propagation. The gap between these two curves will also become larger if we increase the input power, which actually increases the nonlinearities. Lastly,

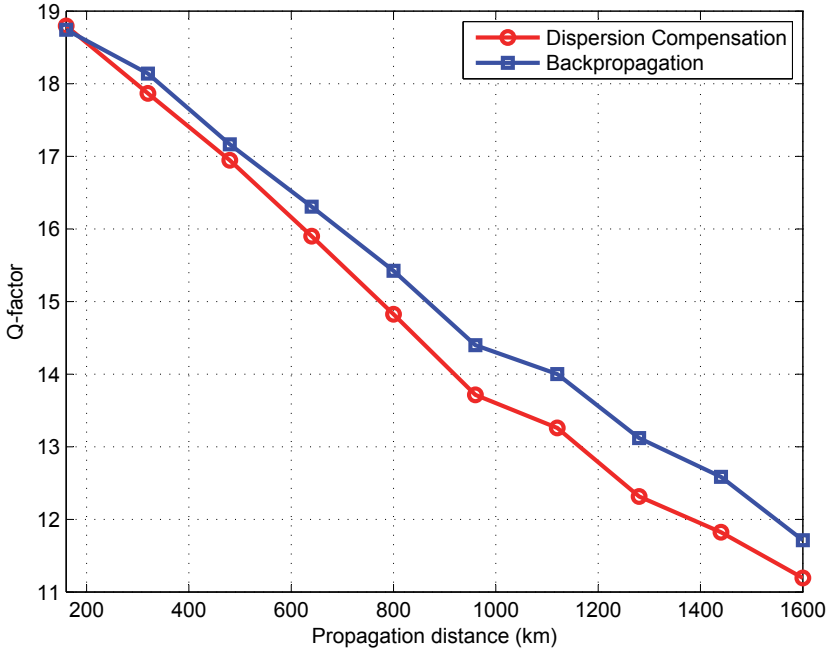

Fig. 5.7 The Q-factor versus propagation distance.

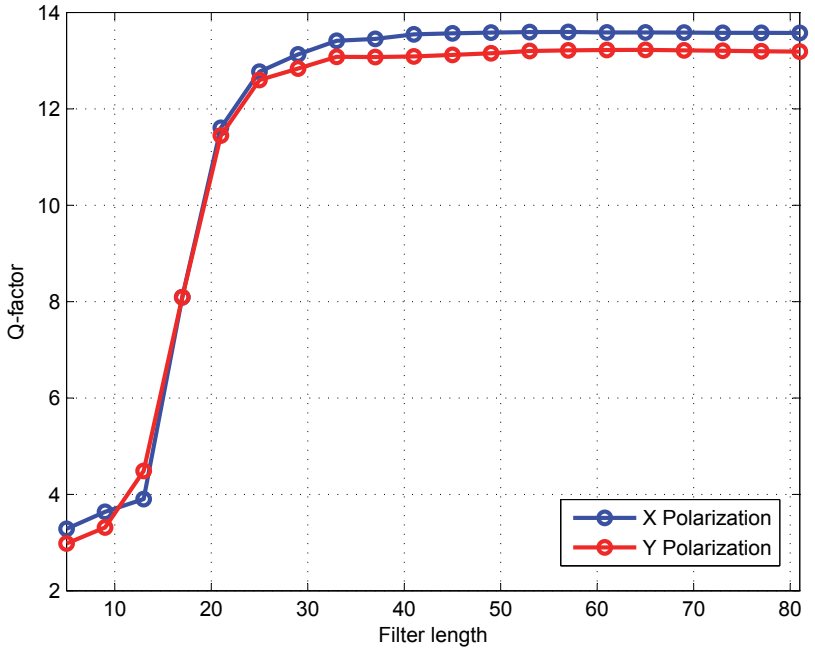

Fig. 5.8 The Q-factor as a function of the FIR filter length.

we plot the relationship between the Q-factor and the filter length for the back propagation algorithm. The step size is chosen as 1km and the input power is 6dBm. The Q-factor is plotted against different filter lengths from 5 to 81 in Fig. 5.8. An insufficient filter length can result in worse performance because the frequency response of the FIR filter deviates from the desired dispersion characteristics. The optimal filter length is around 41 and increasing the filter order any further does not increase the Q-factor. The optimal Q-factor for the X polarization are higher than that for the Y polarization, which is consistent with previous discussions.

