Estimation of Sparse Channels in IR-UWB Systems

Maryam Vala

Department of Electrical & Computer Engineering
McGill University
Montreal, Canada

December 2013

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Master of Engineering.

Abstract

Ultra Wideband (UWB) is a rapidly growing technology for digital wireless communications. UWB utilizes low power, ultra short, pulses and is specifically suited for short-range, high-rate indoor wireless communications as well as fine localization applications. The attractive properties of UWB are a direct consequence of its very wide bandwidth that also implies an extremely high Nyquist sampling rate, so digital processing of UWB signals requires the use of fast and efficient, and therefore, expensive, Analog to Digital Converters. UWB systems are commonly used in indoor environments in which propagation is characterized by long but sparse multipath channels. Multi-path channels cause intersymbol interference (ISI) and introduce distortion to the received signal. To counter the effects of the channel, the channel impulse response must be accurately estimated at the receiver. There are two main estimation approaches for UWB channels: Compressed Sensing (CS)-based and Adaptive Filter (AF)-based. The former combine sampling and compression into a single linear measurement process that operates at sub-Nyquist rate by capitalizing on the sparsity of UWB signals. The latter, uses Least-Mean-Square filtering principles to obtain low complexity channel estimation methods. The problem we consider in this thesis is the estimation of a long sparse multi-path channel in an UWB system. We focus on systems that use Pulse Position Modulation (PPM) that is one of the most commonly used modulation schemes in UWB systems. We review CS and AF based methods and then we propose new AF-type methods specifically for the estimation of sparse multi-path channels with PPM inputs. The main idea behind the proposed methods is to estimate the long channel in sections in order to reduce the computational cost and improve the estimation performance. Finally, we present simulation results showing the superior performance of the proposed algorithms.

Résumé

Ultra Wideband (UWB) est une technologie en croissance rapide de communications numériques sans fil. UWB utilise une faible puissance, ultra court, impulsions et est spécifiquement adapté á courte portée, haut-débit sans fil intérieure communications ainsi que la localisation fine applications. Les propriétés intéressantes de UWB sont une conséquence directe de son très large bande passante, ce qui implique également un taux extrêmement élevé Nyquist (taux d'échantillonnage, traitement numérique des signaux UWB requiert l'utilisation de rapide et efficace, et donc, cher, convertisseurs analogiques-numériques. UWB systèmes sont couramment utilisés dans les environnements intérieurs dont la propagation est caractérisé par un long mais clairsemée multipath canaux. Multi-canaux chemin cause intersymbol causant des interférences (ISI) et introduire de distorsion du signal reçu. Pour contrer les effets du canal, le canal réponse impulsion doit être estimée avec précision au niveau du récepteur. Il existe deux principales approches estimation pour UWB canaux: Comprimé de télédétection (CS)-fondée et filtre adaptatif (AF)-fondée. L'ancienne moissonneuse-batteuse d'échantillonnage et de compression dans une seule mesure linéaire processus qui fonctionne en sous-fréquence Nyquist en capitalisant sur la faible densité des signaux UWB. Ce dernier, utilise moins Mean Square filtrage principes pour obtenir une faible complexité canal méthodes d'estimation. Le problème que nous considérons dans cette thèse est l'estimation d'un long sparse multi-chemin canal d'un système UWB. Nous nous concentrons sur les systèmes qui utilisent Pulse Position Modulation (PPM) qui est l'un des plus couramment utilisé schémas de modulation en systèmes UWB. Nous revoir CS et AF fondées sur des méthodes et ensuite nous proposer de nouvelles AF-type méthodes spécifiquement pour l'estimation des sparse multi-chemin canaux avec PPM entrées. L'idée principale derrière les méthodes proposées est d'estimer le long canal en sections afin de réduire les coûts informatiques et améliorer les méthodes d'estimation des performances. Enfin, nous présentons les résultats de simulation montrant la supériorité de la performance des algorithmes proposés.

Acknowledgments

Appreciation is extended to Dr. Ioannis Psaromiligkos whose recommendations and advices have been essential for the fulfillment of this thesis have been extremely helpful. I would also like to thank my family for their help and support. Finally, I would also like to thank the Department of Electrical and Computer Engineering for its invaluable help in providing equipment and resources.

Contents

List of Figures

List of Tables

List of Acronyms

UWB	Ultra Wideband
IR-UWB	Impulse Radio Ultra Wideband
ADCs	Analog to Digital Conversions
RF	Radio Frequency
ISI	Intersymbol Interference
CS	Compress Sensing
AF	Adaptive Filter
MSE	Mean Square Error
IF	Intermediate Frequency
FIR	Finite Impulse Response
WLANs	Wireless Local Area Networks
BER	Bit Error Rate
GPS	Global Positioning System
PPM	Pulse Position Modulation
PSD	Power Spectral Density
TH-PPM	Time-Hopping Pulse Position Modulation
TH-BPSK	Time-Hopping Binary Phase Shift Keying
LMS	Least Mean Square
ZALMS	Zero Attractor Least Mean Square
RZALMS	Reweighted Zero Attractor Least Mean Square
DSP	Digital Signal Processing
SV	Saleh Valenzuela
SD	Steepest Descent
MMSE	Minimum Mean Square Error

FCC	Federal Communications Commission
MAI	Multiple Access Interference
RIP	Restricted Isometry Property
CIR	Channel Impulse Response
MP	Matching Pursuit
OMP	Orthogonal Matching Pursuit
BS	Basis Pursuit
GSM	Global System for Mobile Communications
IEEE	Institute of Electrical and Electronics Engineers
AWGN	Additive White Gaussian Noise
LP	Linear Programm
MUI	Multiple User Interference

Chapter 1

Introduction

1.1 Overview

Digital wireless communication has experienced a significant growth in recent years driven by increasing demand for more effective and reliable data communication. In particular, the Ultra-Wideband (UWB) technology has gained in popularity because of its high data rate capabilities in indoor wireless applications. UWB signals are formed of ultra-short (nanosecond-scale) pulses whose position or phase encodes information. The short duration of the pulses results in a very wide spectrum in the frequency domain (up to 7GHz), hence the name UWB.

Due to their wide bandwidth UWB signals propagate through frequency selective channels, also called multi-path fading channels, that cause Intersymbol Interference (ISI) in the received signal. Indeed, the received signal comprises of time-shifted and attenuated versions of the transmitted signal. To recover the transmitted signal, accurate knowledge of the channel impulse response is required at the receiver. This information is usually obtained by a channel estimator based on the known sequence of transmitted symbols.

In UWB systems, the very fine time resolution of the transmitted signal results in a very large number of resolvable multi-path components which makes the channel estimation a particularly challenging problem [2]. In addition, the Channel Impulse Responses (CIRs) are typically sparse as they contain a large number of zero or negligible coefficients. These two characteristic properties of UWB channels have motivated research of efficient CIR estimation techniques along two main avenues: Compressed Sensing (CS)-based and Adaptive Filter (AF)-based approaches. Both approaches capitalize on the sparsity of the

CIR to offer improved performance but they exploit this property in different ways.

CS-based methods combine sampling and compression into a single linear measurement process that operates at sub-Nyquist rate. The main motivation behind these methods is cost reduction through the use of inexpensive Analog-to-Digital Converters (ADCs). Due to the large bandwidth of UWB signals, the corresponding Nyquist rate is very high and Analog-to-Digital conversion requires very costly high-speed Analog-to-Digital Converters (ADCs), in addition to very high storage and processing demands. CS-based methods avoid the use of high-speed ADCs; however, they require the use of advanced and very complex signal processing techniques.

AF-based methods, on the other hand, rely on Least-Mean-Square (LMS) system identification principles. They are conceptually simple, have low complexity, and they offer improved channel estimation performance and tracking speed by enforcing a sparse structure on the channel estimate. However, they require sampling at the Nyquist rate.

1.2 Motivation and Objective

The problem we consider in this thesis is the estimation of multi-path channels in UWB systems. We focus on systems that use Pulse Position Modulation (PPM) that is one of the most commonly used modulation schemes in UWB systems. Motivated by recent advances in CS and adaptive signal processing of sparse signals, we propose novel, simple and cost effective methods for UWB channel estimation. Conventional channel estimation methods do not work well for the typically very long and highly sparse UWB channels. The reason is that they assume relatively short and dense CIRs and they cannot take advantage of the sparse nature of the UWB channels in order to effectively estimate them. The proposed methods are based on LMS system identification method and its zero-attracting variants that were developed for the identification of sparse systems. LMS is a low complexity iterative procedure for the estimation of the impulse response of a linear system given pairs of inputs and the corresponding outputs. In the proposed methods, and in contrast to traditional LMS methods the CIR is not estimated as a whole but in parts. This allows us to handle long channels in an efficient manner. We give a detailed analysis of the scenarios where the algorithm can be utilized to its maximum potential to estimate an UWB channel.

1.3 Thesis Outline

The thesis is organized as follows. In Chapter 2, we provide a general overview of UWB systems including the signal model, the UWB channel model, and introduce the notation and terminology used throughout the thesis. In Chapter 3, we provide the essential background material relating to this thesis. More specifically, we describe the basic theory of data compression by CS, and channel estimation using adaptive filtering. In Chapter 4, we describe the proposed channel estimation method in detail and analyze our findings regarding the channel estimation of the very long sparse multi-path channel in the case of UWB PPM input signals. Simulation results investigating the effect of the various design parameters on the performance are presented. Finally, in Chapter 5, we summarize our work and provide directions for future research.

Chapter 2

Ultra-Wideband Systems

An UWB signal consists of a train of ultra-short (nanosecond-scale) pulses called monocycles whose position or phase encodes information. The name UWB is due to the fact that very narrow pulses in the time domain result in a very wide spectrum in the frequency domain.

2.1 UWB definition

An UWB system, according to the Federal Communications Commission (FCC), is a radio system having a bandwidth exceeding the lesser of 500 MHz or 20% of its center frequency (f_c) [3]. The center frequency is defined as [2]

$$f_c = \frac{f_H + f_L}{2}, \tag{2.1}$$

with f_H (f_L) being the upper (lower) frequency of -10 dB bandwidth emission point (i.e., the frequency at which the magnitude spectrum has $1/\sqrt{10}$ of its peak value). The 10-dB bandwidth of the signal is then defined as [4]

$$B_{10dB} = f_H - f_L. \tag{2.2}$$

Prior to 2001 the interest in UWB devices was restricted to radar systems, mostly for military applications. There was a drastic change in 2002 when FCC allowed the operation of UWB radios over a frequency band, namely 3.1 – 10.6 GHz [3], of width up to 7.5

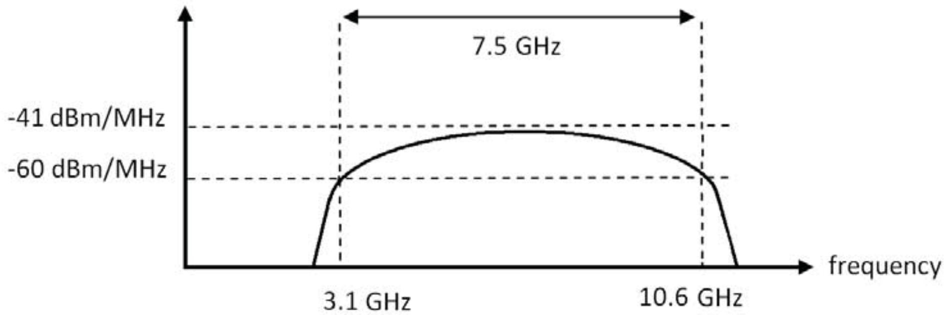

Figure 2.1 Power spectral density of UWB radio system.

GHz. As a result, UWB technology is presently being explored for high capacity wireless communications systems. However, the FCC imposes limits on the Power Spectral Density (PSD) of UWB signals; it should not exceed −41 dBm/MHz and should fall rapidly to −60 dBm/MHz at frequencies below 3.1 GHz [5]. The FCC imposed limits are shown in Fig. 2.1. The advantages, disadvantages, and application of UWB technology are described next.

2.1.1 Advantages and Disadvantages

The main advantages of UWB technology are listed below:

1. High data rate UWB communication technology is capable of achieving very high data rates (greater than 100 Mbps) over short and medium distances ($20 - 50$m) [6].

2. Immunity to multi-path fading. The very wide bandwidth of UWB signals implies a very fine time resolution that, in turn, allows the resolution of multi-path components in dense multi-path environments. Thus, it can result in low vulnerability to multi-path fading in dense multi-path environments [7, 2].

3. Increased security. UWB signals are hard to intercept because of their low average transmission power. In addition, UWB pulses are time-position-modulated with codes unique to each transmitter/receiver pair. This time modulation results in increased

security, because detecting picoseconds-wide pulses without knowing their distinct code is extremely difficult [8].

4. Low power and low interference. The average transmitted power spectral density of UWB signals as dictated by the FCC spectral mask is extremely low (-41.3dBm/MHz). This allows UWB systems to coexist with narrowband radio systems on the same spectrum without creating excessive interference [6].

On the other hand, UWB exhibits several disadvantages in real-world applications. Some of these are listed below:

1. Low transmission power. The low transmission power could be a disadvantage for UWB systems as it limits the communication range. As a result, UWB is not appropriate for long-range applications [6].

2. Potential interference. UWB shares a very wide bandwidth with other radio frequency technologies, resulting in increased risk of interference, from and to existing systems [9, 10].

3. Long synchronization times. The ultra-narrow pulses require use of complicate signal acquisition and time synchronization methods. It can take up to a few milliseconds for a transmitter and receiver to reach synchronization [11].

To counter these drawbacks, advanced Digital Signal Processing (DSP) methods can be utilized [2].

2.1.2 Applications

UWB radio is a promising technology with attractive features in wireless communications, networking, radar, imaging, and positioning systems. The motivations for commercial use of UWB radios are: (a) the rising demand for inexpensive portable devices with high-rate wireless communication capabilities at low power; (b) the crowding of the spectrum and lack of available frequencies [12]. There have been increased efforts by research institutions, industry, and government agencies to utilize the potential of UWB radios in different application areas, such as

1. Localization with centimetre-level precision for indoor environments. GPS only provides location estimates within meters and is unsuitable for indoor applications.

2. High-resolution ground-penetrating radar and through-wall imaging. The lower frequency components allow signal propagation through the ground (1GHz bandwidth) [13] and walls (2GHz bandwidth) [14] while the larger bandwidth provides higher resolution [15].

3. Monitoring patients dynamically and remotely [15].

4. Asset tracking for increased security using distance measurements between radios, UWB signals provide the potential of highly accurate location estimation. Rescue services could benefit by using UWB technology to find people inside buildings during emergencies [15, 2].

2.2 UWB indoor Multi-path channels

In wireless data communication environments, the receiver observes a superposition of attenuated and delayed versions of the original transmitted radio signal, called multi-path signal components. Reflections and refractions from obstacles and buildings are the cause of the multi-path propagation phenomenon. The multi-path effect is demonstrated in Fig. 2.2. Delays that are smaller than the information symbol duration T_s, are combined into a single propagation path with its corresponding delay and attenuation [16, 17]. In some cases, the different propagation paths add destructively and produce significant attenuation called fading in the received signal amplitude. When different frequency components of a transmitted signal experience different amounts of fading then frequency-selective fading occurs. For a mobile radio moving in an area that has many reflective obstacles, the rapid change in the geometry of the reflected propagation path causes fast fading. Fast fading is especially bothersome at frequencies above 1GHz because the relative phases of the multi-path signals can drastically vary with small differences in the lengths of the propagation paths [18]. However, UWB systems transmit ultra-short pulses, they are able to resolve multi-path components with much higher resolution than technologies using longer pulse durations [19].

