SPARSE SIGNAL RECOVERY FROM INCOMPLETE AND PERTURBED DATA

by

Rıfat Volkan Şenyuva

B.S., in Control Engineering, Istanbul Technical University, 2007

M.S., in Electrical & Electronics Engineering, Boğaziçi University, 2009

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

Graduate Program in Electical & Electronics Engineering

Boğaziçi University

2016

SPARSE SIGNAL RECOVERY FROM INCOMPLETE AND PERTURBED DATA

APPROVED BY:

Prof. Emin Anarım (Thesis Supervisor)	.....
Prof. Tayfun Akgül	.....
Assoc. Prof. Güneş Karabulut Kurt	.....
Assoc. Prof. Ali Emre Pusane	.....
Assist. Prof. Ahmet Öncü	.....

DATE OF APPROVAL: 18.04.2016

ACKNOWLEDGEMENTS

I want to thank my advisor Prof. Emin Anarım for his guidance and patience during my Ph.D. I am very grateful to Assoc. Prof. Güneş Karabulut Kurt, who introduced me to the subject of sparse signal recovery, for her guidance during the preparation of this thesis. I have very much benefited from her support and the criticisms on the subject. I greatly appreciate the support of Assist. Prof. Özgür Özdemir for introducing me to electromagnetic inverse scattering problems.

I would also like to thank Prof. Tayfun Akgül, Assoc. Prof. Ali Emre Pusane, and Assist. Prof. Ahmet Öncü for attending my thesis defense.

I would like to express my sincere gratitude to my family for their unconditional love and support. They never hesitated to give me all assistance they can throughout my entire life.

I gratefully acknowledge the funding received during my Ph.D. from TUBİTAK National Ph.D. Scholarship Programme 2211.

ABSTRACT

SPARSE SIGNAL RECOVERY FROM INCOMPLETE AND PERTURBED DATA

Sparse signal recovery consists of algorithms that are able to recover undersampled high dimensional signals accurately. These algorithms require fewer measurements than traditional Shannon/Nyquist sampling theorem demands. Sparse signal recovery has found many applications including magnetic resonance imaging, electromagnetic inverse scattering, radar/sonar imaging, seismic data collection, sensor array processing and channel estimation. The focus of this thesis is on electromagentic inverse scattering problem and joint estimation of the frequency offset and the channel impulse response in OFDM. In the electromagnetic inverse scattering problem, the aim is to find the electromagnetic properties of unknown targets from measured scattered field. The reconstruction of closely placed point-like objects is investigated. The application of the greedy pursuit based sparse recovery methods, OMP and FTB-OMP, is proposed for increasing the reconstruction resolution. The performances of the proposed methods are compared against NESTA and MT-BCS methods. Simulations show that the FTB-OMP method increases the resolution of the regular OMP and is superior to NESTA for less noisy measurements. OFDM is a multicarrier modulation technique that is very sensitive to frequency synchronization and channel estimation errors. Frequency offset destroys the orthogonality of the OFDM carriers and results in intercarrier inteference that causes severe performance degradation. A new approach that represents the channel impulse response as a 1-block sparse signal in a dictionary built by concatenating subspaces of frequency offset values is proposed. Thus the frequency offset and the channel impulse response can be jointly estimated. Only one OFDM training block is used and noise or channel statistics are not required. Its performance is close to maximum likelihood estimation and does not depend on frequency offset.

ÖZET

EKSİK VE BOZUK VERİLER İLE SEYREK SİNYAL GERİÇATIMI

Seyrek sinyal geriçatımı, yüksek boyutlu sinyalleri az sayıda örnek üzerinden tekrar oluşturabilen yöntemlerden meydana gelir. Bu yöntemler ile sinyal geriçatımı için gereksinim duyulan örnek sayısı, geleneksel Shannon/Nyquist örnekleme teoremine kıyasla çok daha az sayıdadır. Seyrek sinyal geriçatımı manyetik rezonans görüntüleme, elektromanyetik ters saçılım problemi, radar/sonar görüntüleme, sismik veri toplama, sensör dizi işleme ve kanal kestirimi olmak üzere birçok uygulamada kullanılmaktadır. Bu tez çalışmasının odağı elektromanyetik ters saçılım problemi ve OFDM için frekans kaymasının ve kanal yanıtının birlikte kestirilmesidir. Ters saçılım probleminde saçılan elektromanyetik alandan hedef cisimlerin özelliklerinin belirlenmesi amaçlanmaktadır. Bu kapsamda birbirine yakın konumlandırılmış nokta cisimlerin yerlerinin bulunması incelenmiştir. Bu problemdeki geriçatım çözünürlüğünün iyileştirilmesi için açgözlü geriçatım yöntemlerinden dikgen uyum kovalama OMP ve esnek ağaç arama yapılı FTB-OMP yöntemleri önerilmiştir. Önerilen yöntemlerin başarımları NESTA ve MT-BCS yöntemleri ile karşılaştırılmıştır. Yapılan benzetimler FTB-OMP yönteminin OMP geriçatım çözünürlüğünü artırdığı ve NESTA' ya göre düşük gürültü seviyelerinde üstünlük sağladığı göstermiştir. OFDM çoklu taşıyıcılı modülasyon yöntemi olarak frekans senkronizasyonuna ve kanal kestirim hatalarına karşı çok hassastır. Frekans kayması OFDM taşıyıcılarının dikgenliğini bozar ve taşıyıcılar arası girişime sebep olarak sistem başarımını düşürür. Önerilen yeni yöntemde frekans kayma değerlerinden oluşturulan bir kitaplık ile kanal yanıtı 1-blok seyrek bir sinyal olarak temsil edilir. Böylece frekans kayması ve kanal yanıtı birlikte kestirilebilmektedir. Sadece bir OFDM öbeği pilot yapılırken gürültü veya kanal istatistiğine gerek duyulmaz. Başarımı en büyük olabilirlik kestirimine çok yakındır ve frekans kaymasından bağımsızdır.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

$\mathbf{A}$	Sensing matrix
$\mathbf{A}^*$	Hermitian of matrix $\mathbf{A}$
$\mathbf{A}_\Lambda$	Submatrix of $\mathbf{A}$ with columns restricted to Λ
$\widetilde{\mathbf{A}}$	Perturbed sensing matrix
$\mathbf{C}_{\epsilon_\mathrm{n}}$	Circular matrix with normalized frequency offset coefficients in frequency domain
C_{ϵ_n}	Normalized frequency offset coefficient in frequency domain
c_{ϵ_n}	Normalized frequency offset coefficient in time domain
$\mathcal{C}_\mathbf{A}$	Computational cost of multiplication with matrix $\mathbf{A}$
E^inc	Incident electric field
E^sc	Scattered electric field
$E_i^\mathrm{tot}(\mathbf{r}')$	Total electric field associated with transmitting source i
$\mathrm{D}_\mathrm{KL}(\mathcal{P}_i, \mathcal{P}_l)$	Kullback-Leibler divergence between distributions $\mathcal{P}_i$ and $\mathcal{P}_l$
$D(\epsilon_\mathrm{n})$	Degradation of signal to noise ratio in an OFDM system
$\mathbf{F}$	Discrete-time Fourier matrix
$\mathbf{F}_{N_\mathrm{g}}$	Submatrix of $\mathbf{F}$ with the first N_g columns
$\mathbf{F}_\mathbf{w}$	Submatrix of $\mathbf{F}$ with rows restricted to $\mathbf{w}$
$G(\mathbf{r}, \mathbf{r}')$	Green function of homogeneous medium
H_0	Zero-order Hankel function of the first kind
$H(\mathbf{x})$	Entropy of $\mathbf{x}$
$H(\mathbf{x}\|\mathbf{y})$	Entropy of $\mathbf{x}$ conditioned on $\mathbf{y}$
$\mathbf{I}$	Identity matrix
$\mathbb{I}$	Indicator function
$I(\mathbf{x}, \mathbf{y})$	Mutual information of $\mathbf{x}$ and $\mathbf{y}$
J_0	Zero order Bessel function of first kind
k_0	Wavenumber of free space
L	Channel length in OFDM
N_c	Total number of cells in discretized investigation domain
N_g	Number of guard samples in an OFDM block

N_{rx}	Number of receiving antennas in electromagnetic imaging
N_{s}	Number of subcarriers in an OFDM system
N_{t}	Total number of samples in an OFDM block
N_{tx}	Number of transmitting antennas in electromagnetic imaging
$\mathbf{p}$	Occupation probability vector
P_h	Total average power of channel impulse response
s	Sparsity level of the signal
$S(\mathbf{x})$	Support set of $\mathbf{x}$
$\mathrm{SNR}(\epsilon_{\mathrm{n}}, k)$	Signal to noise ratio of the k-th subcarrier in an OFDM system
T	Sampling period of OFDM system
$\mathrm{spark}(\mathbf{A})$	Minumum number of linearly dependent columns of $\mathbf{A}$
$\mathbf{X}$	Diagonal matrix containing OFDM symbols
$\mathbf{x}$	Sparse signal of interest
$\mathbf{x}^{(s)}$	Head of the sparse signal with s smallest entries set to zero
$\mathbf{x}^{(t)}$	Tail of the sparse signal with $n - s$ largest entries set to zero
$\mathbf{y}$	Measurement vector
$\tilde{\mathbf{y}}$	Perturbed measurement vector
$\mathbf{Z}$	Multiplicative noise matrix
$\mathbf{z}$	Additive noise vector
α_s	Constant of the null space property for sparsity s
δ_s	Restricted isometry constant for sparsity s
$\tilde{\delta}_s$	Modified restricted isometry constant for $\widetilde{\mathbf{A}}$ and sparsity s
ϵ_{f}	Fractional part of normalized carrier frequency offset
ϵ_{i}	Integer part of normalized carrier frequency offset
ϵ_{n}	Normalized carrier frequency offset
ϵ_r	Relative permittivity
ε	Relative bound for Euclidean norm of additive noise
$\varepsilon_{\mathbf{y}}$	Relative perturbation bound for noise to measurement ratio
$\varepsilon_{\mathbf{A}}$	Relative perturbation bound for sensing matrix
$\varepsilon_{\mathbf{A}}^{(s)}$	Relative perturbation bound for submatrices $\mathbf{A}_\Lambda$ with $\lvert\Lambda\rvert = s$

ε_{tot}	Total relative perturbation bound
$\kappa_{\mathbf{A}}^{(s)}$	Ratio of upper and lower restricted isometry property bounds
$\lambda(\mathbf{A})$	Eigenvalue of matrix $\mathbf{A}$
$\lambda_{\min}(\mathbf{A})$	Minumum eigenvalue of matrix $\mathbf{A}$
$\lambda_{\max}(\mathbf{A})$	Maximum eigenvalue of matrix $\mathbf{A}$
ξ	Wavelength of the electromagnetic source
Λ	Subset of indices
$\lvert\Lambda\rvert$	Cardinality of the subset Λ
$\mu(\mathbf{A})$	Mutual coherence of matrix $\mathbf{A}$
$\mu_B(\mathbf{A})$	Block coherence of matrix $\mathbf{A}$
$\sigma_{\min}^{(s)}$	Minimum singular value of all s-column submatrices of $\mathbf{A}$
$\sigma_{\max}^{(s)}$	Maximum singular value of all s-column submatrices of $\mathbf{A}$
Σ_s	Space of s-sparse signals bounded by ℓ_0 pseudo-norm
$\chi(\mathbf{r}')$	Characteristic of the scatterer
Δf_{c}	Carrier frequency offset

LIST OF ACRONYMS/ABBREVIATIONS

AR	Autoregressive State Space Model
BA	Born Approximation
BCS	Bayesian Compressive Sensing
BOMP	Block Orthogonal Matching Pursuit
BP	Basis Pursuit
BWCB	Best Worst-Case Bound
CFO	Carrier Frequency Offset
CIR	Channel Impulse Response
CP	Cyclic Prefix
CS	Compressive Sensing
CRB	Cramer-Rao Bound
DFT	Discrete Fourier Transform
EM	Expectation Maximization
ERP	Exact Reconstruction Property
FIM	Fisher Information Matrix
FTB-OMP	Flexible Tree-search Based Orthogonal Matching Pursuit
IBI	Interblock Interference
ICI	Intercarrier Interference
IDFT	Inverse Discrete Fourier Transform
LASSO	Least Absolute Shrinkage and Selection Operator
LSP	Least Squares Problem
MLE	Maximum Likelihood Estimate
MRI	Magnetic Resonance Imaging
MSE	Mean Square Error
MT-BCS	Multitask Bayesian Compressive Sensing
NESTA	Nesterov's Algorithm
NSP	Null Space Property
OFDM	Orthogonal Frequency Division Multiplexing