So far, we have investigated the application of the proposed filters to time-domain digital backproagation, which successfully improves the performance of DP-QPSK systems. It should be noted that here we didn't consider random polarization state fluctuations. If these effects are included in the simulations, the polarization demultiplexing can be performed using constant modulus algorithms.

5.3 Implementation and Computational Complexity

For digital backpropagation algorithms, the implementation and computational complexity are of crucial importance. Time-domain methods are superior to the frequency-domain approach when the number of samples to be processed are huge. The key to choosing a time-domain method is to select the right digital filter. The IIR filters have low filter orders but are subject to stability issues due to its feedback components. The FIR filters that we used in this chapter do not have the stability problem and are good for real-time implementations in parallel processors.

Now, we analyze the computational complexity and processing latency of the time-domain digital backpropagation algorithm. The analysis is similar to that in [51]. Let N_{pb} be the number of the parallelization branches, N_{step} be the number of steps, and M' be the FIR filter length. Each FIR filter requires $4M'$ multiplications and $4M'-2$ summations. Taking into consideration 16 multiplications and 7 summations needed by the nonlinear operator, the number of multiply-accumulate (MAC) units is $N_{step} \times N_{pb}(8M'+16)$ for non-iterative S-SSM. Assuming each multiplication or summation requires half a clock cycle $T/2$, each FIR filter requires $(\lceil \log_2 M' \rceil + 2) \times T/2$ computational time whereas each nonlinear operator requires $7 \times T/2$. Thus, the latency is $N_{step}(\lceil \log_2 M' \rceil + 11) \times T/2$. Moreover, the FIR filters can be implemented using overlap-add and overlap-save methods

as discussed before, which can further reduce the complexity of linear convolution.

5.4 Summary

We have verified our proposed filter for time-domain digital backpropagation. The algorithm compensates for both dispersion and nonlinearities. Compared with the algorithm that only compensates dispersion, the Q-factor is improved for DP-QPSK long-haul transmission systems. The proposed FIR filters are suitable for real-time implementation of the digital backpropagation algorithm.

Chapter 6

Conclusions

6.1 Summary

In this thesis, we proposed the least square filters used for time-domain implementation of the linear dispersion operator in split-step methods. Our work presents a systematic and comprehensive study of this problem that provides theoretical and practical insights. The proposed filters can be used in both time-domain simulations of pulse propagation in optical fiber and time-domain digital backpropagation for fiber impairment compensation.

Chapter 2 formulated the least square problem which minimizes the integral of squared error between the FIR frequency response and desired dispersion characteristics. This least square problem has been solved using two approaches: one is normal equation and the other is based on SVD. The normal equation gives an explicit solution whereas SVD approach provides geometrical and mathematical insights. Geometrically, the frequency response of the optimal filter is the orthogonal projection of the desired dispersion filter into the subspace spanned by a set of DPSWF's. In parallel, the optimal filter is a linear combination of the time-domain counterparts of these DPSWF's, namely, a set of index-limited DPSS's.

Chapter 3 investigated the numerical issues of the problem. The theory of DPSS reveals that the least square problem could be ill-conditioned. Adding a regularization term to the objective function and using adaptive quadrature techniques to compute the integral both contribute to overcoming the ill-conditioned property of this problem. We also analyzed the negative effects of overshooting and successfully controlled it either by increasing the length of the regularized LS filter, or imposing a maximum magnitude constraint. The

latter results in a QCQP problem that can be solved by state-of-the-art interior point methods. These filters are easy to extract and the QCQP-based filter saves the filter length by at least 1/3. Theoretical investigation of the filter order based on the concept of group delay and instantaneous frequency, was successfully verified by numerical experiments. The established linear relationship simplifies the task of choosing the step size and the filter length.

In Chapter 4, we applied the proposed filters to simulations of pulse propagation in optical fiber. We discussed the effects of the squared error on the reliability of the simulations. Numerical experiments of single channel and WDM channels verified the validity of the proposed filters. They can generate reliable outputs even after thousands of kilometers of propagation. Finally, we introduce the overlap-add and overlap-save methods that can reduce the computational complexity of linear convolution significantly.