The long delay spread of the channel (typically, several nanoseconds) relative to the monocycle duration implies that UWB systems are subject to less severe fading effects because fewer multi-path components combine during monocycle's duration. However, a long delay spread can have negative influence on the system performance the transmitted signal

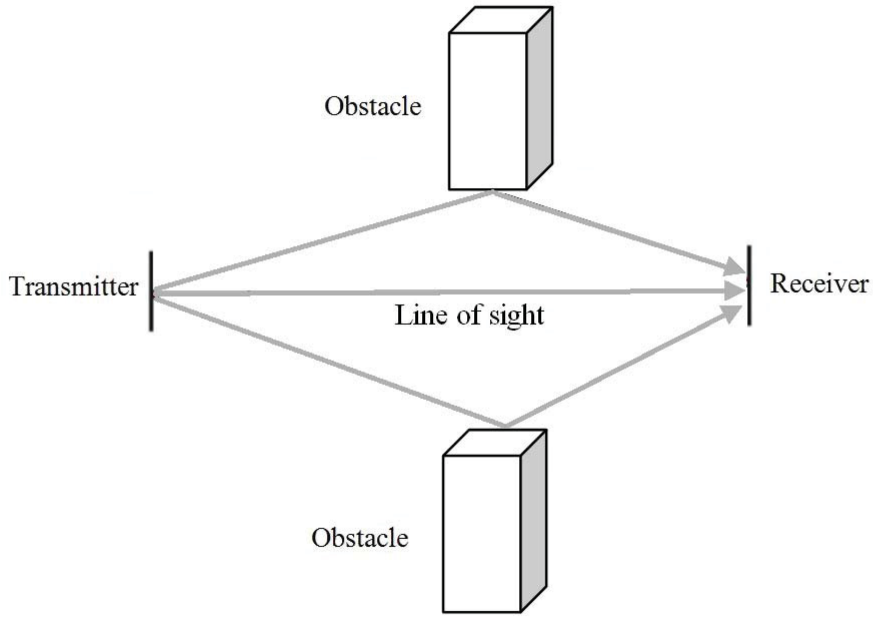

Figure 2.2 Multi-path propagation.

energy is distributed amongst too many multi-path components.

We now briefly describe the wireless channel model that IEEE 802.15.3a standardization group has established for the assessment of UWB communications systems and we consider the measurements that form the basis of this model. UWB multi-path channel are modelled as a linear time-varying system [20, 21]. In addition, it has been found that the channels are sparse, i.e., their impulse response has a large number of zero taps. The impulse response of the channel for the kth is given by [15]

$$h_k(t) = \sum_{\ell=1}^{L} \sum_{k=1}^{K} a_{\ell k}(t)\delta(t - \tau_{\ell k}(t)). \tag{2.3}$$

In Eq. (2.3) t is the observation time, L is the number of the resolvable paths, $\tau_{\ell k}(t)$ is the arrival time of the received signal via ℓth path for the kth user, $a_\ell(t)$ is the random time varying amplitude attenuation, and $\delta(\cdot)$ indicates the Dirac delta function [22]. Fig. 2.3 shows a multi-path channel realization obtained using the Saleh-Valenzuela (SV) [23] model with a pass-band (real) impulse response. The baseband (complex) impulse response of the original SV model has to be converted to a pass-band (real) impulse response by multiplying with a complex exponential of the appropriate carrier frequency and then taking the real part. The parameters involved in the multi-path channel generation are in Table 2.1.

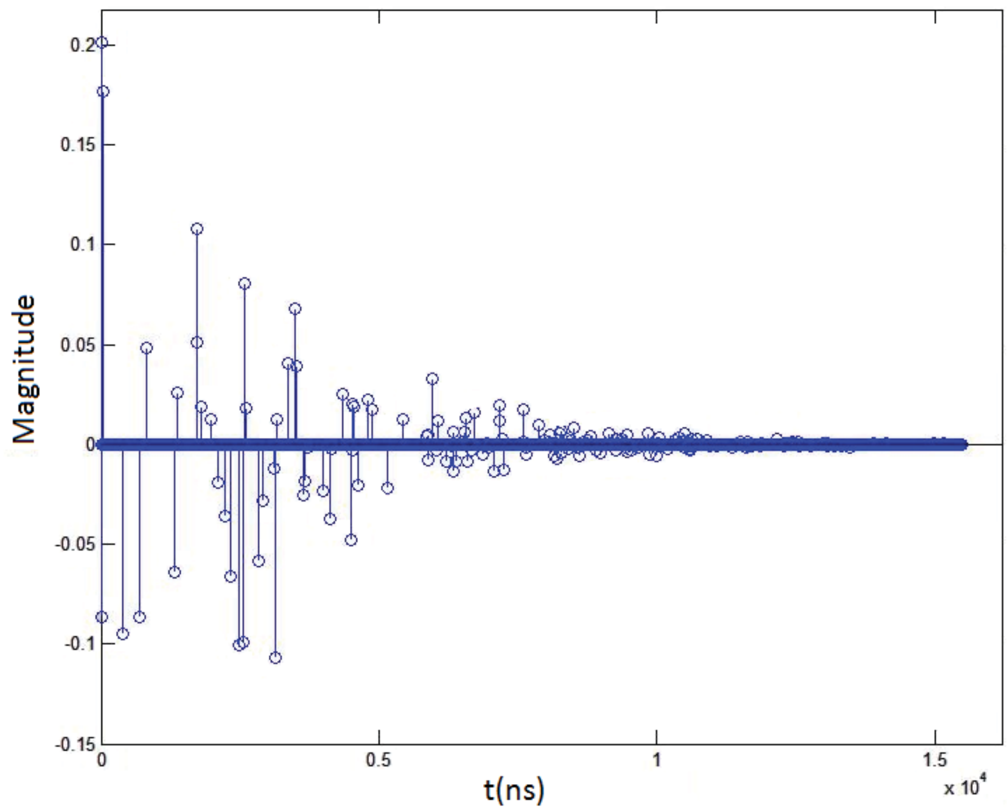

Figure 2.3 A typical multi-path propagation realization.

2.3 UWB Modulations

Time-Hopping Pulse Position Modulation (TH-PPM) and Time-Hopping Binary Phase Shift Keying (TH-BPSK) are the most common forms of signal modulation in UWB communications [7, 24, 25]. In TH-PPM, the information is conveyed by time shifts [19]. In TH-BPSK instead of modulating the signal by a time delay, 180^o phase shifts are used for binary signalling by multiplying each pulse by a user data code of amplitude ± 1 [26]. TH-PPM is the preferred scheme for the system architecture considered in this thesis because of its good spectral properties [27] and its simplicity [28].

Table 2.1 Parameters used for generating the multi-path channel realization in Fig. 2.3 (adopted from [1]).

Parameters definitions and notations	Values of parameters
Number of rays per cluster (κ)	25
Mean number of Clusters ($\bar{L}$)	3
Ray arrival rate for mixture of poisson processes (λ_1)[rays per ns]	1.54
Ray arrival rate for mixture of poisson processes (λ_2)[rays per ns]	0.15
Mixture probability (β)	0.095
Rate of cluster arrival (Λ) [1/ns]	0.047
Ray decay factor (γ_0) [ns]	12.53
Time dependence of ray decay factor (κ_γ)	0
Standard deviation of normally distributed variable for cluster energy ($\sigma_{cluster}$)[dB]	2.75
Cluster decay factor (Γ) [ns]	22.61
Mean of log-normal distributed nakagami-m factor for Small-scale Fading (m_0)[dB]	0.67
Time dependence of m_0 (k_m)	0
Standard deviation of log-normal distributed nakagami-m factor ($\hat{m_0}$)	0.28
Time dependence of $\hat{m_0}$ ($\hat{k_m}$)	0

2.3.1 Monocycles

Commonly used monocycles are the Gaussian monocycles defined as the first derivative of the Gaussian pulse. The Gaussian pulse is given by [29]

$$f(t) = \frac{1}{\sqrt{2\pi}\sigma}\ e^{-\frac{1}{2}(\frac{t-\mu}{\sigma})^2}, \tag{2.4}$$

where t represents time, the parameter σ is the standard deviation and controls the width of the pulse, and μ is the mean and controls the location of the peak of the pulse. Taking the derivative of (2.4) we can see that the Gaussian monocycle (also shown in Fig. 2.4) has the form [30]

$$p(t) = [1 - \frac{(t-\mu)^2}{\sigma^2}]\ e^{-\frac{1}{2}(\frac{t-\mu}{\sigma})^2}. \tag{2.5}$$

The Gaussian monocycle is simple to generate using a step-recovery diode and alternating current [31]. The parameter σ is related to the monocycle's duration T_p, that is, the effective time duration of the pulse over which 99.99% of the total monocycle energy is concentrated. It can be shown that $T_p = 2\pi\sigma$ where $\sigma = 0.1$ ns. When the transmitted pulse is the Gaussian monocycle, the received pulse will have the form of the 2^{nd} derivative of the Gaussian pulse. Indeed, the antenna's filtering effect can be modeled as a differentiator in UWB transmission [32]. The derivative of the Gaussian monocycle is illustrated in Fig. 2.5.

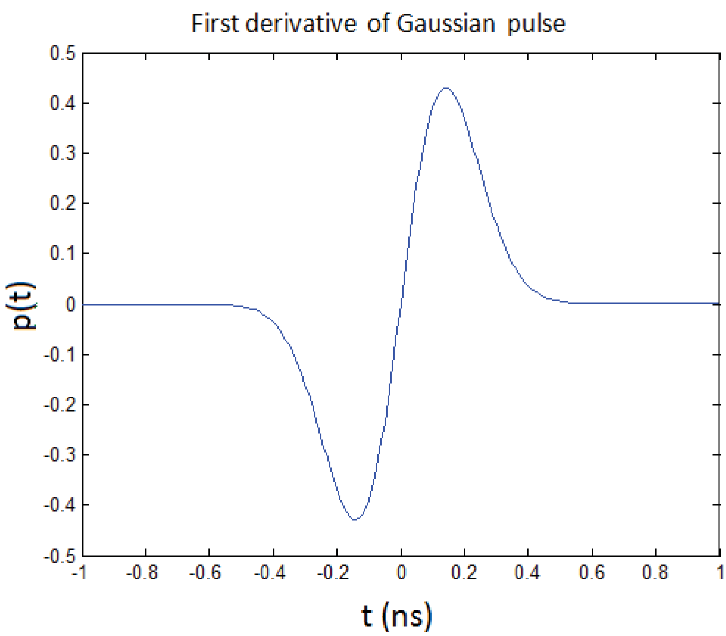

Figure 2.4 Gaussian monocycle pulse (Gaussian monocycle), $\mu = 0$, $\sigma = 0.1$ns.

2.3.2 UWB signal model

In TH-PPM (which is represented in Fig. 2.6), an information bit is transmitted over a symbol time interval T_s, which is divided into N_f frames of length T_f. Each frame is divided to N_h sub-frames called chips with time duration T_c. Moreover, N_s is the number of pulses required to transmit a single data bit that is set to 1 for simplicity [33, 7, 31, 34]. Let the number of users in the system be K. The kth user's, $1 \leq k \leq K$, PPM transmitted signal

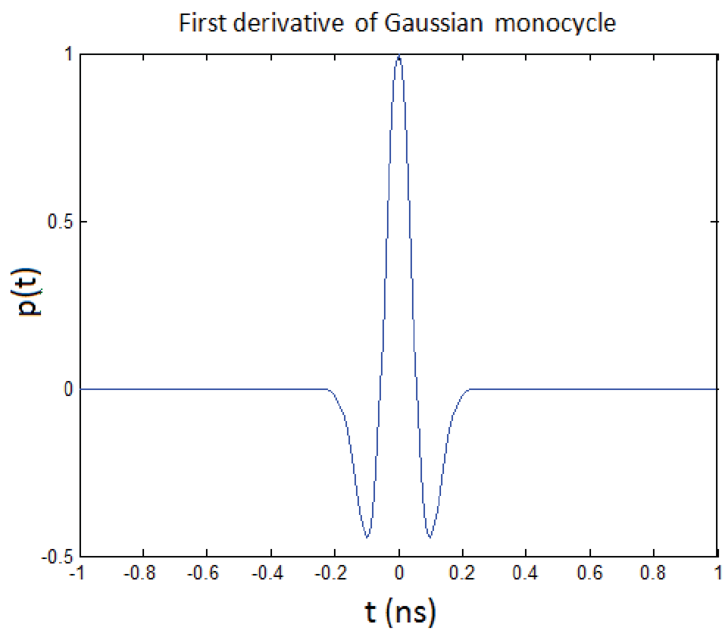

Figure 2.5 First derivative of Gaussian monocycle ($\mu = 0$, $\sigma = 0.1$ns).

$s^{(k)}(t)$ is given by [35]

$$s^{(k)}(t) = \sum_{k=1}^{N_u} \sum_{i=-\infty}^{\infty} \sum_{j=0}^{N_h-1} A^{(k)} p(t - iT_s - jT_f - c_j^{(k)} T_c - \epsilon d^{(k)}), \tag{2.6}$$

where

t represents time;

N_u number of users;

$A^{(k)}$ is the signal amplitude;

$p(t)$ represents the transmitted pulse waveform that begins at time zero at the transmitter. The pulse $p(t)$ is assumed to be normalized, i.e., $\int p^2(t)dt = 1$ [7, 24, 25];

T_f is the frame length, that is, the pulse repetition period. This is usually greater than the pulse width resulting in a low duty cycle signal [36];

N_h is the number of chips or sub-frames of a frame;

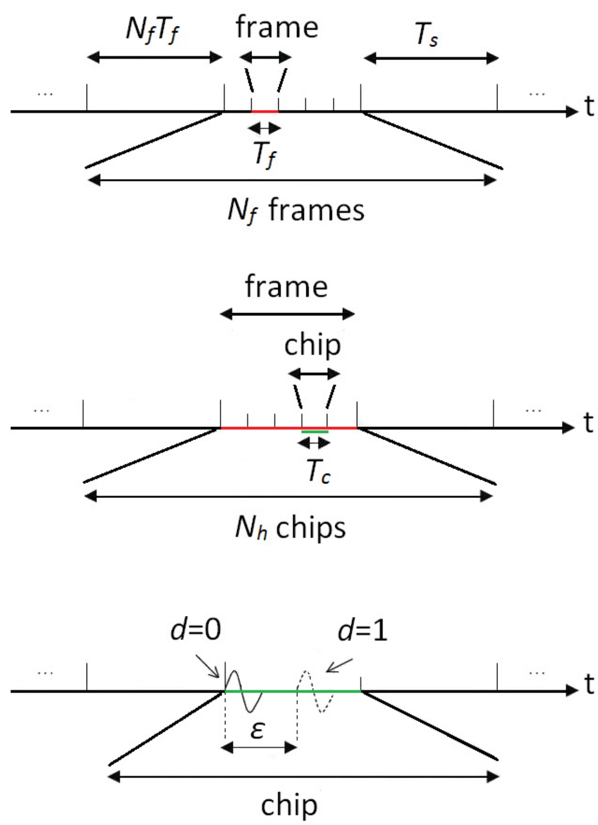

Figure 2.6 Representation of TH-PPM.