OMP	Orthgonal Matching Pursuit
QP	Quadratic Program
RIC	Restricted Isometry Constant
RIP	Restricted Isometry Property
RVM	Relevance Vector Machine
SISO	Single Input Single Output
SOCP	Second Order Cone Program
TM	Transverse magnetic
WWCB	Worst Worst-Case Bound

1. INTRODUCTION

One of the most fundamental building blocks in digital signal processing systems is the analog to digital converter. This block converts an analog signal to bits that can then be used by digital systems to serve different purposes such as transmission, storage, and estimation. The first step of analog to digital conversion is to discretize the signal in time, which is called sampling. The common practice in acquisition of a signal is to sample it according to the Shannon-Nyquist-Whittaker theorem [1–3] and reconstruct it with linear interpolation techniques. Current data acquisition methods are often extremely wasteful: massive amounts of data are acquired first and then discarded by a subsequent compression stage. In contrast, a recent approach known as compressed sensing [4, 5] combines the data collection and compression and reduces the volume of data enormously at much greater efficiency and lower costs. A paradigm of encoding of analog sources by real numbers rather than bits, compressed sensing deals with effcient recovery of sparse vectors from the information provided by linear measurements. By leveraging the prior knowledge of the signal structure, e.g. sparsity, and designing efficient non-linear reconstruction algorithms, effective compression is achieved by taking a much smaller number of measurements than the dimension of the original signal. Signals with sparse structure are encountered in many applications of imaging, signal processing and communications including magnetic resonance imaging, electromagnetic imaging, radar/sonar imaging, sensor array processing, seismic data collection, estimation of sinusoids in noise, locating errors in digital transmissions, and channel estimation. There is a wide variety of available techniques and algorithms, with the emergence of newer sparsity promoting techniques such as basis pursuit or greedy pursuits.

The case of incomplete observations or undersampling corresponds to a signal of interest, $\mathbf{x}$, being observed from a set of linear measurements

$$\mathbf{y} = \mathbf{A}\mathbf{x} \tag{1.1}$$

where the number of rows of the sensing matrix is less than the number of its columns, $m < n$. This means that there are more unknowns than the number of equations and such a system is called an underdetermined system of linear equations. The solution set is either empty, $\mathbf{y}$ is not in the range space of the columns of $\mathbf{A}$, or there exists infinite number of solutions due to the nontrivial right nullspace of the sensing waveforms. Finding a single solution amongst infinitely many can be achieved by imposing additional constraints on the signal of interest, $\mathbf{x}$. The sparsity constraint assumes that the support of $\mathbf{x}$, the set of indices corresponding to the nonzero entries, $S(\mathbf{x}) = \{i : x_i \neq 0\}$, has at most s elements

$$\mathbf{x} \in \Sigma_s := \{\mathbf{x} : \|\mathbf{x}\|_0 = |S(\mathbf{x})| \leq s\} \tag{1.2}$$

where ℓ_0-pseudo-norm is defined as the cardinality of the support set.

The recovery process under the sparsity assumptions is the following combinatorial optimization problem (1.3).

$$\begin{aligned} &\text{minimize} \quad \|\mathbf{x}\|_0 \\ &\text{subject to} \quad \mathbf{y} = \mathbf{A}\mathbf{x} \qquad (P_0) \end{aligned} \tag{1.3}$$

This optimization problem is nonconvex and generally impossible to solve since it requires an intractable combinatorial search. It is now well known that methods such as basis pursuit and matching pursuit promotes sparsity. These sparsity promoting methods are functions of the data $(\mathbf{y}, \mathbf{A})$. They are ideal because the data input $(\mathbf{y}, \mathbf{A})$ is not perturbed. However, in practice both the observation vector, $\mathbf{y}$, and the sensing matrix, $\mathbf{A}$, may be perturbed and the problem under general perturbations, ε_{tot}, can be given as in (1.4).

$$\begin{aligned} &\text{minimize} \quad \|\mathbf{x}\|_0 \\ &\text{subject to} \quad \|\tilde{\mathbf{y}} - \widetilde{\mathbf{A}}\mathbf{x}\|_2 \leq \varepsilon_{\text{tot}} \qquad (P_0^{\varepsilon_{\text{tot}}}) \end{aligned} \tag{1.4}$$

1.1. Literature Review

Convex relaxation methods such as the basis pursuit method and the Bayesian compressive sensing has been applied in the reconstruction of objects for microwave imaging [6–8]. Through-the-wall radar imaging and microwave imaging of sparse scatterers has been recast into the Bayesian compressive sensing framework [9–11]. The multitask Bayesian compressive sensing method which improves upon the Bayesian compressive sensing method by jointly considering multiple correlated measurements has also been applied to time-harmonic two-dimensional inverse scattering problems with both sparse and nonsparse scatterers [12–14].

There exists numerous implementations of the CS based block-sparse signal recovery methods for the estimation of the CIR of the OFDM systems [15–17]. In [16] the block-ISD algorithm is implemented to estimate the CIR of a MIMO OFDM system with multiple transmit antennas. Block-ISD method is based on the iterative support detection (ISD) reconstruction algorithm and uses a support set of the detected nonzero elements of the sparse signal in solving a truncated basis pursuit problem at each iteration. Block sparse equivalent CIR is generated by assuming that the CIRs from different antennas share a common support. In these works, block sparsity is achieved by either assuming that the several channel instantiations are group-sparse, locations of the nonzero channel coefficients are same [17] or concatenating multiple CIRs of different antennas with common support in a block sparse structure [15, 16]. However, frequency offset is not considered in these works, which assume perfect synchronization for the OFDM system and so only CIR can be estimated using these methods.

1.2. Main Contributions

In this thesis the application of the sparse recovery methods under perturbations to both the obeservations and the sensing matrix is considered. The sparse recovery framework is applied to the electromagnetic imaging and to the joint estimation of channel impulse response (CIR) and carrier frequency offset (CFO) for single input single output (SISO) OFDM systems. For electromagnetic imaging, the reconstruc-

tion of two closely spaced point-like objects from measured scattered electromagnetic field is investigated. The performances of the greedy orthogonal matching pursuit (OMP) based algorithms are compared to the performances of the convex relaxation method using Nesterov's smoothing technique (NESTA) and the multi-task Bayesian compressive sensing (MT-BCS) method. It is observed that the application of the flexible tree search based implementation of the orthogonal matching pursuit (FTB-OMP) algorithm increases the resolution of the reconstruction compared to the OMP. The FTB-OMP is also also able to perform superior than the NESTA and MT-BCS in the low noise region while the best performance under high noise belongs to the NESTA. The main contributions of this application can be summarized as:

- To the best of our knowlege this is the first study that applies greedy pursuit based CS methods to an electromagnetic inverse scattering problem.
- The performance of the greedy pursuit based methods scale better with increasing SNR level.
- Due to its lower computational complexity the OMP algorithm can be preferred instead of the more complex methods such as NESTA and MT-BCS when the number of scatters are few and are sufficiently separated.
- The tree search based method FTB-OMP method increases the resolution of the regular OMP method when higher number of scatterers are closely placed.
- While the FTB-OMP method is superior to NESTA in the high SNR region, the best performance belongs to the NESTA method in the low SNR region.

Then the joint estimation of CFO and CIR for SISO OFDM with pilot symbols is investigated. A new joint estimator that exploits sparse recovery framework is proposed for SISO OFDM systems. It is shown that the CIR can be represented as a 1-block sparse signal by using a dictionary constructed by concatenating subspaces of CFO values taken from a search space. The block orthogonal matching pursuit algorithm is implemented as the sparse recovery method. The performance of the proposed method is compared against the well-established pilot based estimators; Moose, Classen, the maximum likelihood estimator (MLE) and the p-algorithm. Numerical results show that the performance of the proposed method is independent of the value of the CFO.

The worst-case analysis of sparse recovery under general perturbations is also applied for the mean squared error of the CIR estimate for a sparse multipath channel. The contributions of this proposed framework are given as:

- The proposed CS framework allows to jointly estimate CFO and CIR.
- The proposed method only needs one OFDM block of training symbols.
- There is no need for any initialization since it does not depend on the statistics of the channel or the noise.
- The performance does not depend on CFO and is very close to the MLE method.
- Sparse recovery bounds from CS perturbation theory are applied from the channel estimation.

1.3. Outline of the Thesis

The organization of this thesis is as follows: Chapter 2 gives a brief review on the uniqueness of the solutions given by the sparse recovery algorithms and the stability analysis of the sparse solutions. In this chapter some of the most important properties used in the stability analysis of sparse recovery methods including spark, mutual coherence, restricted isometry property and null space property and their relations with each other are covered. The tools given in this chapter are also used in Chapter 3 to give the necessary conditions to achieve the performance guarantees of the sparse recovery methods. Chapter 3 reviews some of the existing sparse recovery algorithms and their theoretical performance guarantees. Sparse recovery algorithms are presented in three categories: greedy algorithms, the convex relaxation by ℓ_1 minimization, and the Bayesian compressed sensing method. While orthogonal matching pursuit, flexible tree search based orthogonal matching pursuit and block orthogonal matching pursuit algorithms are considered under greedy methods, interior point methods and the Nesterov's algorithm methods are given under ℓ_1 minimization. The Bayesian compressed sensing approach is covered under its own section. The performance bounds of the sparse recovery methods are given for additive and multiplicative noise in Chapter 4. A best case and a worst case analysis is provided for each perturbation. Chapter 5 investigates the application of the sparse recovery methods for the electromagnetic

imaging of closely placed objects. The reconstruction of closely placed point-like scatterers in inverse scattering problems is analysed. Numerical results regarding a far and a close proximity inverse scattering setup are given and the performances of the sparse recovery methods presented in Chapter 3 are compared. Joint estimation of CIR and CFO for OFDM systems is studied from the perspective of sparse recovery framework in Chapter 6. CFO is modeled as a multiplicative noise and sparse recovery bounds for sparse recovery methods under general perturbations (Chapter 4) are applied to compute worst case channel estimation bounds with CFO. The performance of the proposed method are compared against the well-established Moose, Classen, the maximum likelihood estimation and the p-algorithm for three channel models: a time invariant channel, a time varying channel and a time varying channel with varying carrier frequency offset.

2. UNIQUENESS OF SPARSEST SOLUTION

The uniqueness of a sparsest solution is an important question for sparse signal recovery. In order to analyse the conditions under which a sparse recovery method yields an unique solution, several properties/tools have been developed in the CS literature. This chapter gives a brief review on some of the most important ones including spark, mutual coherence, restricted isometry property and null space property and on how they are related to each other. Since some of these properties are also used in the performance bounds of the sparse recovery methods (Section 4), a thorough understanding of them is vital.

2.1. Spark

A key property to study uniqueness is the spark of the matrix $\mathbf{A}$, a term coined and defined by Donoho and Elad [18]. The spark is a way of characterizing the null space of the matrix using the ℓ_0 pseudo norm. The spark of a given matrix is the smallest number of columns that are linearly independent [18]. The spark of a matrix is far more difficult to obtain compared to its rank since it is related to linearly dependent matrix columns. Inclusion-wise minimal collections of linearly dependent columns are called circuits. A circuit is a set of column indices, $\Lambda \subseteq \{1, ..., n\}$, such that the submatrix constructed by extracting the columns corresponding to the indices in the set Λ, has a nonzero solution, $\mathbf{A}_\Lambda \mathbf{x} = \mathbf{0}$. Thus, finding the size of its smallest circuit is equivalent to computing its spark [19]. The spark gives a simple criterion for uniqueness of sparse solutions. By definition, the vectors in the null space of the matrix, $\mathbf{A}\mathbf{x} = \mathbf{0}$, must satisfy $\|\mathbf{x}\|_0 \geq \text{spark}(\mathbf{A})$, since at least $\text{spark}(\mathbf{A})$ number of columns must combine to give the zero vector.