In Chapter 5, we extended the applicability of the proposed filters by employing them in digital backpropagation algorithms. These algorithms are used to compensate fiber impairments, which is the inverse process of pulse simulation. The time-domain split-step methods based on FIR filters are preferable for real-time implementation. Simulation results show that the digital backpropagation algorithms based on the proposed filters are able to successfully compensate for both dispersion and nonlinearities, and thereby improve system performance. This is especially necessary in the case of large input power and long-haul transmission system. Other implementation and complexity issues are addressed at the end of this chapter.

6.2 Future Works

We have proposed two filters: the unconstrained LS filter and the QCQP-based filter. Firstly, our work can be extended to the scenarios of variable step size. Variable step-size technique is advantageous in accuracy and complexity. As step size changes from one step to another, the FIR filter needs to be updated frequently. Thus, the computational complexity of extracting the filter is a major concern. The unconstrained LS filter can be obtained based on the fast MLD algorithm, and are therefore suitable for variable step-size simulations. However, the overall computational complexity should be further quantified and investigated.

Secondly, PMD can be implemented in the time-domain, along with fiber dispersion.

PMD is typically emulated by random concatenations of birefringent fiber sections. Therefore, in each section, frequency-domain transfer matrices can only be implemented in the time-domain based on FIR filters, which also requires an efficient design of the FIR filters. The unconstrained LS filter may be a good candidate, but more work is needed in complexity analysis, reliability verification, and other practical issues.

The QCQP-based filter is suitable for time-domain digital backpropagation due to its short length. However, the computational complexity of backpropagation is still high, and more improvements and fine-tuning are needed for real-time implementation.

References

[1] A. Gnauck, R. Tkach, A. Chraplyvy, and T. Li, "High-capacity optical transmission systems," *IEEE/OSA Journal of Lightwave Technology*, vol. 26, no. 9, pp. 1032–1045, May 2008.

[2] A. Ellis, J. Zhao, and D. Cotter, "Approaching the non-linear shannon limit," *IEEE/OSA Journal of Lightwave Technology*, vol. 28, no. 4, pp. 423–433, Feb. 2010.

[3] G. P. Agrawal, *Nonlinear fiber optics*, 4th ed. San Diego, CA: Academic Press, 2007.

[4] Q. Chang, E. Jia, and W. Sun, "Difference schemes for solving the generalized nonlinear Schrödinger equation," *Journal of Computational Physics*, vol. 148, pp. 397–415, 1999.

[5] O. Sinkin, Z. Holzlöhner, R., and C. J., Menyuk, "Optimization of the split-step Fourier method in modeling optical-fiber communications systems," *IEEE/OSA Journal of Lightwave Technology*, vol. 21, no. 1, pp. 61–68, Jan. 2003.

[6] A. Carena, V. Curri, R. Gaudino, P. Poggiolini, and S. Benedetto, "A time-domain optical transmission system simulation package accounting for nonlinear and polarization-related effects in fiber," *IEEE Journal on Selected Areas in Communications*, vol. 15, pp. 751–765, May. 1997.

[7] A. Lowery, O. Lenzmann, I. Koltchanov, R. Moosburger, R. Freund, S. Richter, A.and Georgi, D. Breuer, and H. Hamster, "Multiple signal representation simulation of photonic devices, systems, and networks," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 6, pp. 282–296, Mar./Apr. 2000.

[8] T. Kremp and W. Freude, "Fast split-step wavelet collocation method for WDM system parameter optimization," *IEEE/OSA Journal of Lightwave Technology*, vol. 23, no. 3, pp. 1491–1502, Mar. 2005.

[9] M. Delfour, M. Fortin, and G. Payr, "Finite-difference solutions of a non-linear schrödinger equation," *Journal of Computational Physics*, vol. 44, pp. 277–288, 1981.

[10] K. Peddanarappagari and M. Brandt-Pearce, "Volterra series approach for optimizing fiber-optic communications system designs," *IEEE/OSA Journal of Lightwave Technology*, vol. 16, no. 11, pp. 2046–2055, Nov. 1998.