T_c is the chip time duration, where $N_hT_c \leq T_f$;

$c_j^{(k)}$ is the distinct time-hopping sequence, $0 \leq c_j^{(k)} < N_h$;

$d^{(k)} \in \{0, 1\}$ is the data symbol transmitted by the kth user. It introduces an additional shift ϵ to distinguish between pulses carrying bit 0 when $\epsilon = 0$ and bit 1 otherwise;

ϵ is the additional time shift of PPM.

For PPM, we assume constant unit signal amplitude for all users; that is, $A^{(k)} = 1$, in which case (2.6) becomes

$$s^{(k)}(t) = \sum_{k=1}^{N_u} \sum_{j=0}^{N_h-1} p(t - jT_f - c_j^{(k)}T_c - \epsilon d^{(k)}), \text{ where } t \in [0, T_s]. \tag{2.7}$$

The received signal $r(t)$ after transmission over the fading channel described by Eq. (2.3) is given by

$$r(t) = \sum_{\ell=1}^{L} \sum_{k=1}^{K} \left[a_{\ell k} \sum_{j=1}^{N_s} q(t - \tau_{\ell k} - jT_f - c_j^{(k)} T_c - \epsilon d^{(k)}) \right] + w(t). \tag{2.8}$$

where L is the number of resolvable paths, $q(t)$ is the received pulse waveform, $a_{\ell k}$ and $\tau_{\ell k}$ are the attenuation and delay of the ℓth path of kth user's channel, respectively. Finally, $w(t)$ is the Additive White Gaussian Noise (AWGN) with double-sided power spectral density $N_0/2$. When only a single active user is in the system, i.e., $K = 1$, the received signal can be simplified to

$$r(t) = \sum_{\ell=1}^{L} a_\ell \sum_{j=1}^{N_s} q(t - \tau_\ell - jT_f - c_j T_c - \epsilon d) + w(t). \tag{2.9}$$

where a_ℓ and τ_ℓ are the attenuation and delay of the ℓth path [35].

Chapter 3

Background: Processing of Sparse Signals and Channel Estimation

3.1 Compressed sensing

As discussed earlier, UWB communication systems use ultra-short pulses that spread the signal energy over a wide frequency spectrum. According to the Nyquist's theorem, the sampling rate must be at least twice the maximum frequency of a signal to avoid losing information due to aliasing. Therefore, due to the high bandwidth of the received UWB signal, its analog to digital conversion requires high-speed ADCs. However, high speed ADCs are expensive, consume a lot of power, and have low resolution. Even if sampling at the Nyquist rate was possible the resulting large number of samples would require expensive storage systems and processing methods. To achieve the required sampling rates and bit resolution, Compressed Sensing (CS) has been investigated for use in UWB systems. CS is a new signal acquisition technique where sampling is done at a rate considerably below the Nyquist rate. In CS, the compression and the sensing process are performed together, i.e., CS compresses the signal and samples it at a reduced rate. The critical assumption is that the signal is sparse, i.e., it can be written as a linear combination of a few basis vectors [37, 38, 39]. CS utilizes linear projections to project a high-dimensional sparse signal onto a lower-dimensional space preserving enough information for signal reconstruction.

In the following, we elaborate on CS in the context of discrete-time signal processing. A sparse discrete-time signal is a linear combination of a few vectors called atoms chosen from an over-complete dictionary. Formally, an over-complete dictionary is a collection of redundant atoms such that the number of atoms exceeds the dimension of the signal space, so that any signal does not have a unique representation as a linear combinations of atoms. Let $\mathbf{f} \in \mathbb{R}^N$ denote the discrete-time signal to be sampled. We assume that $\mathbf{f}$ is of the form

$$\mathbf{f} = \sum_{i=1}^{N} x_i \boldsymbol{\psi}_i, \tag{3.1}$$

where $\boldsymbol{\psi}_i \in \mathbb{R}^N$, $x_i \in \mathbb{R}$, $i = 1, \ldots, N$. The last equation can be written in matrix form as

$$\mathbf{f} = \mathbf{\Psi x}, \tag{3.2}$$

where

$$\mathbf{\Psi} = [\boldsymbol{\psi}_1, \boldsymbol{\psi}_2, \ldots, \boldsymbol{\psi}_N], \tag{3.3}$$

and

$$\mathbf{x} = [x_1, x_2, x_3, \ldots, x_N]^T. \tag{3.4}$$

We assume that only a few of the coefficients x_i are non-zero, i.e., that $\mathbf{f}$ is sparse on the columns of $\mathbf{\Psi}$. In CS, $\mathbf{f}$ is acquired by projecting it on K vectors $\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots, \boldsymbol{\phi}_K$ with $K \ll N$ to obtain K measurements $\mathbf{y} = [y_1, y_2, y_3, \ldots, y_K]$. By defining

$$\mathbf{\Phi} = [\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots, \boldsymbol{\phi}_K]^T, \tag{3.5}$$

this projection operation can be written in matrix form as

$$\mathbf{y} = \mathbf{\Phi f} = \mathbf{\Phi \Psi x} = \mathbf{A x}, \tag{3.6}$$

where $\mathbf{\Phi} \in \mathbb{R}^{K \times N}$ is the sensing (measurement) matrix, $\mathbf{y} \in \mathbb{R}^K$ the an observation vector $K \ll N$, and $\mathbf{A} = \mathbf{\Phi \Psi}$ is the effective measurement or holographic matrix.

3.1.1 Feasibility of signal recovery

We now focus on the problem of reconstructing the signal $\mathbf{f}$ from the observation vector $\mathbf{y}$. From Eq. 3.6 we see that this is equivalent to recovering $\mathbf{x}$ from $\mathbf{y}$ by solving the linear system $\mathbf{y} = \mathbf{Ax}$. Unfortunately, the system does not have a unique solution unless we take into account the fact that $\mathbf{x}$ is sparse [40].

Restricted isometry property

In recovering the sparse signal, the over-complete matrix $\mathbf{A}$ agrees to the condition known as Restricted Isometry Property (RIP) [41] introduced below.
Let $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_N] \in \mathbb{R}^{K \times N}$, with $K \ll N$. Then, $\mathbf{a}_i$, where $i = 1, \ldots, N$ atoms are normalized, that is, $\|\mathbf{a}_i\|_2 = 1$. Now, for integers $k = 1, 2, \ldots$, where k is sparsity of $\mathbf{x}$, i.e., $\|\mathbf{x}\|_0 = k$. Exact signal recovery can be guaranteed if matrix $\mathbf{A}$ obeys the RIP of order k, that is, there exists kth-restricted isometry constant $\delta_k \in (0, 1)$ such that

$$(1 - \delta_k)||\mathbf{x}||_2^2 \leq ||\mathbf{A}\mathbf{x}||_2^2 \leq (1 + \delta_k)||\mathbf{x}||_2^2. \tag{3.7}$$

If $\mathbf{A}$ has sufficiently small δ_k (e.g., $\delta_k < 1/3$), it approximately maintains ℓ_2 distances between k-sparse signals. Specifically, δ_k quantifies how close to isometrically $\mathbf{A}$ acts on k-sparse vectors [42].

Mutual incoherence

The measurement matrix $\mathbf{\Phi}$ must be incoherent with $\mathbf{\Psi}$ in order to guarantee signal recovery [43, 37, 44]. The measure of mutual incoherence of $\mathbf{\Phi}$ and $\mathbf{\Psi}$ is defined as follows

$$M(\mathbf{\Phi}, \mathbf{\Psi}) = \sup_{\boldsymbol{\phi} \in \mathbf{\Phi}, \boldsymbol{\psi} \in \mathbf{\Psi}} | \langle \boldsymbol{\phi}, \boldsymbol{\psi} \rangle |. \tag{3.8}$$

where the notation, $\boldsymbol{\phi} \in \mathbf{\Phi}$ ($\boldsymbol{\psi} \in \mathbf{\Psi}$) means that $\boldsymbol{\phi}$ ($\boldsymbol{\psi}$) that is a row (column) of $\mathbf{\Phi}$ ($\mathbf{\Psi}$). Therefore, when $\mathbf{\Phi}$ and $\mathbf{\Psi}$ have a very small value of M, that is, they are mutually incoherent and when signal $\mathbf{x}$ obeys the relation

$$\|x\|_0 < 1/2(1 + M^{-1}) \tag{3.9}$$

the k-sparse signal $\mathbf{x}$ can be recovered uniquely using convex optimization techniques [45, 46, 38, 39].

3.2 Sparse approximation with pursuit algorithms

There are two fundamental approaches to reconstruct the original signal from CS measurements: (1) convex optimization and (2) pursuit algorithms. If the measurement matrix obeys the RIP with sufficiently small δ_k and no measurement noise exists, it is possible to exactly recover signals from the measurement vector $\mathbf{y}$ using the convex optimization as follows.

To reconstruct $\mathbf{x}$, we look for the sparsest solution to the linear system $\mathbf{Ax} = \mathbf{y}$. Since the number of non-zero elements of $\mathbf{x}$ is equal to its ℓ_0-norm, $\|\mathbf{x}\|_0$, the signal reconstruction problem can be formulated as

$$\min\|\mathbf{x}\|_0 \text{ subject to } \mathbf{Ax} = \mathbf{y}. \tag{3.10}$$

The reconstruction problem can be rewritten by allowing an approximation error δ

$$\min\|\mathbf{x}\|_0 \text{ subject to } \|\mathbf{Ax} - \mathbf{y}\|_2 \leq \delta, \tag{3.11}$$

where $\delta > 0$ and $\|.\|_2$ denotes the ℓ_2-norm. Since exact solutions to the above problems can only be obtained by high computational complexity optimization methods, less complex methods called pursuit methods are used to solve (3.11) [47, 48].

3.2.1 Basis pursuit

The basis pursuit (BP) replaces the ℓ_0-norm condition in (3.10) by the ℓ_1-norm to be able to solve the problem through linear programming [49]. Eq. (3.10) can be written in terms of ℓ_1-norm minimization as

$$\min\|\mathbf{x}\|_1 \text{ subject to } \mathbf{Ax} = \mathbf{y}, \tag{3.12}$$

Considering the approximation errors $\delta > 0$, the problem can be posed as

$$\min\|\mathbf{x}\|_1 \text{ subject to } \|\mathbf{Ax} - \mathbf{y}\|_2 \leq \delta, \text{ where } \delta > 0. \tag{3.13}$$

Now, these problems are transformed into linear programming problems as

$$\min_{\mathbf{c}} \ \mathbf{c}^T\mathbf{x} \text{ subject to } \mathbf{A}\mathbf{x} = \mathbf{y}, \ \mathbf{x} \geq 0, \tag{3.14}$$

where $\mathbf{c}^T\mathbf{x}$ is the objective function, $\mathbf{A}\mathbf{x} = \mathbf{y}$ is a collection of equality constraints, and $\mathbf{x} \geq 0$ is a set of bounds. By dropping the condition $\mathbf{x} \geq 0$ and doing further manipulations, Eq. (3.14) can be simplified as an equivalent dual linear program in terms of the dual variables ν and e:

$$\max_{\mathbf{b}} \ \mathbf{y}^T\mathbf{b} \text{ subject to } \mathbf{A}^T\mathbf{b} - 2\nu = -e, \ 0 \leq \nu \leq e. \tag{3.15}$$

Now, Eq. (3.15) has more variables than constraints and that can be solved by standard optimization techniques. Although Eq. (3.14) is different from the original problems in (3.10) and (3.11), solving this problem results in approximate solutions to the ℓ_0-norm minimization [50, 51].

3.2.2 Matching pursuit

BP algorithms produce perfect reconstructions of $\mathbf{y}$ under the conditions described in the previous section but their computational cost can be still high. A popular alternative to BP are the Matching Pursuit (MP) algorithm [52] and its variant Orthogonal Matching Pursuit (OMP) [53, 54, 55] that we describe next.

MP is an iterative method to compute a linear expansion of $\mathbf{y}$ over a set of vectors selected from the dictionary matrix $\mathbf{A}$. Let $\mathbf{a}_i$, $1 \leq i \leq N$, denote the ith column of $\mathbf{A}$ [43, 37]. Each iteration in the MP and OMP algorithms consists of two steps: (1) an atom selection step and (2) a residual update step. The atom selection step is done as follows.

At the beginning of the first iteration, the residual or the approximation error $\mathbf{r}_0$ is equal with the given signal $\mathbf{y}$, i.e., $\mathbf{r}_0 = \mathbf{y}$. After the first iteration, the largest projection (or the highest correlation) of $\mathbf{r}_0$ that is in the direction of $\mathbf{a}_1$ (the first column of $\mathbf{A}$) is selected, as shown in Fig. 3.1. Next, in the residual update step, $\mathbf{r}_1$ is obtained from subtracting the correlated part in the direction of $\mathbf{a}_1$ from $\mathbf{r}_0$. As one can see in Fig. 3.1, in the next iteration, $\mathbf{r}_1$ is projected in the direction of $\mathbf{a}_0$ resulting in the residual $\mathbf{r}_2$. $\boldsymbol{\alpha}_n$ is the atom

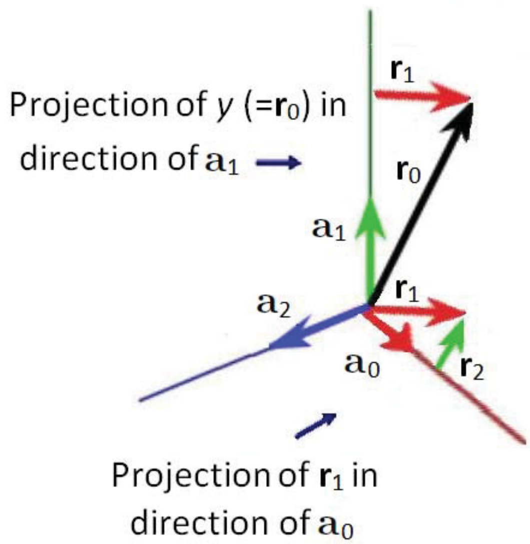

Figure 3.1 First and second iteration of decomposing $\mathbf{y}$ in a three element dictionary by MP.

selected in the nth iteration, $n = 1, 2, \ldots$, is shown as

$$\boldsymbol{\alpha}_n = \arg\max_{\mathbf{a}_i \in \mathbf{A}} |\langle \mathbf{r}_{n-1}, \mathbf{a}_i \rangle|, \tag{3.16}$$

where $\langle \mathbf{r}_{n-1}, \mathbf{a}_i \rangle$ represents the weight of the selected atom $\boldsymbol{\alpha}_n$, and $\mathbf{r}_{n-1}$ denotes the residual at the $(n-1)$th iteration. Then, the algorithm updates the residual as

$$\mathbf{r}_n = \mathbf{r}_{n-1} - c_n \boldsymbol{\alpha}_n. \tag{3.17}$$

where the constant c_n denotes the weight of the selected atom $\boldsymbol{\alpha}_n$. The algorithm stops if the norm of the residual goes below a desired approximation error threshold, or when the number of distinct atoms in the approximation reaches a desired limit. Finally, the selected atoms weighted by their respective correlation values are combined to give the final sparse approximation [56]. The approximation at the nth iteration is given as

$$\mathbf{y}_n = \sum_{k=1}^{n} c_k \boldsymbol{\alpha}_k = [\boldsymbol{\alpha}_1\ \boldsymbol{\alpha}_2\ \ldots\ \boldsymbol{\alpha}_n][c_1\ c_2\ \ldots\ c_n]^T. \tag{3.18}$$