Theorem 2.1. If a system of linear equations has a solution $\mathbf{x}_1$ which has a ℓ_0 pseudo norm less than the half of the spark of the sensing matrix, $\|\mathbf{x}_1\|_0 < \text{spark}(\mathbf{A})/2$, this solution is necessarily the sparsest possible [20].

Proof. Suppose there exists another, $\mathbf{x}_2$ that satisfies $\mathbf{A}\mathbf{x}_2 = \mathbf{y}$. Then the difference of these solutions must be in the null space, $\mathbf{A}(\mathbf{x}_1 - \mathbf{x}_2) = \mathbf{0}$. ℓ_0 pseudo norm satisfies the triangle inequality (2.1).

$$\|\mathbf{x}_1\|_0 + \|\mathbf{x}_2\|_0 \geq \|\mathbf{x}_1 - \mathbf{x}_2\|_0 \geq \text{spark}(\mathbf{A}) \tag{2.1}$$

Since $\|\mathbf{x}_1\|_0 < \text{spark}(\mathbf{A})/2$, the alternative solution must be $\|\mathbf{x}_2\|_0 > \text{spark}(\mathbf{A})/2$ in order to hold for (2.1). □

The value of spark can be informative. Checking the sparsity of the solution and then comparing that with the spark is a way to check global optimality. Large values of spark allow large values of sparsity. Spark must be in the range, $[2, m+1]$ [20].

2.2. Mutual Coherence

Since computing the spark is NP-complete [19], simpler ways to guarantee uniqueness are of interest. Mutual coherence [21–23] of a matrix is defined as the largest absolute normalized inner product between different columns of the sensing matrix (2.2).

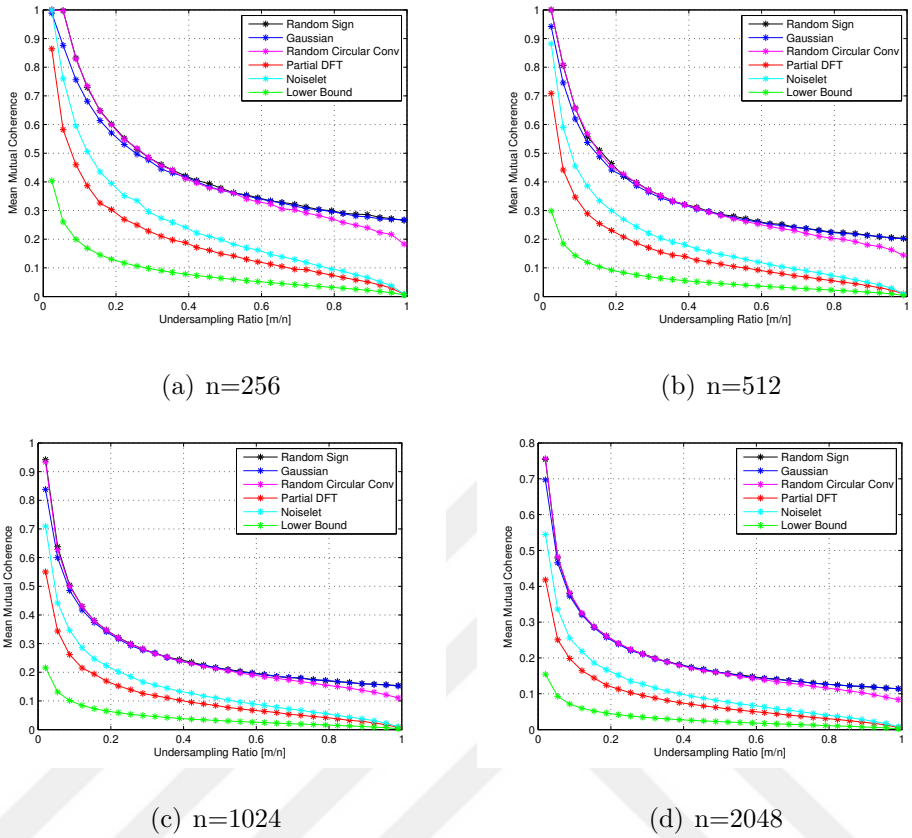
$$\mu(\mathbf{A}) = \max_{1 \leq i,l \leq n \; i \neq l} \frac{|\langle \mathbf{a}_i, \mathbf{a}_l \rangle|}{\|\mathbf{a}_i\|_2 \, \|\mathbf{a}_l\|_2} \tag{2.2}$$

$\mathbf{a}_i$ and $\mathbf{a}_l$ denote the i-th and l-th columns of $\mathbf{A}$ respectively. If the angle between each and every column pairs of the sensing matrix is the same, $|\langle \mathbf{a}_i, \mathbf{a}_l \rangle|$, this type of sensing matrices is called equiangular frames. If a sensing dictionary is built by bringing structured two-ortho matrices side-by-side, then the mutual coherence of this dictionary can be shown to satisfy $1/\sqrt{m} \leq \mu(\mathbf{A}) \leq 1$. Identity-Fourier and Identity-Hadamard pairs achieves the lower bound $\mu(\mathbf{A}) = 1/\sqrt{m}$ [24,25]. For full rank matrices of size $m \times n$, the smallest value of the mutual coherence is given as (2.3).

$$\mu(\mathbf{A}) \geq \sqrt{\frac{n-m}{m(n-1)}} \tag{2.3}$$

This bound (2.3) is also known as the Welch bound [24–27]. For general matrices with more columns than rows, mutual coherence is strictly positive. In order to get closest to the behaviour exhibited by unitary matrices, the sensing matrices with the smallest mutual coherences are desired, $\mu(\mathbf{A}) = O(1/\sqrt{m})$ [27]. Grassmanian frames are of unit norm and minimize the maximum correlation among all frames. Equality is achieved if and only if the sensing matrix is equiangular and tight, $\mathbf{A}\mathbf{A}^* = \mathbf{I}$.

The change in the mutual coherence of some of the well-known sensing dictionaries against the varying undersampling ratio and different signal dimensions can be observed in Figures 2.1(a), 2.1(b), 2.1(c), 2.1(d). The Bernoulli Ensemble, the Gaussian Ensemble, the random convolution matrix, the partial discrete Fourier transform (DFT), and the noiselet basis are used to generate the sensing matrices. The undersampling of the sensing matrices is accomplished by choosing the rows of the sensing matrix uniformly. The undersampling ratio range is kept within the interval of $(0, 1]$. The representation basis where the signal is assumed to be sparse is taken as the identity basis. For all signal dimensions, it is seen that the partial DFT and the noiselet basis yield sensing matrices with smaller mutual coherences while the mutual coherences of the sensing matrices built from the Bernoulli Ensemble, the Gaussian Ensemble and random convolution sensing matrices are greater. The mutual coherences of the Bernoulli and the Gaussian sensing matrices are very close to each other within the whole undersampling ratio range for all signal dimensions. The mutual coherence of the random convolution matrix is also close to these two within the undersampling ratio interval of $[0.4, 0.6]$ for signal dimension $n = 256$ (Figure 2.1(a)). When the undersampling ratio increases beyond this range, the mutual coherence of the random convolution matrix gets smaller. However, as the signal dimension gets larger, the undersampling ratio range for which the mutual coherence of the convolution matrix is close to the Bernoulli and Gaussian sensing matrices increases. It is seen for all sensing matrices that decreasing the undersampling ratio increases the mutual coherence. The mutual coherence of the partial DFT is closest to the Welch bound. On the contrary as the signal dimension increases, it is observed that the mutual coherence for all sensing matrices gets smaller.

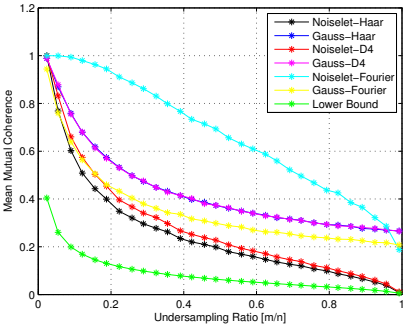


(a) n=256 (b) n=512

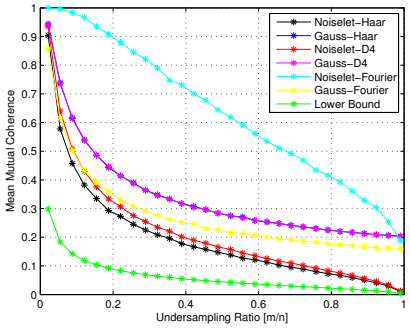
(c) n=1024 (d) n=2048

Figure 2.1. Mutual coherence of sensing matrices with identity basis.

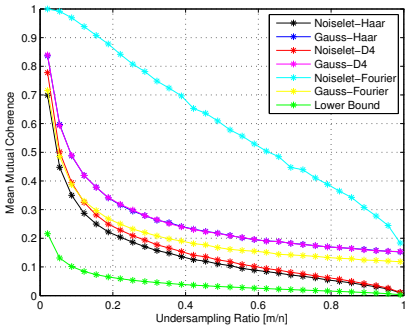
Representation basis where the signal is sparse does not always have to be the identity basis as shown in Figure 2.1 and so can from different basis such as a wavelet basis or a Fourier basis. In order to demonstrate such an example, the mutual coherences of Gaussian and noiselet sensing basis against Haar, Daubechies-4 (D4) and Fourier representation basis are given in Figures 2.2(a), 2.2(b), 2.2(c). For the Gaussian sensing matrices, changing the representation basis from Haar to D4 does not affect the mutual coherence. However, when the representation basis is Fourier, it is observed that the Gaussian sensing matrix yields lower mutual coherence compared to Gauss-Haar and Gauss-D4 pairs for all signal dimensions. When Haar and D4 basis are used as the representation basis for the noiselet sensing matrix, the mutual coherence is closest to the Welch bound. If the Fourier basis is paired with the noiselet sensing matrix, then the mutual coherence becomes the greatest. The mutual coherence of the noiselet sensing matrix also depends on the chosen representation basis. Noiselet-Haar and noiselet-D4 pairs appear to be the most incoherent basis pairs while Gauss-Haar and Gauss-D4 pairs result in larger coherences.



(a) n=256



(b) n=512



(c) n=1024

Figure 2.2. Mutual coherence of sensing matrices with wavelet basis.

Mutual coherence and spark of a sensing matrix are related to each other as shown in the following Theorem 2.2.

Theorem 2.2. For any sensing matrix $\mathbf{A}$, the spark is greater than the inverse of the mutual coherence (2.4).

$$\text{spark}(\mathbf{A}) \geq 1 + \frac{1}{\mu(\mathbf{A})} \tag{2.4}$$

Proof. Modify the matrix $\mathbf{A}$ by normalizing its columns to unit norm and obtain the Grammian matrix, $\mathbf{A}^*\mathbf{A}$. While diagonal entries of the Grammian is equal to 1 due to normalization, off-diagonal entries must be less than the mutual coherence by definition. Choose a subgroup of p columns from $\mathbf{A}$ and compute its Grammian matrix. If the sub-Grammian is diagonally dominant, this sub-matrix is positive definite and the chosen columns are linearly independent. The condition $p < 1 + 1/\mu(\mathbf{A})$ implies the smallest number of columns that leads to linear dependence is $p = 1 + 1/\mu(\mathbf{A})$. □

2.2.1. Minimizing Mutual Coherence via Stochastic Optimization

Choosing the rows of the sensing matrix during undersampling is an important problem since some row configurations may yield sensing matrices with lesser mutual coherence. One particular application is designing the pilot symbols in an OFDM system with pilot assisted channel estimation [28–30]. Let the sensing matrix be the DFT matrix whose rows are to be chosen as

$$\mathbf{F}_{\mathbf{w}} = \frac{1}{\sqrt{n}} \begin{bmatrix} 1 & e^{-j2\pi\omega_1/n} & \dots & e^{-j2\pi\omega_1(n-1)/n} \\ 1 & e^{-j2\pi\omega_2/n} & \dots & e^{-j2\pi\omega_2(n-1)/n} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & e^{-j2\pi\omega_m/n} & \dots & e^{-j2\pi\omega_m(n-1)/n} \end{bmatrix}_{m\times n} \tag{2.5}$$

where $\mathbf{w} = [\omega_1\, \omega_2 \dots \omega_m]^T$. Mutual coherence for this sensing matrix can be computed using Equation (2.6).