[11] R. Hardin and F. Tappert, "Applications of the split-step Fourier method to the numerical solution of nonlinear and variable coefficient wave equations," *SIAM Review*, vol. 15, p. 423, 1973.

[12] X. Li, X. Chen, and M. Qasmi, "A broad-band digital filtering approach for time-domain simulation of pulse propagation in optical fiber," *IEEE/OSA Journal of Lightwave Technology*, vol. 23, no. 2, pp. 864–875, Feb. 2005.

[13] K. He and X. Li, "An efficient approach for time-domain simulation of pulse propagation in optical fiber," *IEEE/OSA Journal of Lightwave Technology*, vol. 28, no. 20, pp. 2912–2918, Oct. 2010.

[14] P. Winzer and R. Essiambre, "Advanced modulation formats for high-capacity optical transport networks," *IEEE/OSA Journal of Lightwave Technology*, vol. 24, no. 12, pp. 4711–4728, Dec. 2006.

[15] H. Ishio, J. Minowa, and K. Nosu, "Review and status of wavelength-division-multiplexing technology and its application," *IEEE/OSA Journal of Lightwave Technology*, vol. 2, no. 4, pp. 448–463, Aug. 1984.

[16] C. Brackett, "Dense wavelength division multiplexing networks: Principles and applications," *IEEE Journal on Selected Areas in Communications*, vol. 8, no. 6, pp. 948–964, Aug. 1990.

[17] E. Ip, A. Lau, D. Barros, and J. Kahn, "Coherent detection in optical fiber systems," *Optics Express*, vol. 16, no. 2, pp. 753–791, Jan. 2008.

[18] P. Winzer, A. Gnauck, C. Doerr, M. Magarini, and L. Buhl, "Spectrally efficient long-haul optical networking using 112-gb/s polarization-multiplexed 16-qam," *IEEE/OSA Journal of Lightwave Technology*, vol. 28, no. 4, pp. 547–556, Feb. 2010.

[19] A. Srivastava, Y. Sun, J. Zyskind, and J. Sulhoff, "Edfa transient response to channel loss in wdm transmission system," *Photonics Technology Letters, IEEE*, vol. 9, no. 3, pp. 386–388, Mar. 1997.

[20] E. Ip and J. Kahn, "Fiber impairment compensation using coherent detection and digital signal processing," *IEEE/OSA Journal of Lightwave Technology*, vol. 28, no. 4, pp. 502–519, Feb. 2010.

[21] A. Chraplyvy, "Limitations on lightwave communications imposed by optical-fiber nonlinearities," *IEEE/OSA Journal of Lightwave Technology*, vol. 8, no. 10, pp. 1548–1557, Oct. 1990.

[22] C. Shannon and W. Weaver, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, no. 379, p. 623, 1948.

[23] P. Mitra and J. Stark, "Nonlinear limits to the information capacity of optical fibre communications," *Nature*, vol. 411, no. 6841, pp. 1027–1030, Jun. 2001.

[24] G. Weiss and A. Maradudin, "The baker-hausdorff formula and a problem in crystal physics," *Journal of Mathematical Physics*, vol. 3, p. 771, Jul./Aug. 1962.

[25] A. Quarteroni, R. Sacco, and F. Saleri, *Numerical mathematics*. Berlin Heidelberg, New York: Springer-Verlag, 2007.

[26] Y. Ye, *Interior point algorithms: theory and analysis*. New York: John Wiley & Sons, 1997.

[27] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge, UK: Cambridge University Press, 2004.

[28] X. Li, X. Chen, G. Goldfarb, E. Mateo, I. Kim, F. Yaman, and G. Li, "Electronic post-compensation of wdm transmission impairments using coherent detection and digital signal processing," *Optics Express*, vol. 16, no. 2, pp. 880–888, Jan. 2008.

[29] E. Mateo, L. Zhu, and G. Li, "Impact of xpm and fwm on the digital implementation of impairment compensation for wdm transmission using backward propagation," *Optics Express*, vol. 16, no. 20, pp. 16 124–16 137, Sep. 2008.

[30] E. Ip and J. Kahn, "Compensation of dispersion and nonlinear impairments using digital backpropagation," *IEEE/OSA Journal of Lightwave Technology*, vol. 26, no. 20, pp. 3416–3425, Oct. 2008.