3.2.3 Orthogonal matching pursuit

OMP does not reselect atoms during the process of signal reconstruction and instead it updates the coefficients of all the previously selected atoms so that the residual at each iteration is minimized. Thus, the last derived residual is orthogonal to all the atoms selected at previous iterations and not only to the last selected one. This way, a faster convergence for OMP is possible compared to MP. At the first iteration, similar to MP, the residual $\mathbf{r}_0$ is equal with the given signal $\mathbf{y}$. The nth iteration the algorithm computes

$$\boldsymbol{\alpha}_n = \arg \max_{\mathbf{a}_i \in (\mathbf{A}/\mathbf{A}_{n-1})} |\langle \mathbf{r}_{n-1}, \mathbf{a} \rangle|, \tag{3.19}$$

where $\mathbf{r}_{n-1}$ denotes the residual at the $(n-1)$th iteration, $\mathbf{A}_{n-1}$ denotes the set of atoms selected up to the $(n-1)$th iteration, and the operator "/" denotes the set difference operator. The approximation at the nth iteration is given as the projection of the original signal vector onto the subspace spanned by the selected atoms, as shown below

$$\mathbf{y}_n = \mathbf{A}_n (\mathbf{A}_n^T \mathbf{A}_n)^{-1} \mathbf{A}_n^T \mathbf{y} = \mathbf{A}_n \mathbf{c}_n, \tag{3.20}$$

where $\mathbf{A}_n = [\mathbf{A}_{n-1}, \boldsymbol{\alpha}_n]$ and $\mathbf{A}_{n-1}$ is the matrix of atoms selected up to the $(n-1)$th iteration. Also, $\mathbf{c}_n$ stands for the coefficient vector at the nth iteration. Next, in the updating step, the algorithm updates the approximation error as

$$\mathbf{r}_n = \mathbf{y} - \mathbf{A}_n \mathbf{c}_n. \tag{3.21}$$

The algorithm does not proceed to the next iteration if the norm of the residual falls below the desired residual threshold. Also, if the desired limit for the number of approximated atoms is reached, that is, iteration number is equal maximum allowed sparsity, $n = k$. Moreover, because the selected atoms by any iteration are always linearly independent, the matrix inverse operations in above computations are valid. Since the selected atoms are all distinct, the nonzero components of the approximated sparse vector are equal to the elements of the coefficient vector at the last iteration.

3.2.4 Matching pursuit versus orthogonal matching pursuit

In the MP algorithm, reselection of atoms of the dictionary is possible. At any iteration step in MP, the newly obtained residual is orthogonal only to the last selected atom and it may not be orthogonal to all the atoms selected at the previous steps. As a result, some atoms selected at an earlier iteration may get reselected, resulting in a slowdown of the convergence speed. However, since MP algorithm yields an approximation error which decreases with each iteration, the algorithm is guaranteed to converge. MP can initially select an atom that is not part of the optimal sparse representation; in many of the subsequent iterations the atoms selected by MP only try to compensate for the poor initial selection [57, 58].

On the other hand, in OMP atom reselection is avoided by projecting the signal vector on to the subspace spanned by all selected atoms. In other words, by updating the coefficients of all previously selected atoms, the residual at each iteration is orthogonal to not only the last selected atom, but also to all the atoms selected at previous iterations and so that atom reselection is prevented resulting in converging in fewer iterations. Although OMP converges in fewer iterations compare to MP, the required computational effort still depends on the choice of the dictionary. In short, MP achieves an accurate decomposition of the signal with a low complexity alternative to BP, but needs an unbounded number of iterations for convergence. OMP converges in a fixed number of iterations but requires the added complexity of the orthogonalization at each step.

3.3 Channel estimation using adaptive filtering

As explained earlier, the transmitted signal undergoes multipath fading before it reaches the receiver. The process of reversing the channel's effect and recovering the transmitted signal is called equalization. Successful equalization relies heavily on the availability of accurate channel estimates. Channel estimation is, therefore, a critical part of a communication system's receiver structure [59, 60, 61]; given the transmitted signal and the channel output, we want to reconstruct the channel. This can be done using gradient descent adaptive filtering [62, 63] that employs a gradient descent optimization procedure that attempts to closely approximate a given system, as shown in Fig. 3.2. The received signal (channel

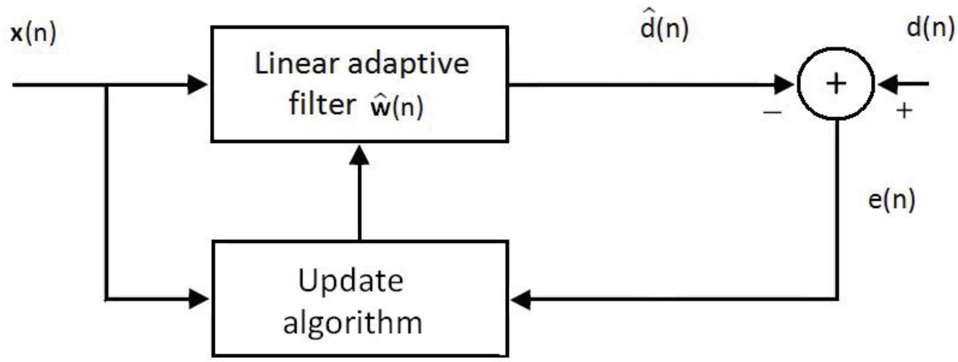

Figure 3.2 Block diagram of a gradient descent adaptive process.

output) $x(n)$ is the sum of the channel output due to the transmitted signal or desired response signal $d(n)$ and the additive noise $v(n)$. For simplicity in the system structure, we assume that the channel and linear adaptive filter are both Finite Impulse Response (FIR). The coefficients for a filter of order $N-1$ are defined as

$$\mathbf{w}(n) = [w_0(n), w_1(n), \ldots, w_{N-1}(n)]^T, \tag{3.22}$$

where $n = 0, 1, 2, \ldots$ is the iteration index, and

$$\hat{d}(n) = \sum_{i=0}^{N-1} w_i(n)x(n-i) = \mathbf{w}(n)^T\mathbf{x}(n). \tag{3.23}$$

where an input signal vector is

$$\mathbf{x}(n) = [x(n), x(n-1), \ldots, x(n-N+1)]^T. \tag{3.24}$$

The goal is to find the coefficients of $\mathbf{w}(n)$ such that its output, $\hat{d}(n)$, is closest to $d(n)$. In other words, we want to find a filter such that the estimation error is

$$e(n) = d(n) - \hat{d}(n). \tag{3.25}$$

reaches the smallest value [62, 64].

3.3.1 Implementation of the steepest descent algorithm

The method of steepest descent (SD) [65, 66] is an iterative procedure that aims at locating the minimum point of a smooth surface that in our case is the cost function $J(\mathbf{w})$ that is the MSE between the filter output and the desired response:

$$J(\mathbf{w}) = E[(d(n) - \mathbf{w}^T\mathbf{x}(n))^2]. \tag{3.26}$$

where $\mathbf{w}$ is the filter tap weight vector and $E[\cdot]$ denotes statistical expectation. The location of the lowest point, $\mathbf{w}_{opt}$, defines the optimum values for the tap weight vector. It can be shown that if the cross-correlation vector of the desired response signal and the channel output $\mathbf{x}(n)$ is given by

$$\mathbf{p}_{d\mathbf{x}} = E[d(n)\mathbf{x}(n)], \tag{3.27}$$

and the autocorrelation matrix of the input vector is given by

$$\mathbf{R}_{\mathbf{xx}} = E[\mathbf{x}(n)\mathbf{x}(n)^T], \tag{3.28}$$

and both are constant (do not change with n) then $J(\mathbf{w})$ is minimized at

$$\mathbf{w}_{opt} = \mathbf{R}_{\mathbf{xx}}^{-1}\mathbf{p}_{d\mathbf{x}}, \tag{3.29}$$

where $\mathbf{R}_{\mathbf{xx}}^{-1}$ is the inverse of the autocorrelation matrix of the input vector. We will provide a proof of this later. The SD algorithm converges at the steady-state to a minimum point, called the Minimum Mean Square Error (MMSE) [62] solution, on the MSE surface. The starting point $\mathbf{w}(0)$ can be arbitrarily chosen to any value. The aim is to iteratively descend to the minimum point of the cost function surface. This iterative procedure is graphically explained in Fig. 3.3, which shows a MSE cost function for a single-tap FIR filter. Clearly, at the optimum point the tangential slope of the cost function is zero and the further from this point we go the larger the slope magnitude becomes. This is a direct consequence of the fact that the MSE error is a quadratic, i.e., a convex function with respect to the filter coefficients. Also, the sign of the cost function to the right and left of the optimal point is positive and negative, respectively. The SD algorithm makes large (small) adjustments of

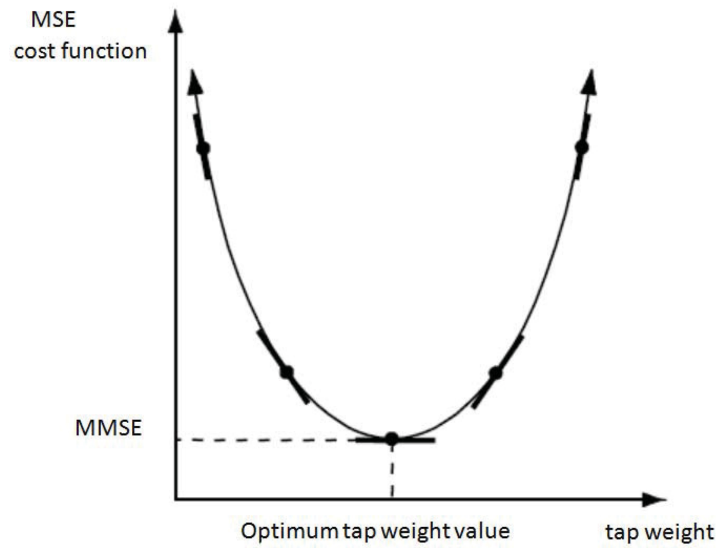

Figure 3.3 MSE cost function for a single-tap FIR filter.

the parameter value when its value is far from (close) to the optimum value. Considering the MSE cost function, the updating coefficients algorithm becomes

$$\partial \mathbf{w}(n+1) = \partial \mathbf{w}(n) - \mu/2 \frac{\partial E[e^2(n)]}{\partial \mathbf{w}}, \tag{3.30}$$

where the factor 1/2 is used for notational convenience, and the step size μ controls the convergence rate and stability of the algorithm. $e(n)$ is the error at the current sample n and $E[e^2(n)]$ is the mean square error. To implement the SD algorithm, we have to evaluate the partial derivatives of the cost function with respect to the coefficient values. We have

$$\begin{aligned} \frac{\partial E[e^2(n)]}{\partial \mathbf{w}} &= E[\frac{\partial e^2(n)}{\partial \mathbf{w}}], \\ &= E[2e(n)\frac{\partial e(n)}{\partial \mathbf{w}}] \\ &= E[2e(n)\frac{\partial (d(n) - \mathbf{w}^T\mathbf{x}(n))}{\partial \mathbf{w}}] \\ &= -2E[e(n)\mathbf{x}(n)]. \end{aligned} \tag{3.31}$$

Thus, the SD iteration can be written as

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu E[e(n)\mathbf{x}(n)]. \tag{3.32}$$

The expectation term $E[e(n)\mathbf{x}(n)]$ is given by

$$E[e(n)\mathbf{x}(n)] = E[\mathbf{x}(n)(d(n) - \hat{d}(n))], \tag{3.33}$$

Then, we have

$$\begin{aligned} E[e(n)\mathbf{x}(n)] &= E[d(n)\mathbf{x}(n)] - E[\mathbf{x}(n)\mathbf{x}(n)^T\mathbf{w}(n)], \\ &= \mathbf{p}_{d\mathbf{x}} - \mathbf{R}_{\mathbf{xx}}\mathbf{w}(n). \end{aligned} \tag{3.34}$$

Therefore, the updating rule becomes

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu(\mathbf{p}_{d\mathbf{x}} - \mathbf{R}_{\mathbf{xx}}\mathbf{w}(n)). \tag{3.35}$$

3.3.2 Steady-State Properties of the algorithm

Assume that the autocorrelation matrix and cross-correlation vector are constant over time, i.e.,

$$\mathbf{R}_{\mathbf{xx}}(n) = \mathbf{R}_{\mathbf{xx}}, \tag{3.36}$$

and

$$\mathbf{p}_{d\mathbf{x}}(n) = \mathbf{p}_{d\mathbf{x}}. \tag{3.37}$$

The SD update can be written as

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu(\mathbf{p}_{d\mathbf{x}} - \mathbf{R}_{\mathbf{xx}}\mathbf{w}). \tag{3.38}$$

Let $\mathbf{w}_{ss}$ be the coefficient values of $\mathbf{w}(n)$ in the steady-state. In the steady-state, the gradient of the cost function, $\mathbf{p}_{d\mathbf{x}} - \mathbf{R}_{\mathbf{xx}}\mathbf{w}_{ss}$ is zero, that is

$$\mathbf{p}_{d\mathbf{x}} - \mathbf{R}_{\mathbf{xx}}\mathbf{w}_{ss} = 0, \tag{3.39}$$

so, $\mathbf{w}(n+1)$ becomes

$$\mathbf{w}(n+1) = \mathbf{w}(n) = \mathbf{w}_{ss}. \tag{3.40}$$

and the system converges. To find the stationary point, we can solve the N system of linear equations defined by Eq. (3.38):

$$\mathbf{R}_{\mathbf{xx}}\mathbf{w}_{ss} = \mathbf{p}_{d\mathbf{x}}. \tag{3.41}$$

If the inverse of the autocorrelation matrix exists, the optimal solution for the MMSE estimation is

$$\mathbf{w}_{ss} = \mathbf{R}_{\mathbf{xx}}^{-1}\mathbf{p}_{d\mathbf{x}} = \mathbf{w}_{opt}. \tag{3.42}$$

3.3.3 Stochastic gradient based adaptive filter Least-Mean-Square algorithm

The SD procedure utilizes the statistics of the input and desired response signals and not the actual measured signals. In practice, the input signal statistics, $\mathbf{R}_{\mathbf{xx}}$ and $\mathbf{p}_{d\mathbf{x}}$, are not known a priori, instead, we have access to pairs of input $\mathbf{x}(n)$ and corresponding outputs $d(n)$. Therefore, the SD estimation procedure is not applicable in most practical situations. Hence, a practical adaptive approach called the stochastic gradient adaptive algorithm is introduced. The least-mean-square (LMS) [67, 62, 62] adaptive filter is the stochastic gradient version of the method of SD. Computational simplicity and robust adaptation properties have made the LMS the most popular algorithm for FIR filters. This algorithm is an approximate implementation of the SD procedure in which an instantaneous estimate of $e(n)\mathbf{x}(n)$ is used instead of the expected value $E[e(n)\mathbf{x}(n)]$, as below. Here, the vector $e(n)\mathbf{x}(n)$ is an approximation of the true gradient of the squared error function $e^2(n)/2$ with respect to the coefficient vector $\mathbf{w}(n)$:

$$\begin{aligned}\frac{\partial e^2(n)}{\partial \mathbf{w}(n)} &= (1/2)2e(n)\frac{\partial(d(n) - \mathbf{w}(n)^T\mathbf{x}(n))}{\partial \mathbf{w}(n)}, \\ &= -e(n)\mathbf{x}(n).\end{aligned} \tag{3.43}$$

The LMS recursion is obtained by using Eq. (3.43) in place of the true gradient in the steepest descent recursion, that is,

$$\mathbf{w}(n+1) = \partial\mathbf{w}(n) - \mu\frac{\partial e^2(n)}{\partial \mathbf{w}(n)}, \tag{3.44}$$

becomes

$$\mathbf{w}(n+1) = \partial \mathbf{w}(n) + \mu e(n)\mathbf{x}(n). \tag{3.45}$$

Denote $\mathbf{R}$ as the covariance matrix of the input vector $\mathbf{x}(n)$ and λ_{max} as its maximum eigenvalue. Then, the convergence condition for LMS is

$$0 < \mu \leq \frac{1}{\lambda_{max}}. \tag{3.46}$$

A value of μ in the above range guarantees that LMS is convergent in mean sense [68]. In practice, the adaptive filter coefficients fluctuate around the optimal filter coefficients and are never equal to them. The excess mean square error is the difference between the MSE introduced by the adaptive filter and the minimum MSE produced by corresponding filter [62]. The steady-state excess MSE of LMS is

$$P_{ex}(\infty) = \frac{\eta}{2-\eta} P_0, \tag{3.47}$$

where P_0 is the power of the observation noise $P_0 = E[\nu^2(n)]$ and $\eta = tr(\mathbf{R}(\mathbf{I} - \mu \mathbf{R}))$.