$$\mu(\mathbf{F}_{\mathbf{w}}) = \max_{1\le i,l\le n} \left| \sum_{v=1}^{m} e^{-2j\pi\omega_v(l-i)/n} \right| \tag{2.6}$$

The problem is to find the rows or in this case the frequency vector, $\mathbf{w}$, which gives the sensing matrix with the least mutual coherence. It is possible to exhaustively search for all of $\binom{n}{m}$ row configurations and find the one with the least mutual coherence but as the signal dimension increases, this is computationally not feasible. A discrete stochastic approximation can be used to sequentially search a near-optimal row placement. The algorithm generates a sequence of the estimates of the optimal row subset where the new estimate is based on the previous one by moving a small step in a good direction towards the global optimizer. An occupation probability vector $\mathbf{p}$ is used to keep an estimate of the occupation probability of one state, i.e. a row subset. The algorithm converges to the state with the largest occupation probability in $\mathbf{p}$ [28, 30]. Let $\mathbf{w}_l, \hat{\mathbf{w}}_l$ and $\tilde{\mathbf{w}}_l$ denote the chosen rows at iteration l and the number of all possible placements

are $n_x = \binom{n}{m}$. The state occupation probability at iteration l is given in (2.7).

$$\mathbf{p} = \left\{ [\mathbf{p}_{l,1}, \mathbf{p}_{l,2}, ..., \mathbf{p}_{l,n_x}]^T \in \mathbb{R}^{n_x} \middle| \sum_{i=1}^{n_x} \mathbf{p}_{l,i} = 1, \mathbf{p}_{l,i} \in [0,1] \right\} \tag{2.7}$$

The algorithm starts at a random row position $\mathbf{w}_0$ and initialize $\mathbf{p}_0 = 0$. Given $\mathbf{w}_l$ a new row placement, $\tilde{\mathbf{w}}_l$ is obtained by sequentially changing one row position of $\mathbf{w}_l$. In every iteration only one row position is changed so that the computational complexity of the algorithm is reduced. $\mu(\mathbf{F}_{\tilde{\mathbf{w}}_l})$ is compared to $\mu(\mathbf{F}_{\mathbf{w}_l})$ and then the algorithm moves towards the global optimizer (Figure 2.3). The frequency of the selected row placement is updated via the state occupation probability vector $\mathbf{p}_{l+1} = \mathbf{p}_l + (\mathbf{r}_{l+1} - \mathbf{p}_l)/(l+1)$ where $\mathbf{r}_{l+1} \in \mathbb{R}^{n_x}$ is defined as a zero vector except for its $(l+1)$th component to be 1.

The output of the stochastic algorithm for the undersampled DFT sensing matrix with dimensions $m = 17$ and $n = 273$ is shown in Figure 2.4. For these dimensions the Welch bound (2.3) exists and the mutual coherence of the DFT matrix with the initial randomly selected rows is approximately two times greater than the Welch bound. However, the algorithm decreases this gap by sequentially changing one row position using the state occupation probability vector (Figure 2.3). When the iteration number reaches to 10^2, the decrease in the mutual coherence of the newly seleceted rows become smaller (Figure 2.4).

2.3. Restricted Isometry Property

One of the most used tools in the analysis of sparse recovery is the restricted isometry property (RIP) [31]. For each integer $s = 1, 2, ...$, the restricted isometry constant (RIC) of a matrix $\mathbf{A}$ is defined as the smallest number such that

$$(1 - \delta_s)\|\mathbf{x}\|_2^2 \leq \|\mathbf{A}\mathbf{x}\|_2^2 \leq (1 + \delta_s)\|\mathbf{x}\|_2^2 \tag{2.8}$$

holds for all s-sparse vectors $\mathbf{x} \in \Sigma_s$ [31]. s-sparse vectors cannot be in the null space, $\mathbf{A}\mathbf{x} \neq \mathbf{0}$. An equivalent description of the RIP (2.8) is that all subsets of s columns

Randomly generate $\mathbf{w}_0, \hat{\mathbf{w}} \leftarrow \mathbf{w}_0$
$\mathbf{p}_0 \leftarrow \mathbf{0}, \mathbf{p}_{0,0} \leftarrow 1, u \leftarrow 0, v \leftarrow 0$
for $i = 0$ to $I_{\text{stoc}} - 1$ **do**
  for $k = 0$ to $m - 1$ **do**
    $l \leftarrow iI_{\text{stoc}} + k$
    generate $\tilde{\mathbf{w}}_l \setminus \mathbf{w}_l$
    if $\mu(\mathbf{F}_{\tilde{\mathbf{w}}_l}) < \mu(\mathbf{F}_{\mathbf{w}_l})$ **then**
      $\mathbf{w}_{l+1} \leftarrow \tilde{\mathbf{w}}_l, u \leftarrow l + 1$
    else if $\mu(\mathbf{F}_{\tilde{\mathbf{w}}_l}) \geq \mu(\mathbf{F}_{\mathbf{w}_l})$ **then**
      $\mathbf{w}_{l+1} \leftarrow \mathbf{w}_l$
    end if
    $\mathbf{p}_{l+1} \leftarrow \mathbf{p}_l + (\mathbf{r}_{l+1} - \mathbf{p}_l)/(l + 1)$
    if $\mathbf{p}_{l+1,u} > \mathbf{p}_{l+1,v}$ **then**
      $\mathbf{w}_{l+1} \leftarrow \tilde{\mathbf{w}}_{l+1}, v \leftarrow u$
    else if $\mathbf{p}_{l+1,u} \leq \mathbf{p}_{l+1,v}$ **then**
      $\mathbf{w}_{l+1} \leftarrow \mathbf{w}_l$
    end if
  end for
end for

Figure 2.3. Discrete stochastic approximation algorithm for row allocation.

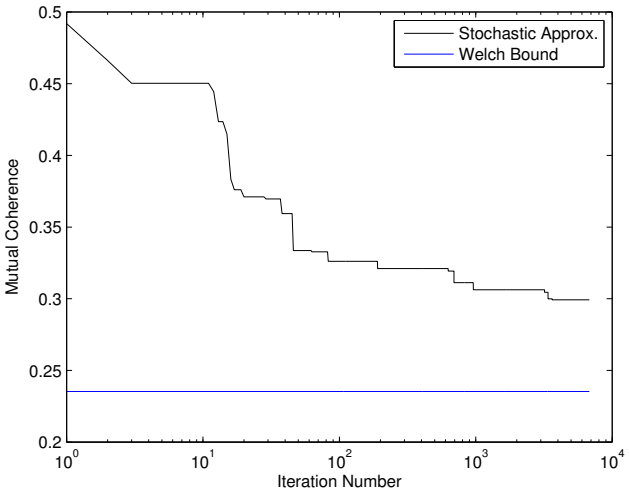


Figure 2.4. Comparison of the algorithm and the Welch bound for $n = 273$.

taken from $\mathbf{A}$ are in fact nearly orthogonal and so

$$(1 - \delta_s)\|\mathbf{c}\|_2^2 \leq \|\mathbf{A}_\Lambda \mathbf{c}\|_2^2 \leq (1 + \delta_s)\|\mathbf{c}\|_2^2 \tag{2.9}$$

must be satisfied for all submatrices $\mathbf{A}_\Lambda$ with cardinality $|\Lambda| \leq s$ and coefficient sequences $\mathbf{c}$. Another result of this property is that the eigenvalues of the Grammian matrix, $\mathbf{A}_\Lambda^* \mathbf{A}_\Lambda$, must be within the interval determined by the RICs (2.10).

$$(1 - \delta_s) \leq \lambda_{\min}(\mathbf{A}_\Lambda^* \mathbf{A}_\Lambda) \leq \lambda_{\max}(\mathbf{A}_\Lambda^* \mathbf{A}_\Lambda) \leq (1 + \delta_s) \tag{2.10}$$

In this context the largest and smallest singular values taken over all s-column submatrices of $\mathbf{A}$ are also related to the RIC as $\|\mathbf{A}\|_2^{(s)} = \sigma_{\max}^{(s)}(\mathbf{A}) \leq \sqrt{1 + \delta_s}$ and $\sigma_{\min}^{(s)}(\mathbf{A}) \geq \sqrt{1 - \delta_s}$. A matrix obeys the RIP of order s if δ_s is not too close to one. The RIC can be equivalently defined as the maximum spectral norm of the difference of the Grammian matrix and the identity matrix [27].

$$\delta_s = \max_{\Lambda \subset \{1,\ldots,n\}, \mathrm{card}(\Lambda) \leq s} \|\mathbf{A}_\Lambda^* \mathbf{A}_\Lambda - \mathbf{I}\|_2 \tag{2.11}$$

Given a matrix $\mathbf{A}$ it is hard to evaluate δ_s for $s \gg 1$ since the computation requires a sweep through all the $\binom{n}{s}$ supports. The RIC can be bounded using the mutual coherence [20] as in (2.12).

$$\delta_s \leq (s-1)\mu(\mathbf{A}) \tag{2.12}$$

For the completely perturbed case where the sensing matrix is also perturbed by the multiplicative noise, $\widetilde{\mathbf{A}} = \mathbf{A} + \mathbf{Z}$, the RIP (2.8) can be redefined as in (2.14).

Theorem 2.3. Given the sparsity level of the signal, the RIC associated with matrix $\mathbf{A}$ and the relative perturbation $\varepsilon_{\mathbf{A}}^{(s)} = \|\mathbf{Z}\|_2^{(s)} / \|\mathbf{A}\|_2^{(s)}$ associated with the perturbation matrix $\mathbf{Z}$, the maximum RIC can be fixed as in (2.13).

$$\tilde{\delta}_{s,\max} = (1+\delta_s)\left(1+\varepsilon_{\mathbf{A}}^{(s)}\right)^2 - 1 \tag{2.13}$$

Then the RIP constant $\tilde{\delta}_s \leq \tilde{\delta}_{s,\max}$ for matrix $\widetilde{\mathbf{A}} = \mathbf{A} + \mathbf{Z}$ is the smallest nonnegative number such that

$$(1-\tilde{\delta}_s)\|\mathbf{x}\|_2^2 \leq \|\widetilde{\mathbf{A}}\mathbf{x}\|_2^2 \leq (1+\tilde{\delta}_s)\|\mathbf{x}\|_2^2 \tag{2.14}$$

holds for any $\mathbf{x} \in \Sigma_s$ [32].

Proof. Assume that $\tilde{\delta}_{s,1}$ and $\tilde{\delta}_{s,2}$ satisfy the RIP for $\widetilde{\mathbf{A}}$

$$(1-\tilde{\delta}_{s,1})\|\mathbf{x}\|_2^2 \leq \|\widetilde{\mathbf{A}}\mathbf{x}\|_2^2 \leq (1+\tilde{\delta}_{s,2})\|\mathbf{x}\|_2^2 \tag{2.15}$$

Using the triangle inequality, the relative perturbation bound $\varepsilon_{\mathbf{A}}^{(s)} = \|\mathbf{Z}\|_2^{(s)} / \|\mathbf{A}\|_2^{(s)}$

and $\|\mathbf{A}\|_2^{(s)} = \sigma_{\max}^{(s)} \leq \sqrt{1+\delta_s}$ the right hand side of (2.15) can be written as:

$$\|\widetilde{\mathbf{A}}\mathbf{x}\|_2^2 \leq [\|\mathbf{A}\mathbf{x}\|_2 + \|\mathbf{Z}\mathbf{x}\|_2]^2 \tag{2.16}$$

$$\leq \left[\sqrt{1+\delta_s} + \|\mathbf{Z}\|_2^{(s)}\right]^2 \|\mathbf{x}\|_2^2 \tag{2.17}$$

$$\leq (1+\delta_s)\left[1+\varepsilon_{\mathbf{A}}^{(s)}\right]^2 \|\mathbf{x}\|_2^2 \tag{2.18}$$

If the right hand side of (2.15) is compared to (2.18) using the fact that $\tilde{\delta}_{s,2}$ must be the smallest nonnegative number in order to satisfy the RIP, then $\tilde{\delta}_{s,2}$ should satisfy (2.19).

$$(1+\tilde{\delta}_{s,2}) \leq (1+\delta_s)\left[1+\varepsilon_{\mathbf{A}}^{(s)}\right]^2 \tag{2.19}$$