[31] G. W. Stewart, *Afternotes Goes to Graduate School*. Philadelphia, PA: SIAM, 1998.

[32] Z. Battles and L. Trefethen, "An extension of MATLAB to continuous functions and operators," *SIAM Journal on Scientific Computing*, vol. 25, no. 5, pp. 1743–1770, 2004.

[33] L. Trefethen, "Householder triangularization of a quasimatrix," *IMA Journal of Numerical Analysis*, vol. 30, no. 4, pp. 887–897, 2010.

[34] J. Hunter and B. Nachtergaele, *Applied analysis*. Singapore: World Scientific, 2001.

[35] J. Nocedal and S. Wright, *Numerical optimization*, 2nd ed. New York: Springer-Verlag, 2006.

[36] Å. Björck, *Numerical methods for least squares problems*. Philadelphia, PA: SIAM, 1996.

[37] D. Slepian, "Prolate spheroidal wave functions, Fourier analysis, and uncertainty—V: The discrete case," *Bell System Technical Journal*, vol. 57, no. 5, pp. 1371–1430, May/Jun. 1978.

[38] G. Golub and C. Van Loan, *Matrix computations*. Baltimore, MD: Johns Hopkins Univ. Press, 1996.

[39] E. Phan-huy Hao, "Quadratically constrained quadratic programming: Some applications and a method for solution," *Mathematical Methods of Operations Research*, vol. 26, no. 1, pp. 105–119, 1982.

[40] R. Farhoudi and K. Mehrany, "Time-domain split-step method with variable step-sizes in vectorial pulse propagation by using digital filters," *Optics Communications*, vol. 283, no. 12, pp. 2518–2524, 2010.

[41] S. Savory, "Digital filters for coherent optical receivers," *Optics Express*, vol. 16, no. 2, pp. 804–817, Jan. 2008.

[42] G. Bosco, A. Carena, V. Curri, R. Gaudino, P. Poggiolini, and S. Benedetto, "Suppression of spurious tones induced by the split-step method in fiber systems simulation," *Photonics Technology Letters, IEEE*, vol. 12, no. 5, pp. 489–491, May 2000.

[43] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-time signal processing*, 2nd ed. Upper Saddle River, NJ: Prentice Hall, 1999.

[44] A. Elrefaie, R. Wagner, D. Atlas, and D. Daut, "Chromatic dispersion limitations in coherent lightwave transmission systems," *IEEE/OSA Journal of Lightwave Technology*, vol. 6, no. 5, pp. 704–709, May 1988.

[45] J. Gordon and L. Mollenauer, "Effects of fiber nonlinearities and amplifier spacing on ultra-long distance transmission," *IEEE/OSA Journal of Lightwave Technology*, vol. 9, no. 2, pp. 170–173, Feb. 1991.

[46] H. Bulow, F. Buchali, and A. Klekamp, "Electronic dispersion compensation," *IEEE/OSA Journal of Lightwave Technology*, vol. 26, no. 1, pp. 158–167, Jan. 2008.

[47] A. Lowery, "Fiber nonlinearity pre-and post-compensation for long-haul optical links using ofdm," *Optics Express*, vol. 15, no. 20, pp. 12 965–12 970, Sep. 2007.

[48] G. Li, "Recent advances in coherent optical communication," *Advances in optics and photonics*, vol. 1, no. 2, pp. 279–307, Feb. 2009.

[49] R.-J. Essiambre, G. Kramer, P. Winzer, G. Foschini, and B. Goebel, "Capacity limits of optical fiber networks," *IEEE/OSA Journal of Lightwave Technology*, vol. 28, no. 4, pp. 662 –701, Feb. 2010.

[50] A. Lowery, L. Du, and J. Armstrong, "Performance of optical ofdm in ultralong-haul wdm lightwave systems," *IEEE/OSA Journal of Lightwave Technology*, vol. 25, no. 1, pp. 131–138, Jan. 2007.

[51] L. Zhu, X. Li, E. Mateo, and G. Li, "Complementary fir filter pair for distributed impairment compensation of wdm fiber transmission," *Photonics Technology Letters, IEEE*, vol. 21, no. 5, pp. 292–294, Mar. 2009.