3.3.4 Zero attractor LMS algorithm

From Eq. (3.45), we can find that the LMS-based channel estimation method never takes advantage of a possible sparse structure of the filter. We can exploit the channel sparsity to improve performance by introducing ℓ_p-norm penalties $(0 \leq p \leq 1)$ to the LMS-based cost function. The Zero attractor LMS (ZA-LMS) [69] cost function denoted as $J_{ZA}(n)$ is constructed by merging the instantaneous square error with the ℓ_1-norm penalty of the coefficient vector

$$J_{ZA}(n) = (1/2)e^2(n) + \gamma_{ZA} \|\mathbf{w}(n)\|_1, \tag{3.48}$$

where γ_{ZA} is a regularization parameter which balances the adaptive estimation error and sparseness penalty of $\mathbf{w}(n)$. The corresponding update equation of ZA-LMS is

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \mu \frac{\partial J_{ZA}(n)}{\partial \mathbf{w}(n)}. \tag{3.49}$$

Then, by substituting for $J_{ZA}(n)$ we have

$$\begin{aligned}\mathbf{w}(n+1) =& \mathbf{w}(n) - \mu \frac{\partial((1/2)e^2(n) + \gamma_{ZA}\|\mathbf{w}(n)\|_1)}{\partial \mathbf{w}(n)} \\ &= \mathbf{w}(n) - \rho_{ZA}\mathrm{sgn}(\mathbf{w}(n)) + \mu e(n)\mathbf{x}(n),\end{aligned} \tag{3.50}$$

where $\rho_{ZA} = \mu\gamma_{ZA}$ and $\mathrm{sgn}(\cdot)$ is the component-wise sign function. The main difference between the ZA-LMS update in Eq. (3.50) and the standard LMS update in Eq. (3.45) is the additional term

$$-\rho_{ZA}\ \mathrm{sgn}(\mathbf{w}(n)). \tag{3.51}$$

This term attracts the small filter coefficients towards zero, which accelerates convergence when a lot of the channel coefficients in $\mathbf{w}(n)$ are zeros. The strength of this zero attractor is controlled by ρ_{ZA}. The convergence conditions of the ZA-LMS are presented in the following theorems:

Theorem 1. *([69]) Assume that μ satisfies Eq. (3.46). Then, the mean coefficient vector $E[\boldsymbol{w}(n)]$ converges to a vector $\boldsymbol{w}(\infty)$ such that*

$$E[\boldsymbol{w}(\infty)] = \boldsymbol{w} - \frac{\rho_{ZA}}{\mu}\boldsymbol{R}^{-1}E[sgn(\boldsymbol{w}(\infty))]. \tag{3.52}$$

□

According to Eq. (3.52) the ZA-LMS filter gives a biased estimate of the true coefficient vector. However, for sparse systems choosing a suitable ρ_{ZA} results in lower MSE for the ZA-LMS compare to the standard LMS.

Theorem 2. *([69]) Let the index set of non-zero taps be denoted by $\mathbb{NZ}$, that is, $w_i \neq 0$ when $i \in \mathbb{NZ}$. Assuming ρ_{ZA} is sufficiently small so that for every $i \in \mathbb{NZ}$ in steady-state we have*

$$E[sgn(w_i(\infty))] = sgn(w_i), \tag{3.53}$$

The steady-state excess MSE of the ZA-LMS is

$$P_{ex}(\infty) = \frac{\eta}{2-\eta}P_0 + \frac{\alpha_1}{(2-\eta)\mu}\rho_{ZA}(\rho_{ZA} - \frac{2\ \alpha_2}{\alpha_1}), \tag{3.54}$$

where α_1 and α_2 are given by

$$\alpha_1 = E[sgn(\boldsymbol{w}(\infty)^T)(\boldsymbol{I} - \mu\boldsymbol{R})^{-1}\ sgn(\boldsymbol{w}(\infty))] \in (0,\ \frac{N}{1-\mu\lambda_{max}}) \tag{3.55}$$

and

$$\alpha_2 = E[\|\boldsymbol{w}(\infty)\|_1] - \|\boldsymbol{w}(n)\|_1. \tag{3.56}$$

□

The first term on the RHS of Eq. (3.54) is the excess MSE of the standard LMS in Eq. (3.47). So, the excess MSE of ZA-LMS in Eq. (3.54) is lower than that of the standard LMS when $\alpha_2 > 0$, or, equivalently, when ρ_{ZA} lies between 0 and $2\alpha_2/\alpha_1$. To specify α_2 further, we have the following result.

Lemma 1. *([69]) Let $\mathbb{Z}$ and $\mathbb{NZ}$ be the index sets of the zero and non-zero taps, respectively. A first-order approximation of α_2 while $\mathbf{w}(n)$ is assumed to be Gaussian distributed, is given by*

$$\alpha_2 \simeq \sum_{i\in\mathbb{Z}} \sqrt{\frac{2}{\pi}\boldsymbol{\Phi}_{ii}(\infty)} - \frac{\rho_{ZA}}{\mu}\sum_{i\in\mathbb{NZ}} |b_i|, \tag{3.57}$$

where $\Phi_{ii}(\infty)$ and b_i are the ith element of the diagonal of $\boldsymbol{\Phi}(\infty)$ and $\boldsymbol{b}$, respectively, described as follows:

$$\boldsymbol{\Phi}(\infty) = E[(\boldsymbol{w}(\infty) - \boldsymbol{w}(n))(\boldsymbol{w}(\infty) - \boldsymbol{w}(n))^T], \tag{3.58}$$

$$\boldsymbol{b} = \boldsymbol{R}^{-1}E[sgn(\boldsymbol{w}(\infty))]. \tag{3.59}$$

□

The first term in the RHS of Eq. (3.57) is related to the taps associated with zero coefficients of $\mathbf{w}(n)$ and its value always varies about zero. The second term is a bias which is related to the shrinkage of the taps associated with non-zero coefficients of $\mathbf{w}(n)$. When the number of taps with zero coefficients is larger than the number of taps with non-zero coefficients, the first term of Eq. (3.57) dominates the second one and positive α_2 can be obtained.

3.3.5 Reweighted zero attractor LMS algorithm

To obtain a better performance gain compared to LMS and a wider margin for choosing ρ_{ZA}, a large α_2 is essential for the ZA-LMS. However, the bias term in Eq. (3.57) reduces α_2 and the MSE performance deteriorates. The reason of this behaviour is that in the ZA-LMS the shrinkage does not distinguish between zero taps and non-zero taps. Therefore, for less sparse systems, the ZA-LMS performance worsens because all the channel taps are forced to zero uniformly. In order to refine the zero attractor behavior, RZA-LMS is introduced in [69] that is motivated by Reweighting in CS [70]. The RZA-LMS is derived using the cost function $J_{RZA}(n)$:

$$J_{RZA}(n) = \frac{1}{2}e^2(n) + \gamma' \sum_{i=1}^{N} \log(1 + |w_i|/\epsilon'). \tag{3.60}$$

where $\gamma' > 0$ is a regularization parameter which trades off the estimation error and channel sparsity and ϵ' is the threshold. The log-sum penalty $\sum_{i=1}^{N} \log(1 + |w_i|/\epsilon')$ acts more similarly to the ℓ_0-norm than ℓ_1. Let

$$\rho_{RZA} = \mu\gamma'/\epsilon', \tag{3.61}$$

and

$$\epsilon = 1/\epsilon'. \tag{3.62}$$

then, the coefficient vector update equation is

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \rho_{RZA}\frac{\mathrm{sgn}(\mathbf{w}(n))}{1 + \epsilon|\mathbf{w}(n)|} + \mu e(n)\mathbf{x}(n). \tag{3.63}$$

RZA-LMS selectively shrinks taps with small magnitudes comparable to $1/\epsilon$; and there is shrinkage applied on the taps with coefficient weight $|w_i(n)| \gg 1/\epsilon$, $i = 1, \ldots, N$. Consequently, the bias of the RZA-LMS can be reduced [69].

The performance of the ZA-LMS (Eq. (3.50)) and the RZA-LMS (Eq. (3.63)) are compared with that of the standard LMS (Eq. (3.45)) in the figures below. An experiment is designed to demonstrate their tracking and steady-state performances. In Fig. 3.4, there are 16 coefficients in the system and the 5th tap is set to value 1 and other taps to zero, i.e., a sparsity of 1/16. One can see from the results, when the filter is very sparse, both the ZA-LMS and the RZA-LMS result in a faster convergence rate and better steady-state performances than the standard LMS. In addition, the RZA-LMS achieves lower MSE than the ZA-LMS. The LMS approaches towards the optimal filter weights by updating the filter weights in an ascending/descending manner down the MSE versus filter weight curve, as seen in Fig. 3.3.

In Fig. 3.5, all the odd taps of adaptive filter are set to 1 while all the even taps are zero, i.e., a sparsity of 8/16. It can be seen that as the number of non-zero taps increases, the performance of the ZA-LMS gets worst while the RZA-LMS maintains the best performance among the three filters. In Fig. 3.6, all the 16 taps are set to value -1 while all the odd taps are maintained to be 1, resulting in a completely non-sparse filter. Here, the adaptive filter is non-sparse and the RZA-LMS performs comparably with the standard LMS. Moreover, the performance of the ZA-LMS deteriorates compared to standard LMS and RZA-LMS in both convergence rates and steady-state behaviors. In all three experiments, the input signal and the observed noise are white Gaussian random sequences with variance of 1 and 0.001, respectively. The parameters are set as $\mu = 0.05$, $\rho_{ZA} = \rho_{RZA} = 0.0005$, and $\epsilon = 10$. The three filters are run 200 times to obtain an average estimate of MSE.

Effects of the step size μ

The choice of the step size μ considerably affects how the adaptive filtering system will behave in terms of convergence. A small value of the step size ensures convergence, however, the trade off is slowing down the response time of the system. If the value of the step size is too small, the adaptive filter will not adapt fast enough because of the excessive number of iterations needed for reaching the vicinity of the minimum point on the error surface. On the other hand, in the case of μ being too large, the adaptive filter may not converge; even if it does, the behavior of the coefficient vector will be erratic. Therefore, the value μ should be chosen considering the trade off between tracking speed and convergence.

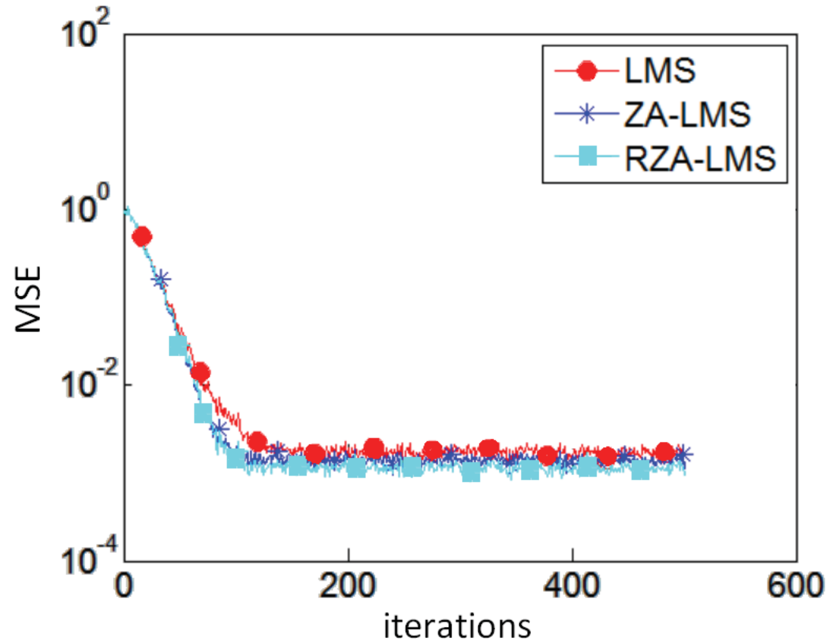

Figure 3.4 MSE of LMS vs ZA-LMS and RZA-LMS, considering sparse filters. There are 16 coefficients in the system and the 5th tap is set to value 1 and other taps to zero, i.e., a sparsity of 1/16.

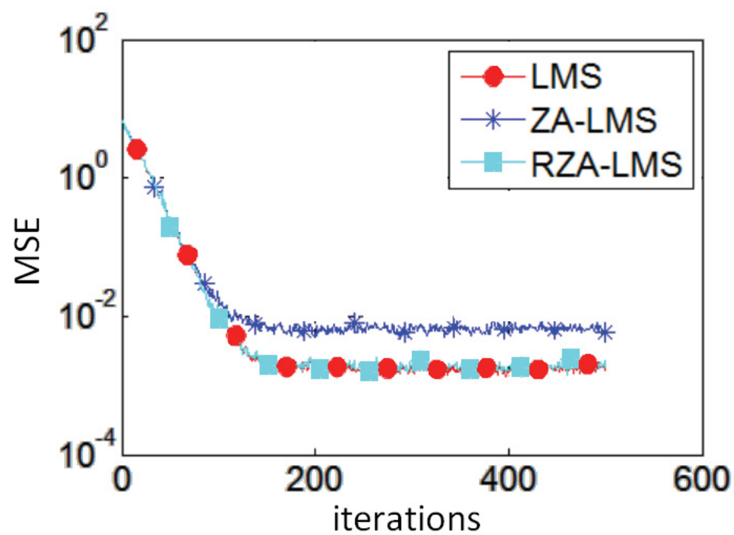

Figure 3.5 MSE of LMS vs ZA-LMS and RZA-LMS, considering less sparse filters. All the odd taps of adaptive filter are set to 1 while all the even taps are zero, i.e., a sparsity of 8/16.