When $\mathbf{Z}$ is a positive real valued multiple of $\mathbf{A}$, the inequalities (2.16) and (2.18) become equality. The bound (2.18) can be attained by using (2.20).

$$\tilde{\delta}_{s,2} = (1+\delta_s)\left[1+\varepsilon_{\mathbf{A}}^{(s)}\right]^2 - 1 \tag{2.20}$$

The left hand side of (2.15) is obtained in the same way using similar arguments (2.21).

$$\tilde{\delta}_{s,1} = 1 - (1-\delta_s)\left[1+\varepsilon_{\mathbf{A}}^{(s)}\right]^2 \tag{2.21}$$

Since $(1-\tilde{\delta}_{s,2}) \leq (1-\tilde{\delta}_{s,1})$ and $(1+\tilde{\delta}_{s,1}) \leq (1+\tilde{\delta}_{s,2})$, $\tilde{\delta}_{s,\max}$ can be chosen as

$$\tilde{\delta}_{s,\max} = \tilde{\delta}_{s,2} = (1+\delta_s)\left[1+\varepsilon_{\mathbf{A}}^{(s)}\right]^2 - 1 \tag{2.22}$$

and the proof is finalized [32]. □

The bound (2.13) represents a worst case deviation of $\tilde{\delta}_s$. For a given $\varepsilon_{\mathbf{A}}^{(s)}$ there are infinitely many $\mathbf{Z}$ which satisfy it. It is possible to construct nonzero perturbations which result in $\tilde{\delta}_s = \delta_s$. For example suppose $\widetilde{\mathbf{A}} = \mathbf{A}\mathbf{U}$ for some unitary matrix $\mathbf{U} \neq \mathbf{I}$. Here $\mathbf{Z} = \mathbf{A}(\mathbf{U}-\mathbf{I}) \neq \mathbf{0}$ and yet $\tilde{\delta}_s = \delta_s$. In such cases more information on $\mathbf{Z}$ is

needed to estimate more accurate value of $\varepsilon_{\mathbf{A}}^{(s)}$ [32].

2.4. Null Space Property

A matrix is said to satisfy the null space property (NSP) of order s, if for all vectors $\mathbf{x}$ in the nullspace, $\mathbf{Ax} = \mathbf{0}$, it holds that

$$\|\mathbf{x}\|_{s,1} \leq \alpha_s \|\mathbf{x}\|_1 \tag{2.23}$$

where $\|\mathbf{x}\|_{s,1}$ denotes the sum of the s largest absolute values of entries in $\mathbf{x}$. α_s is a constant and for $\alpha_s < 1/2$ exact recovery of s sparse solutions are guaranteed [23, 33].

The NSP (2.23) is also related to the RIP (2.8) as presented in the following theorem.

Theorem 2.4. Assume that $\mathbf{A}$ satisfies the RIP of sparsity order $S = s + s^{'}$ with RIC constant $\delta_S \in (0, 1)$. Then $\mathbf{A}$ has the NSP (2.23) of sparsity order s with the NSP constant given as $\alpha_s = \sqrt{[s(1+\delta_S)]/[s^{'}(1-\delta_S)]}$ [27].

Proof. Let $\mathbf{x} \in \{\mathbf{x} \,|\, \mathbf{Ax} = 0\}$, $\Lambda \subset \{1, \ldots, n\}$ and card$\{\Lambda\} \leq s$. Define $\Lambda_0 = \Lambda$ and $\Lambda_1, \ldots, \Lambda_r$ to be disjoint sets of index size at most $s^{'}$ associated to a nonincreasing rearrangement of the entries of $\mathbf{x}$ (2.24).

$$|\mathbf{x}[l]| \leq |\mathbf{x}[k]| \qquad l \in \Lambda_{l'},\, k \in \Lambda_{k'},\, l^{'} \geq k^{'} \geq 1 \tag{2.24}$$

$\mathbf{Ax} = 0$ implies that $\mathbf{Ax}_{\Lambda_0 \cup \Lambda_1} = -\sum_{l=2}^{r} \mathbf{Ax}_{\Lambda_l}$. Then using the Cauchy-Schwarz inequality, the RIP (2.8) and the triangle inequality, the following sequence of inequalities

can be deduced:

$$\begin{aligned}\|\mathbf{x}_\Lambda\|_1 &\leq \sqrt{s}\|\mathbf{x}_\Lambda\|_2 \leq \sqrt{s}\|\mathbf{x}_{\Lambda_0\cup\Lambda_1}\|_2 \\ &\leq \sqrt{\frac{s}{1-\delta_S}}\|\mathbf{A}\mathbf{x}_{\Lambda_0\cup\Lambda_1}\|_2 = \sqrt{\frac{s}{1-\delta_S}}\|\mathbf{A}\mathbf{x}_{\Lambda_2\cup\cdots\cup\Lambda_r}\|_2 \\ &\leq \sqrt{\frac{s}{1-\delta_S}}\sum_{l=2}^{r}\|\mathbf{A}\mathbf{x}_{\Lambda_l}\|_2 \leq \sqrt{\frac{1+\delta_S}{1-\delta_S}}\sqrt{s}\sum_{l=2}^{r}\|\mathbf{x}_{\Lambda_l}\|_2 \end{aligned} \tag{2.25}$$

It follows from (2.24) that $|\mathbf{x}[l]| \leq |\mathbf{x}[k]|$ for all $l \in \Lambda_{l'+1}$ and $k \in \Lambda_{l'}$. Taking the sum over $k \in \Lambda_{l'}$ and then the ℓ_2 norm over $l \in \Lambda_{l'+1}$ yields $|\mathbf{x}[l]| \leq (s')^{-1}\|\mathbf{x}_{\Lambda_{l'}}\|_1$, and $\|\mathbf{x}_{\Lambda_{l'+1}}\|_2 \leq (s')^{-1/2}\|\mathbf{x}_{\Lambda_{l'}}\|_1$. Using (2.25) gives

$$\|\mathbf{x}_\Lambda\|_1 \leq \sqrt{\frac{s}{s'}\frac{(1+\delta_S)}{(1-\delta_S)}}\sum_{l=1}^{r-1}\|\mathbf{x}_{\Lambda_l}\|_1 \leq \sqrt{\frac{s}{s'}\frac{(1+\delta_S)}{(1-\delta_S)}}\|\mathbf{x}_{\Lambda^c}\|_1 \tag{2.26}$$

and the proof is completed. □

Some of the most important CS stability analysis tools; spark, mutual coherence, RIP and NSP are covered in this chapter. While spark, RIP and NSP are computationally inefficient for higher signal dimensions, mutual coherence can be computed for any given sensing matrix. Although it is possible to use mutual coherence to give bounds for the other tools such as spark and RIP, these bounds are not tight.

3. RECONSTRUCTION METHODS

Since there exists a wide variety of sparse recovery methods, the algorithms that are given in this chapter do not cover all of them but they include some of the important well-established methods. Sparse reconstruction methods can be presented in three categories: greedy pursuits, convex relaxation methods, and Bayesian compressed sensing.

3.1. Greedy Pursuits

Greedy pursuits constructs an approximation to the sparse signal of interest one term at a time by making locally optimal choices at each step of the algorithm. These methods selects the column of the sensing matrix which correlates most strongly with the residual signal in a greedy fashion. Greedy pursuits consists of the orhogonal matching pursuit (OMP) [34–36] algorithm and modified versions of the OMP. Compressive sampling matching pursuit (COSAMP) [37], regularized OMP [38, 39] and stagewise OMP [40] are some of the examples of greedy pursuits that are based on OMP.

3.1.1. Orthogonal Matching Pursuit

In order to identify the signal $\mathbf{x}$, the columns of $\mathbf{A}$ which participate in the measurement vector $\mathbf{y}$ must be determined. The idea of OMP is to pick columns of $\mathbf{A}$ in a greedy fashion. OMP algorithm iteratively chooses the columns of the matrix by selecting the column whose correlation with the remaining part of $\mathbf{y}$ has the maximum value [36]. Then the contribution of the chosen column is subtracted from $\mathbf{y}$ and the new residual is used for the next iteration. At each iteration the new signal estimate is calculated by least squares projection which can be implemented with Gram-Schmidt scheme. The Gram-Schmidt orthogonalization procedure uses the selected column to create an orthogonal set of vectors. A new residual is computed as the difference between the previous residual and the projection of the previous residual onto the new vector obtained through Gram-Schmidt scheme (Figure 3.1). The algorithm is

expected to identify the correct support set of the signal $\mathbf{x}$ in s iterations.

OMP algorithm initializes by setting the residual equal to the measurement vector. Then OMP attempts to reconstruct the support set of $\mathbf{x}$ iteratively by starting with an empty set of selected columns $\Lambda = \emptyset$ (Figure 3.1). In the i-th iteration, a new element which denotes the column of $\mathbf{A}$ that has the largest absolute value of inner products with the residual $\mathbf{r}$ is added to Λ. The residual is updated at the end of each iteration by representing the components of the measurements $\mathbf{y}$ that cannot be represented by $\mathbf{A}_\Lambda$. The computational complexity of the OMP algorithm depends upon whether the sensing matrix, $\mathbf{A}$, is a dense matrix that can only be implemented with regular matrix multiplication or a matrix that can be implemented with a faster matrix multiplication. For dense sensing matrices and iteration number set to sparsity, $I_{\max} = s$, the total complexity of the OMP algorithm is $(4/3)s^3 + smn + \mathrm{O}(n)$ at most. If the sensing matrix allows fast matrix multiplication with computational cost $\mathcal{C}_\mathbf{A}$, then the complexity of the OMP method is $(4/3)s^3 + 2s\mathcal{C}_\mathbf{A} + \mathrm{O}(n)$. The running time of the OMP algorithm is dominated by the Gram-Schmidt procedure which has a marginal cost of $\mathrm{O}(im)$ for i iterations and can be maintained with Cholesky factorization [36].

3.1.2. Flexible Tree Search Based OMP

Although the performance of OMP algorithm is proven in various problems [36], it may not always provide the sparsest solution when an erroneous decision is made on the initial iterations. This shortcoming of the OMP algorithm prevents the algorithm to find a good approximation to the signal of interest and leads to suboptimum representations. The tree search based techniques can be integrated to matching pursuit (MP) algorithms in order to avoid such error propagations [41]. These algorithms are referred as tree search based matching pursuit (TB-MP) algorithms and they can greatly improve the approximation and detection performance of the MP algorithms at the expense of increased computational cost.

Such a tree search based algorithm, referred to as flexible tree search based orthogonal matching pursuit (FTB-OMP), is proposed in [42]. Through the design parame-

Input: $\mathbf{y}$, $\mathbf{A}$, $I_{\max}$

Initial State: $\mathbf{r}_0 = \mathbf{y}, \Lambda_0 = \emptyset$

for $i = 1$ to $I_{\max}$ **do**

$\boldsymbol{\varphi} \leftarrow \mathbf{A}^* \mathbf{r}_{i-1}$

$l \leftarrow \arg \max_{k \in \{1,\ldots,n\}} |\varphi_k|$

$\Lambda \leftarrow \Lambda \cup \{l\}$

$\mathbf{q}_i \leftarrow \mathbf{a}_l - \sum_{k=0}^{i-1} \langle \mathbf{a}_l, \mathbf{q}_k \rangle \mathbf{q}_k$

$\mathbf{q}_i \leftarrow \mathbf{q}_i / \|\mathbf{q}_i\|_2$

$\mathbf{r}_i \leftarrow \mathbf{r}_{i-1} - \langle \mathbf{r}_{i-1}, \mathbf{q}_i \rangle \mathbf{q}_i$

end for

$\hat{\mathbf{x}} \leftarrow \arg\min \|\mathbf{y} - \mathbf{A}_\Lambda \mathbf{x}\|_2$

Output: $\Lambda, \hat{\mathbf{x}}$

Figure 3.1. Orthogonal Matching Pursuit Algorithm.

ters, FTB-OMP provides the flexibility to establish a tradeoff between the performance of the reconstruction and the size of the search tree which effects the computational complexity. FTB-OMP is implemented by substituting a single vector selection at each iteration by considering ϑ number of possible vectors. In this way the greedy pursuit is substituted by a search on a tree with the branching factor, ϑ. The complexity of the search may be decreased by reducing the number of children as the depth of the search tree increases. In addition the size of the search tree can be further controlled by employing a correlation based pruning. Branch pruning is controlled by a decaying parameter, ι, which affects the tree structure by reducing the branching factor at each iteration. A maximum number of branches are searched for each node in order to find a partial solution. In the initial iteration, the branching factor is set to the given parameter value and then for each iteration the algorithm recursively calls itself with the new branching factor given as $\lceil \vartheta / \iota^{(i-1)} \rceil$ (Figure 3.2). The idea is to construct the search tree with a large number of branches at the initial iteration where an erroneous selection is more likely to appear, and then reduce the branching factor as the number of iterations increases. Further reduction on the tree size may be achieved by

using a correlation threshold. This may help in reducing the size of the search tree by prunnig the tree branches that are heuristically believed to be unnecessary and keep only the branches among which are highly correlated with OMP's first choice branch. At the output of the algorithm, the indices of the selected vectors representing the measurement signal in the most accurate manner provide the support of the signal.