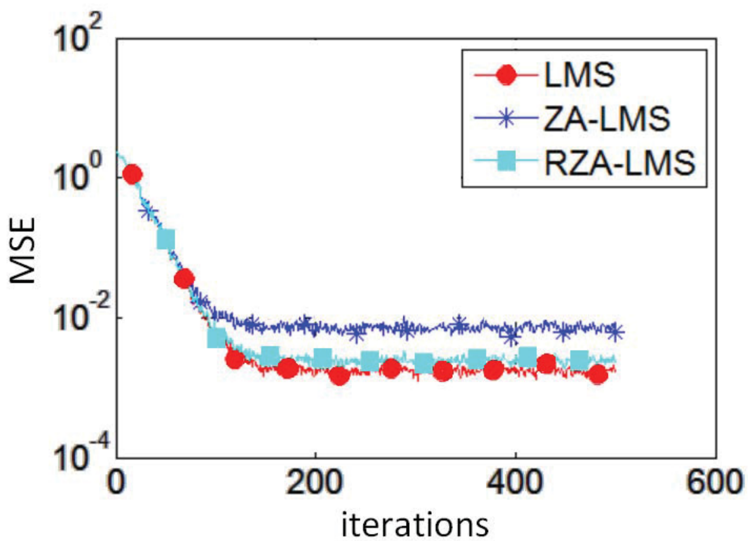

Figure 3.6 MSE of LMS vs ZA-LMS and RZA-LMS, considering non-sparse filters. All the 16 taps are set to value −1 while all the odd taps are maintained to be 1, resulting in a completely non-sparse filter.

Chapter 4

Estimation of highly sparse channels

Channel estimation is a long standing problem in communications and signal processing. The current study aims at developing new techniques for the estimation of a long and sparse indoor multi-path channel whose input is an IR-UWB PPM signal. According to the experimental studies in [7, 71, 72, 73, 70], when the impulse response is very long, the convergence time and MSE of traditional least-squares estimate methods can deteriorate substantially. Similarly, channel estimation using MP-type algorithms can be impractical when there is high interference from dense clusters within a multi-path sparse channel [17]. In this study the LMS, ZA-LMS and, RZA-LMS algorithms will be used to impose a sparse structure on the multi-path channel estimate. Moreover, the LMS, ZA-LMS and RZA-LMS algorithms will be modified by incorporating a novel sectional LMS method that improves the convergence rate through section-by-section probing and estimation of the channel.

4.1 Sectional LMS-based channel estimation

We consider a PPM input signal denoted as $\mathbf{x}(n)$ that goes through a very sparse multi-path channel $\mathbf{w}(n)$ of size J, as shown in Fig. 4.1. At the receiver side, we observe the noisy desired output of the channel, $d(n)$. Now, the aim is to estimate the unknown multi-path channel knowing the input. In order to do this, in place of the adaptive filter in Fig. 4.1, we propose to use the Sectional-LMS-type (S-LMS) adaptive filter. This method helps in estimating the very sparse channel section by section to prevent losing any information resulting from applying the all-zero parts of the PPM input signal to non-zero elements of the channel and vice versa. Besides, the selection of initial vector for each set of S-LMS-

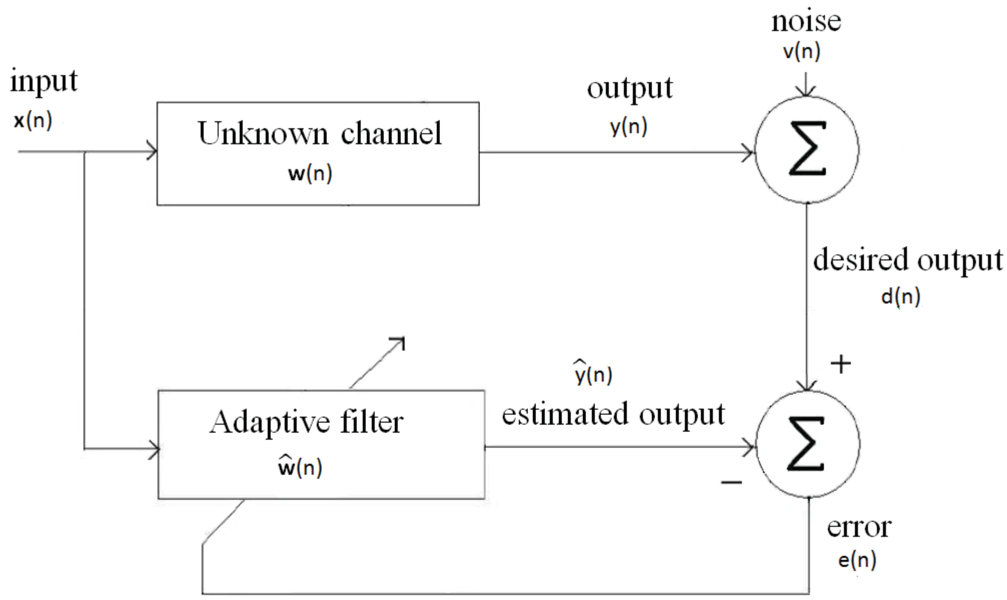

Figure 4.1 Block diagram of S-LMS algorithm

type iterations will be very accurate; that is, each iteration uses the latest estimation of the channel as the part of its initial vector for estimating the next section of the channel. The input signal $\mathbf{x}(n)$, when passed through the channel $\mathbf{w}(n)$ gives the output $y(n)$:

$$y(n) = \mathbf{w}^T(n)\mathbf{x}(n). \tag{4.1}$$

As shown in the block diagram of Fig. 4.1, $\hat{y}(n)$ is the output of the adaptive filter, $\hat{\mathbf{w}}(n)$, with input $\mathbf{x}(n)$. The adaptive filter output is given by:

$$\hat{y}(n) = \hat{\mathbf{w}}^T(n)\mathbf{x}(n). \tag{4.2}$$

The adaptive filter attempts to mimic the unknown multipath channel by adapting its coefficients so that the MSE between its output, $\hat{y}(n)$, and $d(n)$ is minimized. The signal $d(n)$ that the filter tracks, is obtained from the addition of the unknown sparse channel output $y(n)$ and Additive White Gaussian Noise (AWGN) $v(n)$ that are both not directly observable signals, that is,

$$d(n) = y(n) + v(n). \tag{4.3}$$

Our hope is that, when the filter produces outputs that are close to the observed channel outputs then the filter coefficients will also be close to the channel tap weights.

As shown in Fig. 4.2, $\mathbf{x}(n)$ is passed through a filter $\hat{\mathbf{w}}_1(n)$ (i.e., $k = 1$, where $k = 1, \ldots, K$) of size N that is initialized to an all-zero vector, i.e., $\mathbf{w}_1(0) = [0, 0, \ldots, 0]^T$. Then, the estimate of the first section of channel called $\mathbf{w}_1(n)$ of size N, is obtained recursively through the following LMS-type update equation:

$$\hat{\mathbf{w}}_1(n+1) = \hat{\mathbf{w}}_1(n) + \mu e_1(n)\mathbf{x}(n), \tag{4.4}$$

where $\hat{y}_1(n)$ is the filter output due to the input $\mathbf{x}(n)$ passing through the filter $\hat{\mathbf{w}}_1(n)$, that is,

$$\hat{y}_1(n) = \hat{\mathbf{w}}_1^T(n)\mathbf{x}(n), \tag{4.5}$$

and its corresponding error is given by

$$e_1(n) = d(n) - \hat{y}_1(n), \tag{4.6}$$

where $e_1(n)$ is the error related to the input $\mathbf{x}(n)$ passed through $\mathbf{w}_1(n)$.

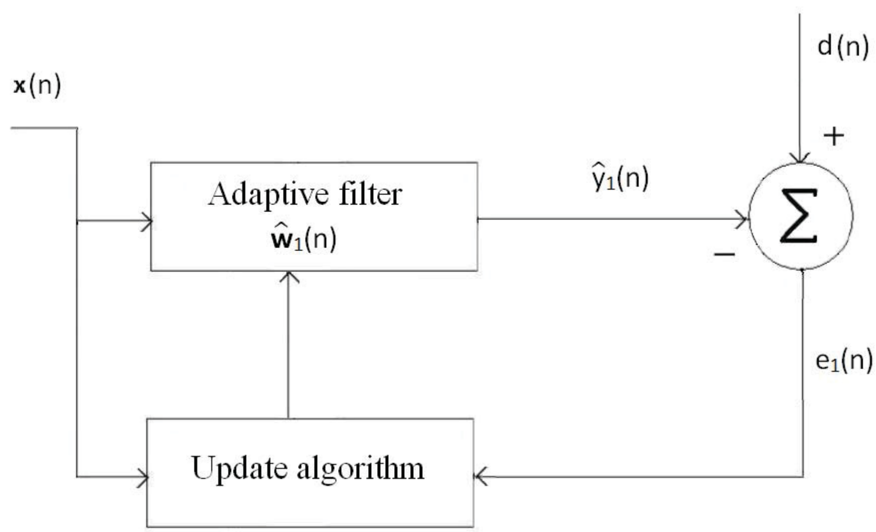

Figure 4.2 The adaptive filter structure

At the second set of S-LMS iterations, we are going to estimate $\mathbf{w}_2(n)$ (of size $2N$)

that is the first section of the sparse channel of size $2N$. Therefore, the size of the vector $\hat{\mathbf{w}}_2(n)$ entering the second set of S-LMS is set to $2N$ that is the size of the last updated estimate of the adaptive filter $\hat{\mathbf{w}}_1(n+1)$ appended with N zeros. The filter output $\hat{y}_2(n)$ is the output of $\mathbf{x}(n)$ passed through the filter $\hat{\mathbf{w}}_2(n)$, i.e.,

$$\hat{y}_2(n) = \hat{\mathbf{w}}_2^T(n)\mathbf{x}(n), \tag{4.7}$$

and its corresponding error is given by

$$e_2(n) = d(n) - \hat{y}_2(n), \tag{4.8}$$

where $e_2(n)$ is the error related to the input $\mathbf{x}(n)$ passed through the first section of the channel of size $2N$, $\mathbf{w}_2(n)$. Moreover, the updating equation for the channel estimates of the first k sections of channel of size kN is given by:

$$\hat{\mathbf{w}}_k(n+1) = \hat{\mathbf{w}}_k(n) + \mu e_k(n)\mathbf{x}(n). \tag{4.9}$$

The sectional channel estimation continues until the Kth set of S-LMS iterations when the entire size of the channel ($KN = J$) is reached.

4.1.1 Zero-attracting S-LMS type algorithms

Sectional-ZA-LMS and Sectional-RZA-LMS are obtained by similar modifications for S-LMS, as was mentioned in Section 4.1. Accordingly, the filter update equation for S-ZA-LMS and S-RZA-LMS are given by:

$$\hat{\mathbf{w}}_k(n+1) = \hat{\mathbf{w}}_k(n) - \rho_{SZA}\,\mathrm{sgn}(\hat{\mathbf{w}}_k(n))e_k(n) + \mu e_k(n)\mathbf{x}(n), \tag{4.10}$$

$$\hat{\mathbf{w}}_k(n+1) = \hat{\mathbf{w}}_k(n) - \rho_{SRZA}\,\mathrm{sgn}(\hat{\mathbf{w}}_k(n))/(1+\epsilon\|\hat{\mathbf{w}}_k(n)\|)e_k(n) + \mu e_k(n)\mathbf{x}(n). \tag{4.11}$$

respectively. As one can observe, an additional factor $e_k(n)$ is added to the second term of the right-hand side formula for S-ZA-LMS and S-RZA-LMS. The motivation behind the introduction of the additional factor is described next. In S-ZA-LMS and S-RZA-LMS, many elements of the filter are attracted to zero and the number of zero coefficients presenting the sparsity rises considerably compared to S-LMS. The reason for this additional term is to compensate for the effect of the zero sections in the UWB PPM input signal and the channel. Throughout the adaptive filtering process, if a section of input signal $\mathbf{x}(n)$ is all-zero, the expression $\mu e_k(n)\mathbf{x}(n)$ is forced to zero. Likewise, the same situation occurs when $\mathbf{x}(n)$ is non-zero but the corresponding part of the sparse multi-path is all-zero. Thus, to prevent any loss of information and any unnecessary update for $\hat{\mathbf{w}}(n)$ (where $\mathbf{x}(n)$ is a vector of zeros), we add the error $e_k(n)$ (equal to zero for this case) to S-ZA-LMS updating term $-\rho_{SZA}\text{sgn}(\hat{\mathbf{w}}(n))e_k(n)$ and to that of S-RZA-LMS, $-\rho_{SRZA}\,\text{sgn}(\hat{\mathbf{w}}_k(n))/(1+\epsilon\|\hat{\mathbf{w}}_k(n)\|)e_k(n)$ thereby forcing this term to be zero.

4.2 Channel estimation performance of S-LMS-type algorithms

In Section 4.1, we stated the specific claims that were to be investigated in this thesis. In this section, we will give details on the corresponding experiments and analyze the obtained results. We will look at the effects of different factors on the performance of the modified adaptive filtering. The length of the filter specifies how accurately a given system can be modeled by the adaptive filter. Increasing the filter length also affects reversely the convergence rate by increasing the computational time. In the case of increasing the channel length, the step size will have to be decreased to preserve the stability of the algorithm.