Input: $\mathbf{y}$, $\mathbf{A}$, $\mathbf{Q}$, Λ, ϑ, ι, $I_{\max}$

Initial State: $\mathbf{r} = \mathbf{y}$, $\mathbf{\Phi} = \mathbf{A}$, $i = 1$, $\Lambda = \emptyset$

$\boldsymbol{\varphi} \leftarrow \mathbf{A}^{*}\mathbf{r}$

$(\sim, I) \leftarrow \text{sort}\,(\varphi_k)$

if $i < I_{\max}$ **then**

  for $l = 1$ to $\lceil \vartheta / \iota^{(i-1)} \rceil$ **do**

    $\mathbf{q}_{\text{new}} \leftarrow \mathbf{a}_{I_l} - \mathbf{Q}\mathbf{Q}^{*}\mathbf{a}_{I_l}$

    $\mathbf{q}_{\text{new}} \leftarrow \mathbf{q}_{\text{new}} / \left\|\mathbf{q}_{\text{new}}\right\|_2$

    $\mathbf{r}_{\text{new}} \leftarrow \mathbf{r} - \langle \mathbf{r}, \mathbf{q}_{\text{new}} \rangle \, \mathbf{q}_{\text{new}}$

    $\mathbf{Q}_{\text{new}} \leftarrow [\, \mathbf{Q} \; \mathbf{q}_{\text{new}} \,]$

    $\Lambda \leftarrow \Lambda \cup \{I_l\}$

    $(\mathbf{r}_l, \Lambda_l) \leftarrow$ **FTB-OMP** $(\mathbf{r}, \mathbf{\Phi}, \mathbf{Q}_{\text{new}}, \Lambda, \vartheta, \iota, i+1)$

    $\mathbf{e}_l \leftarrow \left\|\mathbf{r}_l\right\|_2$

  end for

  $p \leftarrow \arg\min e_k$

  $\mathbf{r} \leftarrow \mathbf{r}_p$

else if $i \geq I_{\max}$ **then**

  $\mathbf{q} \leftarrow \mathbf{a}_{I_1} - \mathbf{Q}\mathbf{Q}^{*}\mathbf{a}_{I_1}$

  $\mathbf{q} \leftarrow \mathbf{q} / \left\|\mathbf{q}\right\|_2$

  $\mathbf{r} \leftarrow \mathbf{r} - \langle \mathbf{r}, \mathbf{q} \rangle \, \mathbf{q}$

  $\Lambda \leftarrow \Lambda \cup \{I_1\}$

  Return

end if

Figure 3.2. Flexible tree search based orthogonal matching pursuit algorithm.

3.1.3. Block Orthogonal Matching Pursuit

Sparse signals may have additional structure in the form of nonzero coefficients occurring in clusters. Such signals are referred to as block-sparse signals and they can be encountered in multi-band signals, measurements of gene expression levels, and multiple measurement vector problems [43, 44]. Block-sparse signals are concatenation of n_b number of blocks with length l_b

$$\mathbf{x} = [x_1 \ \dots \ x_{l_\mathrm{b}} \, x_{l_\mathrm{b}+1} \ \dots \ x_{2l_\mathrm{b}} \ \dots \ x_{n-l_\mathrm{b}+1} \dots \ x_n]^T \tag{3.1}$$

where $n = n_\mathrm{b} l_\mathrm{b}$ and $\mathbf{x}[i]$ denotes the i-th block $\mathbf{x}[i] = [x_{(i-1)l_\mathrm{b}+1} \dots x_{il_\mathrm{b}}]^T$. The sensing matrix for block sparse signals is similarly represented as concatenation of column blocks (3.2).

$$\mathbf{A} = [\mathbf{a}_1 \ \dots \ \mathbf{a}_{l_\mathrm{b}} \, \mathbf{a}_{l_\mathrm{b}+1} \ \dots \, \mathbf{a}_{2l_\mathrm{b}} \ \dots \, \mathbf{a}_{n-l_\mathrm{b}+1} \dots \, \mathbf{a}_n] \tag{3.2}$$

$\mathbf{A}[k]$ denotes the k-th block matrix $\mathbf{A}[k] = [\mathbf{a}_{(k-1)l_\mathrm{b}+1} \dots \mathbf{a}_{kl_\mathrm{b}}]$. A block s-sparse signal is defined as a vector that satisfies $\|\mathbf{x}\|_{2,0} \leq s$. $\|\mathbf{x}\|_{2,0}$ is defined as

$$\|\mathbf{x}\|_{2,0} = \sum_{i=1}^{n_\mathrm{b}} \mathbb{I}\left(\|\mathbf{x}\,[i]\|_2 > 0\right) \tag{3.3}$$

where $\mathbb{I}(\cdot)$ is the indicator function which is zero when its argument is zero and is one otherwise. The indicator function counts the number of nonzero blocks in the signal. The problem of finding a sparse representation which uses the minimum number of blocks of the sensing matrix is given as the following optimization problem (3.4).

$$\begin{aligned} &\text{minimize} \quad \|\mathbf{x}\|_{2,0} \\ &\text{subject to} \quad \mathbf{y} = \mathbf{A}\mathbf{x} \end{aligned} \tag{3.4}$$

However, solving (3.4) is an NP hard problem since it involves searching over all choices of a few blocks of $\mathbf{A}$ [43, 44]. A number of algorithms such as Group Lasso [45], mixed

ℓ_2/ℓ_1 program [46], block orthogonal matching pursuit (BOMP) algorithm [43], block sparse Bayesian learning [47], block iteratively reweighted least squares (BIRLS) [48] and block iterative support detection (block-ISD) [16] have been shown to recover block sparse signals.

BOMP algorithm is an extension of the OMP algorithm used in conventional sparse recovery. The steps of the BOMP algorithm are shown in Figure 3.3. First step of the BOMP is to initialize the residual $\mathbf{r} = \mathbf{y}$ [43]. At each iteration the block that is best matched to the residual from the previous iteration is chosen according to (3.5).

$$l = \arg \max_{k=1,\ldots,n_{\mathrm{b}}} \|\mathbf{A}[k]^{*}\mathbf{r}\|_2 \tag{3.5}$$

Once the block index l is chosen, the signal estimate $\hat{\mathbf{x}}_i$ is computed as

$$\hat{\mathbf{x}}_i = \arg\min \left\| \mathbf{y} - \sum_{k\in\Lambda} \mathbf{A}[k]\mathbf{x}[k] \right\|_2 = \mathbf{A}_{\Lambda}^{\dagger}\mathbf{y} = (\mathbf{A}_{\Lambda}^{*}\mathbf{A}_{\Lambda})^{-1}\mathbf{A}_{\Lambda}^{*}\mathbf{y} \tag{3.6}$$

where Λ is the set of indices $\Lambda = \{l_1, ..., l_i\}$ chosen up to iteration i. Then the residual is updated as in (3.7).

$$\mathbf{r} = \mathbf{r} - \mathbf{A}\hat{\mathbf{x}}_i \tag{3.7}$$

The computational complexity of the algorithm for 1 block-sparse signals is O(mnd). The BOMP method [43] selects the block that is most correlated with the current residual and then applies LS. So at any given iteration of the BOMP method, the solution is guaranteed to be i block-sparse.

The BOMP algorithm is shown to recover any s-block sparse vector $\mathbf{x}$ if the following condition [43] is satisfied

$$sl_{\mathrm{b}} < \frac{\mu_B^{-1}(\mathbf{A}) + l_{\mathrm{b}} - (l_{\mathrm{b}} - 1)\nu\mu_B^{-1}(\mathbf{A})}{2} \tag{3.8}$$

Input: $\mathbf{A}$, $\mathbf{y}$, s

Initial State: $\Lambda = \emptyset$ and $\mathbf{r} = \mathbf{y}$

for $i = 1$ to s **do**

$l \leftarrow \arg \max_{k=1,\ldots,n_{\mathrm{b}}} \|\mathbf{A}[k]^{*}\mathbf{r}\|_2$

$\Lambda \leftarrow \Lambda \cup \{l\}$

$\hat{\mathbf{x}}_i \leftarrow \mathbf{A}_{\Lambda}^{\dagger}\mathbf{y}$

$\mathbf{r} \leftarrow \mathbf{r} - \mathbf{A}\hat{\mathbf{x}}_i$

end for

Output: $\hat{\mathbf{x}}$

Figure 3.3. Block Orthogonal Matching Pursuit Algorithm.

where b represents the block length for $\mathbf{A}$ and $\mu_B(\mathbf{A})$ denotes the block-coherence

$$\mu_B(\mathbf{A}) = \max_{l,k \neq l} \frac{1}{n} \sigma_{\max}(\mathbf{A}[l]^{*}\mathbf{A}[k]) \tag{3.9}$$

and ν represents the subcoherence of $\mathbf{A}$ as $\nu = \max_i \mu(\mathbf{A}[i])$. If the columns of $\mathbf{A}[i]$ are orthonormal for each i, then the subcoherence is zero.

3.2. Convex Relaxation by ℓ_1 Minimization

Convex relaxation methods are based on the equivalance between ℓ_1-norm and ℓ_0-pseudo-norm under certain conditions. Another effective method which is called the least absolute shrinkage and selection operator (LASSO) [49] is the solution to the following optimization problem (3.10).

$$\begin{aligned} &\text{minimize} && \|\mathbf{x}\|_1 \\ &\text{subject to} && \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \leq \varepsilon \end{aligned} \tag{3.10}$$

An unconstrained adaptation of the LASSO (3.10) can be given as

$$\text{minimize} \quad \frac{1}{2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2+\zeta\|\mathbf{x}\|_1 \tag{3.11}$$

which the hard constraint $\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2 \leq \varepsilon$ is replaced with a quadratic penalty scaled by $1/2\zeta$ as $(1/2\zeta)\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2$. This problem is also known as basis pursuit (BP) or basis pursuit denoising [50].

3.2.1. Interior Point Methods

The LASSO (3.10) or the BP (3.11) problem is also known as ℓ_1-regularized least squares problem (LSP) and there is no analytic formula or expression for the optimal solution to ℓ_1-regularized LSP [51]. In this case the solution must be computed numerically. These problems can be transformed to a convex quadratic program (QP) for real data or to a second order cone program (SOCP) [52] for complex data. Both QP and SOCP are solved by standard convex optimization methods called interior-point methods [53]. Interior-point methods are implemented in general purpose solvers such as the CVX package [54, 55] for the MATLAB programming environment. In order to apply an interior point method to a SOCP, the first step is to reformulate (3.10) or (3.11) as a convex problem

$$\begin{aligned} \text{minimize} \quad & \left\| \mathbf{A}^{'} \begin{bmatrix} \mathbf{x}_{\Re} \\ \mathbf{x}_{\Im} \end{bmatrix} - \mathbf{y}^{'} \right\|_2^2 + \zeta\, \mathbf{1}^T\mathbf{u} \\ \text{subject to} \quad & \sqrt{(\mathbf{x}_{\Re}[i])^2+(\mathbf{x}_{\Im}[i])^2} \leq \mathbf{u}[i] \end{aligned} \tag{3.12}$$

where $i = 1, ..., n$. The variables are $\mathbf{x}_{\Re} = \Re\{\mathbf{x}\}$, $\mathbf{x}_{\Im} = \Im\{\mathbf{x}\}$, $\mathbf{u} \in \mathbf{R}^n$ and the new problem data are shown in (3.13).

$$\mathbf{A}^{'} = \begin{bmatrix} \Re\{\mathbf{A}\} & -\Im\{\mathbf{A}\} \\ \Im\{\mathbf{A}\} & \Re\{\mathbf{A}\} \end{bmatrix} \in \mathbf{R}^{2m\times 2n} \qquad \mathbf{y}^{'} = \begin{bmatrix} \Re\{\mathbf{y}\} \\ \Im\{\mathbf{y}\} \end{bmatrix} \in \mathbf{R}^{2m} \tag{3.13}$$

Using the standard barrier function for second order constraints, the convex function to be minimized is given as

$$\phi_t(\mathbf{x}_\Re, \mathbf{x}_\Im, \mathbf{u}) = t \left\| \mathbf{A}^{'} \begin{bmatrix} \mathbf{x}_\Re \\ \mathbf{x}_\Im \end{bmatrix} - \mathbf{y}^{'} \right\|_2^2 + t\,\zeta\,\mathbf{1}^T\mathbf{u} + \Phi(\mathbf{x}_\Re, \mathbf{x}_\Im, \mathbf{u}) \tag{3.14}$$

where $\Phi(\mathbf{x}_\Re, \mathbf{x}_\Im, \mathbf{u})$ is the logarithmic barrier function defined in (3.15).

$$\Phi(\mathbf{x}_\Re, \mathbf{x}_\Im, \mathbf{u}) = -\sum_{i=1}^{n} \log\left((\mathbf{u}[i])^2 - (\mathbf{x}_\Re[i])^2 - (\mathbf{x}_\Im[i])^2\right) \tag{3.15}$$

An optimal solution is achieved by computing a sequence of points $(\mathbf{x}_\Re, \mathbf{x}_\Im, \mathbf{u})$ on the central path (3.14) for an increasing sequence of values of t. A typical method is to select $t = t_0, \varrho t_0, \varrho^2 t_0, ...$ where ϱ is chosen within the range $[2, 50]$ [51]. The search direction of the barrier method is found via solving the Newton equation (3.16).

$$\mathbf{H} \begin{bmatrix} \Delta\mathbf{x}_\Re \\ \Delta\mathbf{x}_\Im \\ \Delta\mathbf{u} \end{bmatrix} = -\mathbf{g} \tag{3.16}$$

The Hessian can be computed as $\mathbf{H} = \nabla^2\phi_t(\mathbf{x}_\Re, \mathbf{x}_\Im, \mathbf{u})$, while the gradient computation is given as $\mathbf{g} = \nabla\phi_t(\mathbf{x}_\Re, \mathbf{x}_\Im, \mathbf{u})$. The exact solution of (3.16) gives the search direction for the sequence of points on the central path. The computational complexity of obatining the exact solution of the Newton system is rather high. Instead an approximate solution using preconditioned conjugate gradients is often preferred. This overall method is called the truncated Newton method and its algorithm is shown in Figure 3.4. Given the search direction, the new point is calculated using the step size parameter. The parameter t is updated at the end of each iteration. For dense sensing matrices the interior point methods can be implemented with computational complexity $I_{\text{cont}}(2n)^3/3 + \mathrm{O}(mn)$. The computational complexity with fast matrix multiplication, $\mathcal{C}_\mathbf{A}$, can be reduced to $2n\mathcal{C}_\mathbf{A} + \mathrm{O}(mn)$.

Input: $\varepsilon_{\mathrm{r}} > 0$

Initial State: $t = 1/\zeta, \mathbf{x} = \mathbf{0}, \mathbf{u} = \mathbf{1}$

for $i = 1$ to I_{cont} **do**

Compute the search direction, $(\Delta\mathbf{x}_{\Re}, \Delta\mathbf{x}_{\Im}, \Delta\mathbf{u})$, by solving (3.16)

Compute the step size by backtracking line search

Update the iterate as $(\mathbf{x}_{\Re}, \mathbf{x}_{\Im}, \mathbf{u}) = (\mathbf{x}_{\Re}, \mathbf{x}_{\Im}, \mathbf{u}) + s(\Delta\mathbf{x}_{\Re}, \Delta\mathbf{x}_{\Im}, \Delta\mathbf{u})$

Construct a dual feasible point of the Lagrange dual problem

if Duality gap $\leq \varepsilon_{\mathrm{r}}$ **then**

Quit

end if

Update t

end for

Figure 3.4. Custom interior point method.

3.2.2. Nesterov's Algorithm

Nesterov's algorithm also known as NESTA [56, 57] is a first-order method with an improved gradient method aiming to improve the speed, accuracy and robustness of ℓ_1 regularized LSP (Section 3.2.1). NESTA is based on an algorithm to minimize any smooth convex function f on the primal feasible convex set $\mathcal{Q}_p$ (3.17).

$$\min_{\mathbf{x}\in\mathcal{Q}_p} f(\mathbf{x}) \tag{3.17}$$

The function is assumed to be differentiable and its gradient $\nabla f(\mathbf{x})$ is Lipschitz

$$\|\nabla f(\mathbf{x}_1) - \nabla f(\mathbf{x}_2)\|_2 \leq L_\varrho \|\mathbf{x}_1 - \mathbf{x}_2\|_2 \tag{3.18}$$

for $\mathbf{x}_1$ and $\mathbf{x}_2$ with Lipschitz constant, L_ϱ. For CS problems NESTA minimizes the smooth approximation to the ℓ_1 norm

$$f_\varrho(\mathbf{x}) = \max_{\mathbf{u}\in\mathcal{Q}_d} \langle \mathbf{u}, \mathbf{x}\rangle - (\varrho/2)\|\mathbf{u}\|_2^2 \tag{3.19}$$

over $\mathcal{Q}_p = \{\mathbf{x} : \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \leq \varepsilon\}$ where the dual feasible set is given as the infinity ball, $\mathcal{Q}_d = \{\mathbf{u} : \|\mathbf{u}\|_\infty \leq 1\}$. f_ϱ in (3.19) is the Huber function and its gradient, $\nabla f_\varrho(\mathbf{x})$, is Lipschitz with constant $L_\varrho = 1/\varrho$. The gradient can be computed as in (3.20).

$$\nabla f_\varrho(\mathbf{x})[i] = \begin{cases} \varrho^{-1}\mathbf{x}[i] & \text{if } |\mathbf{x}[i]| < \mu \\ \operatorname{sgn}(\mathbf{x}[i]) & \text{otherwise} \end{cases} \tag{3.20}$$

The steps of the NESTA algorithm is shown in Figure 3.5. The NESTA algorithm makes use of the continuation and iteratively solves a sequence of CS problems with decreasing smoothing parameters, $\varrho_0 \geq \cdots \geq \varrho_{I_{\text{cont}}}$. The solution of each problem is used as a initial solution for the next problem in the sequence. The number of continuation steps is denoted as I_{cont}. For each continuation step, NESTA iteratively estimates three sequences $\{\mathbf{x}_l\}$, $\{\hat{\mathbf{x}}_l\}$, and $\{\mathbf{v}_l\}$. NESTA computes the current guess of the optimal solution by solving (3.21).

$$\hat{\mathbf{x}}_l = \arg\min_{\mathbf{x}\in Q_p} \frac{L_\varrho}{2}\|\mathbf{x}_l - \mathbf{x}\|_2^2 + \langle \nabla f_\varrho(\mathbf{x}_l), \mathbf{x} - \mathbf{x}_l\rangle \tag{3.21}$$

After applying the Karush-Kuhn-Tucker (KKT) conditions to the Lagrangian of this problem, the primal-dual solution $(\hat{\mathbf{x}}_l, \zeta_\varepsilon)$ can be found as in (3.22).

$$\begin{aligned} \hat{\mathbf{x}}_l &= \left(\mathbf{I} - \frac{\zeta_\varepsilon \mathbf{A}^*\mathbf{A}}{\zeta_\varepsilon + L_\varrho}\right)\left(\frac{\zeta_\varepsilon \mathbf{A}^*\mathbf{y}}{L_\varrho} + \mathbf{x}_l - \frac{1}{L_\varrho}\nabla f_\varrho(\mathbf{x}_l)\right) \\ \zeta_\varepsilon &= \max\left(0, \varepsilon^{-1}\left\|\mathbf{y} - \mathbf{A}\left[\mathbf{x}_l - L_\varrho^{-1}\nabla f_\varrho(\mathbf{x}_l)\right]\right\|_2 - L_\varrho\right) \end{aligned} \tag{3.22}$$

It is assumed that $\mathbf{A}^*\mathbf{A}$ is an orthogonal projector and the rows of $\mathbf{A}$ are orthonormal. NESTA makes use of the previous iterations by computing the sequence $\mathbf{v}_l$ which is

the solution to (3.23).

$$\mathbf{v}_l = \arg\min_{\mathbf{x}} \frac{L_\varrho}{2}\|\mathbf{x}-\mathbf{x}_0\|_2^2 + \frac{\zeta}{2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 + \langle \sum_{k\leq l} \upsilon_k \nabla f_\varrho(\mathbf{x}_k), \mathbf{x}-\mathbf{x}_l\rangle \tag{3.23}$$

KKT conditions of this problem gives the primal-dual solution $(\mathbf{v}_l, \zeta_\varepsilon)$ (3.24).

$$\begin{aligned} \mathbf{v}_l &= \left(\mathbf{I} - \frac{\zeta_\varepsilon \mathbf{A}^*\mathbf{A}}{\zeta_\varepsilon + L_\varrho}\right)\left(\frac{\zeta_\varepsilon \mathbf{A}^*\mathbf{y}}{L_\varrho} + \mathbf{x}_0 - \frac{1}{L_\varrho}\sum_{k\leq l}\upsilon_k \nabla f_\varrho(\mathbf{x}_k)\right) \\ \zeta_\varepsilon &= \max\left(0, \varepsilon^{-1}\left\|\mathbf{y} - \mathbf{A}\left(\mathbf{x}_0 - L_\varrho^{-1}\sum_{j\leq k}\upsilon_j \nabla f_\varrho(\mathbf{x}_j)\right)\right\|_2 - L_\varrho\right) \end{aligned} \tag{3.24}$$

The sequence $\mathbf{v}_l$ keeps memory of the previous iterations since it is calculated using a weighted cumulative gradient $\sum_{k\leq l}\upsilon_k \nabla f_\varrho(\mathbf{x}_k)$. The weight υ_l can be chosen as $\upsilon_l = 1/2(l+1)$ [56]. In the last step of NESTA, the current point $\mathbf{x}_l$ is updated using a weighted average between $\mathbf{v}_l$ and $\hat{\mathbf{x}}_l$ (3.25).

$$\mathbf{x}_{l+1} = \tau_l \mathbf{v}_l + (1-\tau_l)\hat{\mathbf{x}}_l \tag{3.25}$$

The weight τ_l can be chosen as $\tau_l = 2/(l+3)$ which provides a faster convergence rate than standard gradient-based optimization techniques according to the theoretical analysis [56]. Stopping criteria of the NESTA algorithm is to check whether the relative variation of f_ϱ is small enough. For this purpose Δf_ϱ is defined as in (3.26).

$$\Delta f_\varrho = \frac{|f_\varrho(\mathbf{x}_l) - \frac{1}{\min\{10,l\}}\sum_{k=1}^{\min\{10,l\}} f_\varrho(\mathbf{x}_{l-k})|}{\frac{1}{\min\{10,l\}}\sum_{k=1}^{\min\{10,l\}} f_\varrho(\mathbf{x}_{l-k})} \tag{3.26}$$

The convergence can be claimed when $\Delta f_\varrho < \varepsilon_{\mathrm{g}}$. ε_{g} can be chosen from $\varepsilon_{\mathrm{g}} \in [10^{-5}, 10^{-8}]$ depending on the desired accuracy [56]. The complexity of each iteration of the NESTA method is dominated by three steps. The computational complexity of calculating ∇f_ϱ is $\mathrm{O}(n)$. Then the computation of both $\hat{\mathbf{x}}_l$ and $\mathbf{v}_l$ requires the application of the matrix $\mathbf{A}$ three times for each of them. The total cost of finding them is $6\,\mathcal{C}_{\mathbf{A}}$ with $\mathcal{C}_{\mathbf{A}}$ denoting the cost of the matrix multiplication for $\mathbf{A}$. Thus the complexity of a single NESTA

iteration is $6\,\mathcal{C}_{\mathbf{A}} + \mathrm{O}(n)$ where $\mathcal{C}_{\mathbf{A}}$ is dominant.