Fig. 4.3 shows the performance of the proposed S-LMS-type algorithms: S-LMS, S-ZA-LMS and S-RZA-LMS channel estimation schemes. To generate the MSE in this experiment for S-LMS-type algorithms, variance of noise is 10^{-7}, $\epsilon = 10$, $\mu = 0.005$, $\rho_{SZA} = 0.0005$, and $\rho_{ZA} = 0.0005$. Also, the long highly sparse multi-path channel, as explained in Chapter 2, is considered. The number of iteration needed for algorithms to converge is 3×10^6. As one can see, the MSE is erratic at the beginning, but with more iterations the MSE tends to converge closer to zero. S-RZA-LMS performance in estimation of the long sparse multi-path channel is superior to that of S-LMS and S-ZA-LMS. The reason is that its reweighting zero attraction property accelerates the convergence and improves the adaptive filtering. The sudden jumps on the graph, when MSEs go to zero, are from applying the all-zero parts of the PPM input signal to non-zero elements of the channel and vice versa. S-LMS-type algorithms minimize the number of times these situations happen and result in smoother MSE curves. The same experiment with similar setting is done on the LMS-type algorithms, however, there is no result obtained because of the high error. In Fig. 4.4, comparing the estimated multi-path channel from S-RZA-LMS algorithm and the channel obtained from

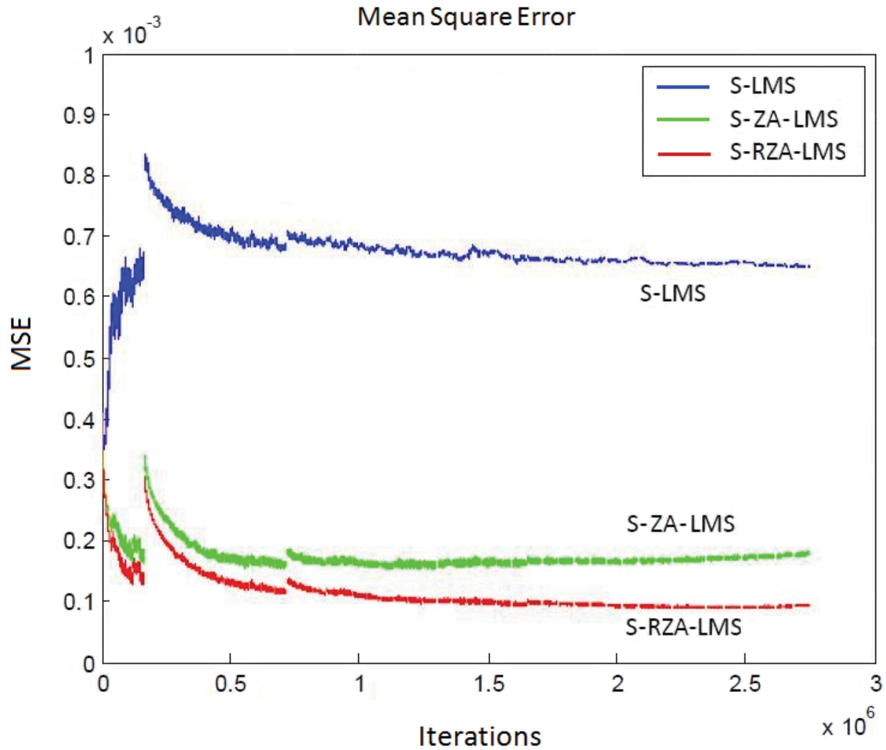

Figure 4.3 MSE of S-LMS, S-ZA-LMS, and S-RZA-LMS. The MSE of S-RZA-LMS exhibits superior performance compare to that of S-ZA-LMS and S-LMS, respectively.

the theoretical values is done. The impulse response of the indoor multi-path channel is as modeled by IEEE 802.15-02/279r0-SG3a and based on the Saleh-Valenzuela model from Chapter 2 and compared with our estimated channel. The plots are closely matched that shows the small error and good estimation using S-RZA-LMS. The graph in Fig. 4.5 shows the result of applying the proposed S-LMS, S-RZA-LMS, LMS, and RZA-LMS algorithms on three short filters of different sparsities with non-modulated signal as the input. The input signal and the observed noise are white Gaussian random sequences with variance of 1 and 0.001, respectively. The parameters are set as $\mu = 0.05$, $\rho_{RZA} = 0.0005$, and $\epsilon = 10$. For the first 500 iterations, sparsity of filter is set to 0.063 and the filter is of length 16. So, one can see that with very sparse filter the convergence of MSE is very fast. When the sparsity of the filter is set to 0.5, from the iterations 500 to 1000, the performance of S-RZA-LMS has the fastest convergence rate among all other algorithms. Also, for iterations after 1000, the filter is set to be non-sparse and still the algorithm S-RZA-LMS is the most

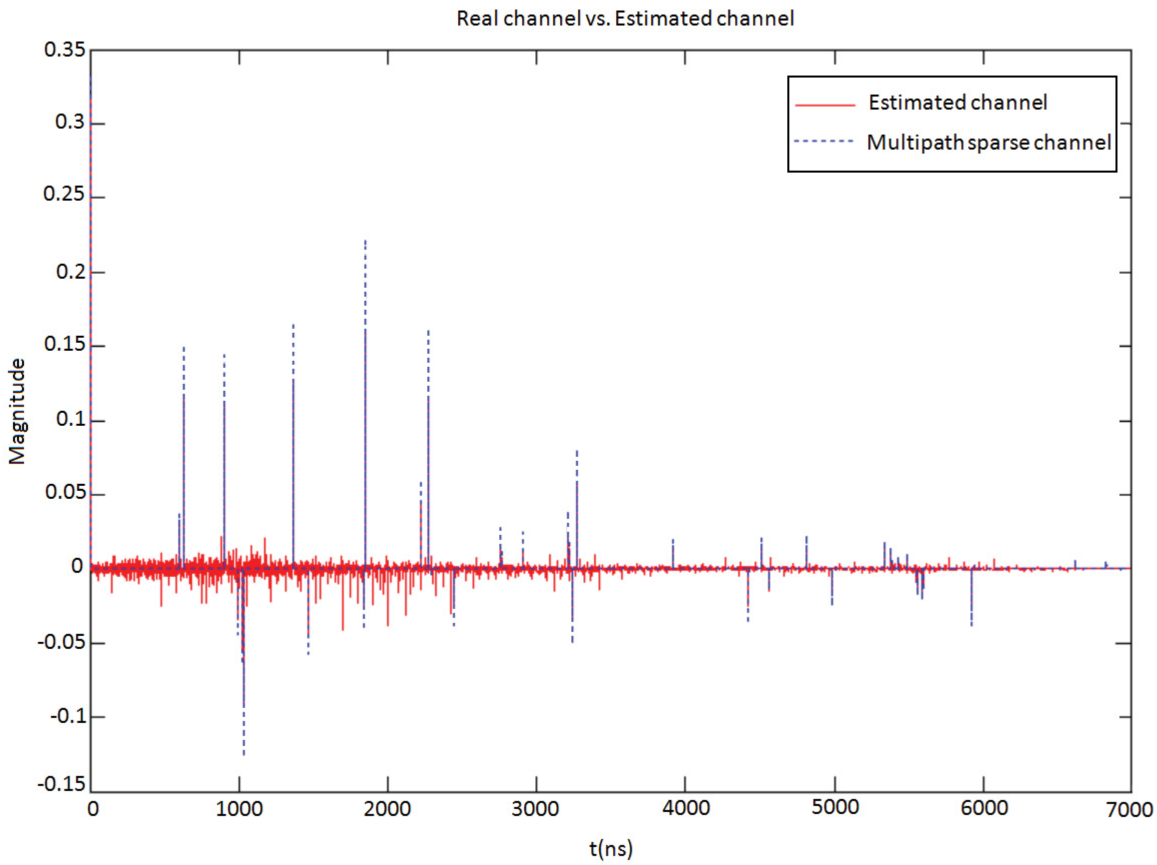

Figure 4.4 Estimated multi-path channel using S-RZA-LMS closely matches the multi-path channel (the Saleh-Valenzuela model)

efficient compare to others. The shrinkage property and the sectional adaptive filtering of S-RZA-LMS result in its improved performance. Moreover, one can see that the MSE is higher when the filter is non-sparse compare to the sparse filter. Fig. 4.6 presents the effect of using a very large value of $\mu = 0.001$ on the MSE convergence of S-LMS, S-ZA-LMS, and S-RZA-LMS in the case of the sparse multi-path channel. Except μ, all the other parameters are the same as those in the experiment of Fig. 4.3. The number of iteration is 9×10^5. The MSE graph shows a very fast convergence at the beginning for S-LMS, S-ZA-LMS, and S-RZA-LMS, however, as expected from the theory explained in Section 3.3, a large μ leads to the instability of the algorithms around the minimum value and leads to an erroneous result. S-LMS shows the least affected in the case of a large μ compare to S-ZA-LMS and S-RZA-LMS This might be because of its slower convergence rate compare to that

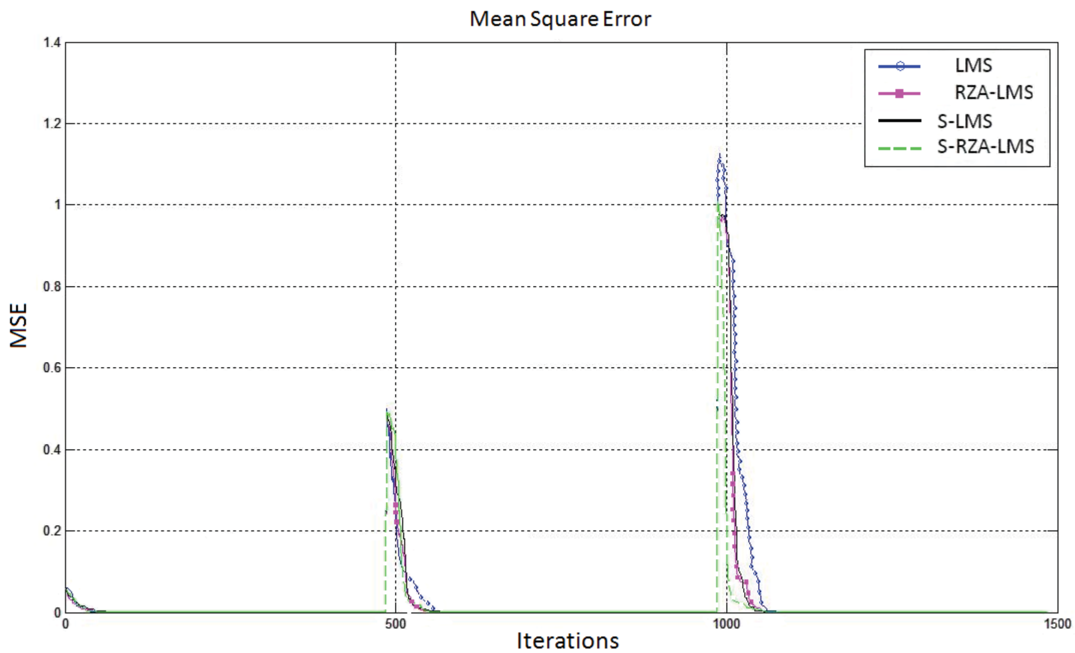

Figure 4.5 MSE of S-LMS, S-RZA-LMS, LMS, and RZA-LMS algorithms on three filters with different sparsities with a non-modulated input signal.

of S-ZA-LMS and S-RZA-LMS, so that it can handle the larger μ better and has a more stable performance. Fig. 4.7 presents the effect of using a very small value of $\mu = 0.0005$ on the MSE convergence of S-LMS, S-ZA-LMS, and S-RZA-LMS. The experiment setting is as the experiment related to Fig. 4.3. The number of iteration is 9×10^5. The MSE graph shows that S-LMS, S-ZA-LMS, and S-RZA-LMS algorithms converge very slowly because of the excessive number of iterations needed to reach the minimum point, as expected from the theory. As mentioned earlier, the jumps on the graph are related to the times when the all-zero parts of the PPM input signal (or channel) are applied to non-zero elements of the channel (or the PPM input signal).

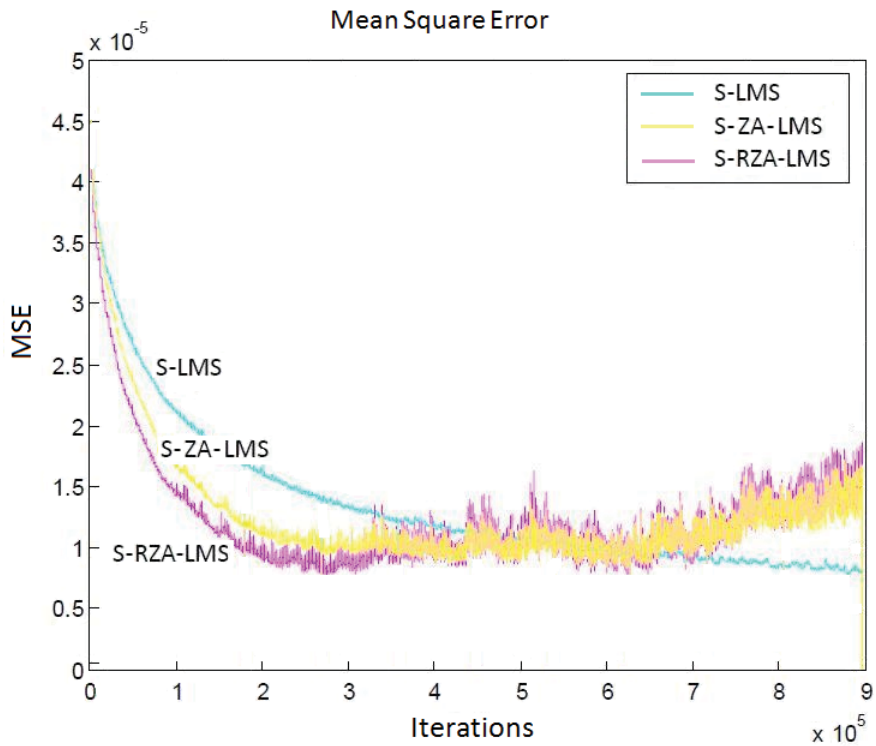

Figure 4.6 Effect of a very large μ in the behaviour of the MSE curve for S-LMS, S-ZA-LMS, and S-RZA-LMS. The large μ leads to the instability of the algorithms around the minimum value and leads to an erroneous result.

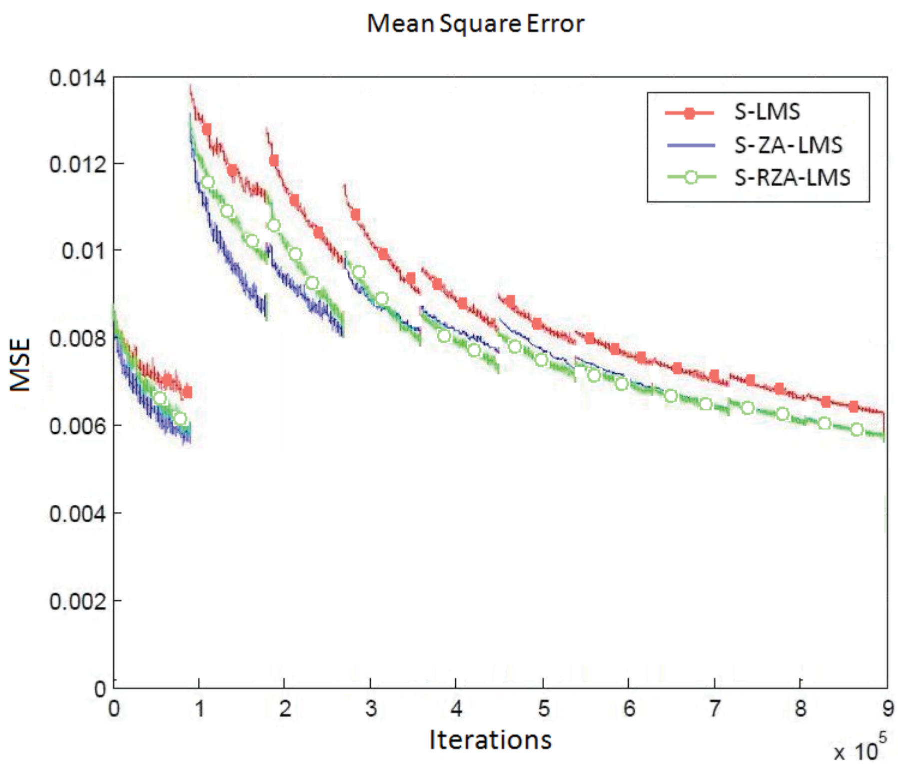

Figure 4.7 Effect of a very small μ in the behaviour of the MSE curve for S-LMS, S-ZA-LMS, and S-RZA-LMS. The MSE of S-LMS, S-ZA-LMS, and S-RZA-LMS algorithms converge very slowly because of the excessive number of iterations needed to reach the minimum point due to the very small μ.

Chapter 5

Summary and Conclusions

In this thesis we considered the problem of multi-path channel estimation in UWB systems. First, we presented the essential background material relating to this thesis. This material included a general overview of UWB systems, the UWB signal and channel models, and CS theory. In addition, we described in detail previous work on AF-based channel estimation methods, namely the LMS, ZA-LMS and RZA-LMS algorithms that formed the basis for the methods developed in this thesis. It was demonstrated that the ZA-LMS and RZA-LMS algorithms that enforce sparsity on the channel estimate throughout the estimation process, offer significant speedups in identifying sparse systems. Then, we applied the three above-mentioned AF methods to the channel estimation problem of UWB channels and we presented modifications of them that are capable of handling estimation of long multi-path channels with sparse inputs such as PPM input signals. The principal algorithm proposed is the S-LMS algorithm, described in Section 4.1. In contrast to traditional LMS methods, in the S-LMS algorithm the CIR is not estimated as a whole but in parts. This approach allows the efficient estimation of long channels. We extended S-LMS using ideas borrowed from ZA-LMS and RZA-LMS algorithms to the problem of sparse channel estimation and we developed the S-ZA-LMS and RZA-LMS methods. Finally, we demonstrated using simulations results and comparisons the channel estimation performance of the proposed methods.

The proposed methods have been designed specifically for long sparse channels with sparse inputs, a scenario where traditional methods fail as they struggle to deal with the large amount of parameters (channel tap weights) to be estimated in the presence of information loss. The results in this thesis show that significant performance gains can be attained by developing algorithms that capitalize on certain structural properties of the channel being estimated and its input signal.

Future research directions

In this work, we have focussed on the UWB multi-path channels estimation with AF-type algorithms. However, there are many issues, some of which are listed below, that warrant additional investigation.

- Even though the S-LMS channel estimator shows a promising performance, its implementation is restricted by the Nyquist sampling rate requirement. Therefore, it would be interesting to investigate the combination of CS theory that can circumvent this requirement, with S-LMS towards the development of a CS variant of S-LMS. To that end, we can use a random filter structure to form the compressed version of the original impulse response. This FIR filter is added to the transmitter front end. Then, we can interpolate the transmitted pilot (input) signal at the transmitter. The original sparse system can be reconstructed by the conventional recovery algorithms where the de-noising property of CS can be deployed. This approach has the potential to offer accurate channel estimation at low computational (due to use of AF techniques) as well as low implementation (due to the use of inexpensive ADCs) costs.

- Another suggestion for future research is the extension of the developed methods to the case of multiuser systems where many users communicate simultaneously over the UWB spectrum. In these systems, the users utilize unique time-hopping or direct-sequence codes that can be used to decouple their transmissions. A possible approach towards that end could be to precede the channel estimator by a multiple user interference (MUI) reduction block that uses the knowledge of the code of a user to cancel the interference from other users; the channel estimator would have to be modified to be able to handle the residual MUI at its input signal.

- Further research can take place on estimation of the highly sparse multipath channel using different modulation schemes, such as TH-BPSK that is another common form of signal modulation in pulsed UWB communications. The BER performance of TH-PPM and TH-BPSK should be compared for highly sparse channel estimation.

Bibliography

[1] G. Llano, J. C. Cuellar, and A. Navarro, *Ultra Wideband Communications: Novel Trends - Antennas and Propagation*, ch. Frequency *UWB* Channel. InTech, August 2011.

[2] L. Yang and G. B. Giannakis, "Ultra-wideband communications - an idea whose time has come," *IEEE Signal Processing Magazine*, vol. 21, pp. 26–54, November 2004.

[3] F. C. C. (FCC), "First report and order in the matter of revision of part 15 of the commission's rules regarding ultra-wideband transmission systems," *Report series ET Docket 98-153, FCC 02-48*, April 2002.

[4] V. Lottici, A. D'Andrea, and U. Mengali, "Channel estimation for ultra-wideband communications," *IEEE Journal on Selected Areas in Communications*, vol. 20, pp. 1638–1645, December 2002.

[5] S. Jogi and M. Choudhary, *Ultra-Wideband Demystified Technologies, Applications, and System Design Considerations*, vol. 4. River publishers, 2009.

[6] T. Mitchell, "Broad is the way," *IEE Review*, vol. 47, pp. 35–39, January 2001.

[7] M. Win and R. Scholtz, "Ultra-wide bandwidth time-hopping spread spectrum impulse radio for wireless multiple-access communications," *IEEE Signal Processing Magazine*, vol. 48, pp. 679–689, April 2000.

[8] F. Nekoogar, *Ultra-Wideband Communications: Fundamentals and Applications*. Prentice Hall, 1st ed., 2005.

[9] L. Miller, "Why *UWB*? a review of ultra-wideband technology," *Report to NETEX Project Office, DARPA by Wireless Communication Technologies Group National Institute of Standards and Technology, Gaithersburg, Maryland*, April 2003.

[10] D. Iyengar and P. L. Mcknight, "Understanding ultra-wideband technology: Advantages, applications, and regulatory policy," Master's thesis, Masters thesis, The Fletcher School of Law and Diplomacy, Tufts University, Boston, May 2002.

[11] A. Gupta and P. Mohapatra, “A survey on ultra-wideband medium access control schemes,” *Computer Networks*, vol. 51, pp. 2976–2993, August 2007.

[12] D. Leeper, “A long term-view of short-range wireless,” *Computer*, vol. 34, pp. 39–44, June 2001.

[13] J. Homer, H. Tang, and I. Longstaff, “Radar imaging of shallow buried objects,” in *Proceedings of 1999 IEEE International Geoscience and Remote Sensing Symposium (IGARSS'99)*, vol. 5, pp. 2477–2479, 1999.

[14] S. Nag, H. Fluhler, and M. Barnes, “Preliminary interferometric images of moving targets obtained using a time-modulated ultra-wideband through-wall penetration radar,” in *Proceedings of the 2001 IEEE Radar Conference*, pp. 64–69, 2001.

[15] H. Zhang, T. Udagawa, T. Arita, and M. Nakagawa, “A statistical model for small-scale multipath fading characteristics of UWB indoor channel,” in *Proceedings of IEEE Conference on Ultra-Wideband Systems and Technologies*, pp. 81–86, May 2002.

[16] W. U. Bajwa, J. Haupt, G. Raz, and R. Nowak, “Compressed channel sensing,” in *42nd Annual Conference on Information Sciences and Systems (CISS 2008)*, pp. 5–10, 2008.

[17] S. F. Cotter and B. D. Rao, “Sparse channel estimation via matching pursuit with application to equalization,” *IEEE Transactions on Communications*, vol. 50, pp. 374–377, March 2002.

[18] B. Sklar, “Rayleigh fading channels in mobile digital communication systems characterization,” *IEEE Communications Magazine*, vol. 35, pp. 90–100, July 1997.

[19] C. M. Canadeo, “Ultra-wideband multiple access performance using TH-PPM and DS-$BPSK$ modulations,” Master's thesis, Air Force Institute of Technology, Wright-Patterson Air Force Base, Ohio, March 2003.

[20] A. Goldsmith, *Wireless Communications*. New York, NY: Cambridge University Press, 2nd ed., 2005.

[21] P. Bello, “Characterization of randomly time-variant linear channels,” *IEEE Transactions on Communication*, vol. 11, pp. 360–393, December 1963.

[22] L. Zhao and A. M. Haimovich, “Capacity of m-ary PPM ultra-wideband communications over $AWGN$ channels,” in *Proceedings of the 54th IEEE Vehicular Technology Conference (VTC '01)*, vol. 2, pp. 1191–1192, October 2001.

[23] A. A. M. Saleh and R. Valenzuela, "A statistical model for indoor multipath propagation," *IEEE Journal on Selected Areas in Communications*, vol. 5, pp. 128–137, February 1987.

[24] K. A. Handi and G. Xuanye, "Bit error rate analysis for *TH-CDMA/PPM* impulse radio networks," in *Proceedings of the 2003 IEEE Wireless Communications and Networking (WCNC 2003)*, vol. 1, pp. 167–172, March 2003.

[25] A. Taha and K. Chugg, "A theoretical study on the effects of interference uwb multiple access impulse radio," in *Conference Record of the Thirty-Sixth Asilomar Conference on Signals, Systems and Computers*, vol. 1, pp. 728–732, November 2002.

[26] J. R. Foerster, "The performance of a direct-sequence spread ultra-wideband system in the presence of multipath, narrowband interference, and multiuser interference," in *Proceedings of 2002 IEEE Conference on Ultra-Wideband Systems and Technologies*, pp. 87–91, May 2002.

[27] D. Cassioli, M. Z. Win, and A. F. Molisch, "The ultra-wide bandwidth indoor channel: from statistical model to simulations," *IEEE Journal on Selected Areas in Communications*, vol. 20, pp. 1247–1257, August 2002.

[28] W. Su, "Ultra-wideband signals and systems in communication engineering," *IEEE Signal Processing Magazine*, vol. 25, pp. 122–123, September 2008.

[29] J. G. Proakis, *Digital Communications*. New York, USA: McGraw-Hill, Incorporation, 1st ed., 1995.

[30] J. M. Cramer, R. A. Scholtz, and M. Win, "On the analysis of *UWB* communication channels," in *Proceedings of IEEE Military Communications Conference (MILCOM 1999)*, vol. 2, pp. 1191–1195, October 1999.

[31] G. Maggio, N. Rulkov, and L. Reggiani, "Pseudo-chaotic time hopping for uwb impulse radio," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 48, pp. 1424–1435, December 2001.

[32] A. H. Jodar, "On the coexistence of the ultra-wideband wireless communication systems with narrowband interference," Master's thesis, School of Engineering Science-Simon Fraser University, Kharagpur, India, 2005.

[33] R. A. Scholtz in *IEEE Conference recording Communications on the Move. Military Communications Conference (MILCOM'93)*, vol. 2, pp. 447–450, October 1993.

[34] M. G. D. Benedetto and G. Giancola, *Understanding Ultra-Wideband Radio Fundamentals*. Pearson Education, June 2004.

[35] H. Zhang, W. Li, and T. A. Gulliver, "Pulse position amplitude modulation for time-hopping multiple-access UWB communications," *IEEE Transactions on Communications*, vol. 53, pp. 1269–1273, August 2005.

[36] G. Durisi and S. Benedetto, "Performance evaluation of TH-PPM UWB systems in the presence of multiuser interference," *IEEE Communications Letters*, vol. 7, pp. 224–226, May 2003.

[37] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, pp. 1289–1306, April 2006.

[38] E. J. Candes, J. K. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, pp. 489–509, January 2006.

[39] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Transactions on Information Theory*, vol. 52, pp. 5406–5425, November 2006.

[40] R. Chartrand and V. Staneva, "Restricted isometry properties and non-convex compressive sensing," *Inverse Problems*, vol. 24, no. 3, pp. 1–14, 2008.

[41] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

[42] E. J. Candes, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, vol. 346, pp. 589–592, February 2008.

[43] M. Z. Win and R. A. Scholtz, "On the robustness of ultra-wide bandwidth signals in dense multipath environments," *IEEE Communications Letters*, vol. 2, pp. 51–53, February 1998.

[44] S. Subramanian, "Compressed sensing methods in sparse channel estimation," Master's thesis, Indian Institute of Technology, Kharagpur, India, Summer 2008.

[45] E. J. Candes, "Compressive sampling," in *Proceedings of the International Congress of Mathematicians: Madrid*, pp. 1433–1452, August 2006.

[46] Y. Tsaig and D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, pp. 1289–1306, April 2006.

[47] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM Journal on Computing*, vol. 24, pp. 227–234, April 1995.

[48] G. Davis, S. Mallat, and M. Avellaneda, "Adaptive greedy approximations," *Constructive approximation*, vol. 13, pp. 57–98, March 1997.

[49] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," Tech. Rep. 1, February 1998.

[50] V. Chvatal, *Linear Programming.* W. H. Freeman and Company, New York, New York, 1st ed., 1993.

[51] K. M. Cheman, *Optimization Techniques for Solving Basis Pursuit Problems.* PhD thesis, North Carolina State University, Raleigh, North Carolina, March 2006.

[52] S. Mallat and Z. Zhang, "Matching pursuit with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, pp. 3397–3415, August 1993.

[53] Y. Pati, R. Rezaiifar, and P. Kirishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with application to wavelet decomposition," in *Proceedings of the 27th Annual Asilomar Conference on Signals, Systems, and Computers*, vol. 1, pp. 40–44, November 1993.

[54] J. A. Tropp, Anna, and C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, pp. 4655–4666, December 2007.

[55] G. Davis, S. Mallat, and Z. Zhang, "Adaptive time-frequency decompositions with matching pursuits," *Optical Engineering*, vol. 33, October 1994.

[56] G. Rath and A. Sahoo, "A comparative study of some greedy pursuit algorithms for sparse approximation," vol. 19, pp. 37–53, August 2009.

[57] P. S. Huggins and S. W. Zucker, "Greedy basis pursuit," *IEEE Transactions on Signal Processing*, vol. 55, pp. 3760–3772, June 2007.

[58] S. Mallat, *A Wavelet Tour of Signal Processing.* Academic Press, 3rd ed., 1998.

[59] S. Y. Park, Y. G. Kim, and C. G. Kang, "Iterative receiver for joint detection and channel estimation in $OFDM$ systems under mobile radio channels," *IEEE Transactions on Vehicular Technology*, vol. 53, pp. 450–460, December 2004.

[60] M. H. Hsieh and C. H. Wei, "Channel estimation for $OFDM$ systems based on comb-type pilot arrangement in frequency selective fading channels," *IEEE Transactions on Consumer Electronics*, vol. 44, pp. 217–225, December 2004.

[61] W. Dongming, H. Bing, Z. Junhui, G. Xiqi, and Y. Xiaohu, "Channel estimation algorithms for broadband *MIMO-OFDM* sparse channel," in *14th IEEE Proceedings on Personal, Indoor and Mobile Radio Communications (PIMRC 2003)*, pp. 1929–1933, September 2003.

[62] S. Haykin, *Adaptive Filter Theory*. Prentice Hall, 2002.

[63] M. L. Honig and D. G. Messerschmitt, *Adaptive Filters, Structures, Algorithms and Applications*. Springer, September.

[64] V. J. Mathews and S. C. Douglas, *Adaptive Filters*. Upper Saddle River, NJ: Prentice Hall, 2003.

[65] H. B. Curry, "The method of steepest descent for nonlinear minimization problems," *Quarterly of Applied Mathematics*, vol. 2, pp. 258–261, 1944.

[66] D. Luenberger, *Introduction to Linear and Nonlinear Programming*. Addison Wesley, 1984.

[67] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*, vol. 15. Prentice Hall, 1985.

[68] P. S. R. Diniz, *Adaptive Filtering: Algorithms and Practical Implementation*. Boston, MA: Springer-Verlag US, 2008.

[69] Y. Chen, Y. Gu, and A. O. H. Iii, "Sparse *LMS* for system identification," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3125–3128, April 2009.

[70] E. J. C, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted l1 minimization," *Journal of Fourier Analysis and Applications*, vol. 14, pp. 877–905, October 2008.

[71] R. C. Qiu, H. P. Liu, and X. Shen, "Ultra-wideband for multiple access," *IEEE Signal Processing Magazine*, vol. 43, pp. 80–87, February 2005.

[72] R. Qiu, R. Scholtz, and X. Shen, "Ultra-wideband wireless communications- new horizon," *IEEE Transactions on Vehicular Technology*, vol. 54, September 2005.

[73] W. Zhuang, X. Shen, Xuemin, and Q. Bi, "Ultra-wideband wireless communications," *Wireless Communications and Mobile Computing*, vol. 3, no. 6, pp. 663–685, 2003.