Input: $\mathbf{y}$, $\mathbf{A}$

Initial State: $\mathbf{x}_\varrho = \mathbf{A}^*\mathbf{y}$, $\varrho_0 = 0.9\|\mathbf{A}^*\mathbf{y}\|_\infty$

for $i = 1$ to I_{cont} **do**

$\mathbf{x}_0 \leftarrow \mathbf{x}_\varrho$

for $l = 0$ to $I_{\max}$ **do**

$$\hat{\mathbf{x}}_l = \left(\mathbf{I} - \frac{\zeta_\varepsilon \mathbf{A}^*\mathbf{A}}{\zeta_\varepsilon + L_\varrho}\right)\left(\frac{\zeta_\varepsilon \mathbf{A}^*\mathbf{y}}{L_\varrho} + \mathbf{x}_l - \frac{1}{L_\varrho}\nabla f_\varrho(\mathbf{x}_l)\right)$$

$$\mathbf{v}_l = \left(\mathbf{I} - \frac{\zeta_\varepsilon \mathbf{A}^*\mathbf{A}}{\zeta_\varepsilon + L_\varrho}\right)\left(\frac{\zeta_\varepsilon \mathbf{A}^*\mathbf{y}}{L_\varrho} + \mathbf{x}_0 - \frac{1}{L_\varrho}\sum_{k\le l} \upsilon_k \nabla f_\varrho(\mathbf{x}_k)\right)$$

$\mathbf{x}_{l+1} \leftarrow \tau_l \mathbf{v}_l + (1-\tau_l)\hat{\mathbf{x}}_l$

$\mathbf{x}_\varrho \leftarrow \mathbf{x}_{l+1}$

if $\Delta f_\varrho < \varepsilon_{\text{g}}$ **then**

Return

end if

end for

$\varrho_{i+1} \leftarrow \gamma \varrho_i$

end for

Output: $\hat{\mathbf{x}}$

Figure 3.5. Nesterov's Algorithm (NESTA).

3.3. Bayesian Compressive Sensing

The Gaussian likelihood model for the CS measurement vector under zero mean additive Gaussian noise with variance σ^2 is given in (3.27).

$$p(\mathbf{y}|\mathbf{x}, \sigma^2) = (2\pi\sigma^2)^{-m/2} \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2\right) \tag{3.27}$$

The constraint that the signal is sparse is formalized by placing a sparseness promoting prior on $\mathbf{x}$ and then the CS problem can be solved via a manimum a posteriori (MAP) estimate of $\mathbf{x}$ using the sparseness prior. An estimate of the full posterior on $\mathbf{x}$ and σ^2

can be realized using the relevance vector machine (RVM). By defining a zero-mean Gaussian prior with α_i inverse-variance on each element of the signal, $\mathcal{N}\left(\mathbf{x}[i]\middle|0,\alpha_i^{-1}\right)$, and then marginalizing over a Gamma prior $\Gamma\left(\alpha_i\middle|a,b\right)$, the overall prior can be written as in (3.28).

$$p\left(\mathbf{x}\middle|a,b\right)=\prod_{i=1}^{n}\int_0^\infty \mathcal{N}\left(\mathbf{x}[i]\middle|0,\alpha_i^{-1}\right)\Gamma\left(\alpha_i\middle|a,b\right)\mathrm{d}\alpha_i \tag{3.28}$$

The integral (3.28) corresponds to the Student-t distribution and can be evaluated analytically [58]. The Student-t distribution favors most $\mathbf{x}[i]$ being zero and so promotes sparseness when the parameters a and b are chosen appropriately. This structural form allows the implementation of iterative algorithms for the evaluation of the posterior density function of $\mathbf{x}$ and α_0. The RVM applies the type-II maximum-likelihood (ML) procedure to estimate the parameters $\boldsymbol{\alpha}$ and α_0. The posterior for $\mathbf{x}$ can be expressed as a multivariate Gaussian distribution with mean vector $\mathbf{m}=\alpha_0\mathbf{K}\mathbf{A}^T\mathbf{y}$ and covariance matrix $\mathbf{K}=\left(\alpha_0\mathbf{A}^T\mathbf{A}+\mathbf{E}\right)^{-1}$. The matrix $\mathbf{E}$ is diagonal with $\mathbf{E}=\mathrm{diag}(\alpha_1,...,\alpha_n)$. The logarithm of the marginal likelihood for parameters $\boldsymbol{\alpha}$ and α_0 is given in (3.29).

$$\log p\left(\mathbf{y}|\boldsymbol{\alpha},\alpha_0\right)=-\frac{1}{2}\left[m\log 2\pi+\log\left|\sigma^2\mathbf{I}+\mathbf{A}\mathbf{E}^{-1}\mathbf{A}^T\right|+\mathbf{y}^T\left(\sigma^2\mathbf{I}+\mathbf{A}\mathbf{E}^{-1}\mathbf{A}^T\right)^{-1}\mathbf{y}\right] \tag{3.29}$$

The type-II ML approximation uses the point estimates of $\{\alpha_i\}$

$$\hat{\alpha}_i=\frac{1-\alpha_i\mathbf{K}[i,i]}{(\mathbf{m}[i])^2} \tag{3.30}$$

for $i=1,...,n$ in order to maximize (3.29). Since the inverse variance of the noise is $\sigma^2=1/\alpha_0$, α_0 can be estimated as in (3.31).

$$\hat{\alpha}_0=\frac{m-\sum_{i=1}^{n}1-\alpha_i\mathbf{K}[i,i]}{\|\mathbf{y}-\mathbf{A}\mathbf{m}\|_2^2} \tag{3.31}$$

BCS algorithm uses the point estimates of $\{\hat{\alpha}_i\}$ (3.30) and $\hat{\alpha}_0$ (3.31) to compute the new mean vector, $\mathbf{m}=\alpha_0\mathbf{K}\mathbf{A}^T\mathbf{y}$, and the new covariance matrix, $\mathbf{K}=\left(\alpha_0\mathbf{A}^T\mathbf{A}+\mathbf{E}\right)^{-1}$, of

the posterior for the signal. Then the newly computed mean and covariance is put into (3.30) and (3.31) to update the point estimates of $\{\hat{\alpha}_i\}$ and $\hat{\alpha}_0$. The BCS algorithm iterates between computation of the mean and covarince and the point estimates (3.30) and (3.31) [58].

When CS measurements are reapeated for similar scenes or for the same type of diagnostic task, e.g. repeated MRI images, many of the measurements $\{\mathbf{y}_i\}$ are statistically related. The Bayesian model can be applied to this multi-task CS problem to exploit statistical inter-relationships between the sensing tasks [59]. For this case the likelihood function of the received signals can be written as in (3.32).

$$p\left(\mathbf{y}_i\middle|\mathbf{x}_i,\alpha_0\right) = (2\pi/\alpha_0)^{-m_i/2}\exp\left(-\frac{\alpha_0}{2}\left\|\mathbf{y}_i-\mathbf{A}_i\mathbf{x}_i\right\|_2^2\right) \tag{3.32}$$

The signal coefficients $\mathbf{x}_i[l]$ of each task is assumed to be drawn from a product of zero mean Gaussian distributions

$$p\left(\mathbf{x}_i\middle|\boldsymbol{\alpha}\right) = \prod_{l=1}^{n}\mathcal{N}\left(\mathbf{x}_i[l]\middle|0,\alpha_0^{-1}\alpha_l^{-1}\right) \tag{3.33}$$

which is shared by M number of tasks. Gamma prior is placed on the noise precision parameter α_0 (3.34).

$$p(\alpha_0|a,b) = \mathrm{Ga}(\alpha_0|a,b) = \frac{b^a}{\Gamma(a)}\alpha_0^{(a-1)}e^{-b\alpha_0} \tag{3.34}$$

The likelihood function of the signal, $\mathbf{x}_i$, given the measurements $\mathbf{y}_i$

$$\begin{aligned} p(\mathbf{x}_i|\mathbf{y}_i,\boldsymbol{\alpha}) &= \int p(\mathbf{x}_i|\mathbf{y}_i,\boldsymbol{\alpha},\alpha_0)p(\alpha_0|a,b) \\ &= \frac{\Gamma(a+n/2)\left[1+\frac{1}{2b}(\mathbf{x}_i-\mathbf{m}_i)^T\mathbf{K}_i^{-1}(\mathbf{x}_i-\mathbf{m}_i)\right]^{-(a+n/2)}}{\Gamma(a)(2\pi b)^{n/2}|\mathbf{K}_i|^{1/2}} \end{aligned} \tag{3.35}$$

where the mean and the covariance for each task can be computed as in (3.36) and

(3.37) respectively.

$$\mathbf{m}_i = \mathbf{K}_i \mathbf{A}_i^T \mathbf{y}_i \tag{3.36}$$

$$\mathbf{K}_i = (\mathbf{A}_i^T \mathbf{A}_i + \mathbf{E})^{-1} \tag{3.37}$$

This representation allows us to integrate α_0 out and to change the multivariate Gaussian distribution to the multivariate Student-t distribution. The parameters $\boldsymbol{\alpha}$ can be estimated by maximizing the logarithm of the marginal likelihood function

$$\begin{aligned}\mathcal{L}(\boldsymbol{\alpha}) &= \sum_{i=1}^{M} \log p(\mathbf{y}_i|\boldsymbol{\alpha}) = \sum_{i=1}^{M} \log \int p(\mathbf{y}_i|\mathbf{x}_i, \alpha_0) p(\mathbf{x}_i|\boldsymbol{\alpha}, \alpha_0) p(\alpha_0|a, b) \mathrm{d}\mathbf{x}_i \mathrm{d}\alpha_0 \\ &= -\frac{1}{2} \sum_{i=1}^{M} \left[(m_i + 2a) \log(\mathbf{y}_i^T \mathbf{B}_i^{-1} \mathbf{y}_i + 2b) + \log |\mathbf{B}_i| \right] + \text{constant} \end{aligned} \tag{3.38}$$

where $\mathbf{B}_i = \mathbf{I} + \mathbf{A}_i \mathbf{E}^{-1} \mathbf{A}_i^T$. A point estimate of $\boldsymbol{\alpha}$ can be obtained by applying either direct differentiation or the EM algorithm

$$\alpha_l = \frac{M}{\sum_{i=1}^{M} \mathbf{m}_i^2[l](n_i + 2a)/(\mathbf{y}_i^T \mathbf{B}_i^{-1} \mathbf{y}_i + 2b) + \mathbf{K}_i[l,l]} \tag{3.39}$$

for $l = 1, \ldots, n$. An iterative algorithm can be built using the equations (3.39) and (3.36)-(3.37) until convergence is achieved. The total computational complexity of the MT-BCS method can be reduced to $O(m^2 n)$ with a fast implementation for the inverse operation (3.37) [59].

All sparse recovery methods have their tradeoffs. The greedy pursuits have lower computational complexity in general. When compared to other CS methods, they can be considered intermediate regarding the number of measurement samples needed to obtain a sparse solution. Some of the greedy pursuit methods may require an initial guess of the sparsity level of the unknown signal. The running time of the convex relaxation methods, on the other hand, is greater than those of the greedy pursuit methods. They don't require apriori knowledge about the sparsity level of the signal and they work with lower undersampling ratios compared to greedy pursuits.

4. SPARSE SIGNAL RECOVERY BOUNDS

The performance of the sparse recovery methods reviewed in Chapter 3 depends on a number of parameters such as the number of measurement samples, the choice and the dimensions of the sensing matrix and the perturbations to both the measurement vector and the sensing matrix. In order to accurately estimate the signal of interest by exploiting sparsity, certain conditions relating to these parameters must be met. The sparse signal recovery bounds are useful in showing limitations of CS and designing the sensing matrices for CS applications.

The variation in the sensing matrix can be modelled using a set-based, worst-case approach. The uncertainty of the sensing matrix can be described by a set of possible values $\widetilde{\mathbf{A}} \in \mathcal{A}$ which is assumed to be nonempty and bounded. The associated worst-case error of a candidate approximate solution $\mathbf{x}$ may be given as

$$e_{\mathrm{wc}}(\mathbf{x}, \widetilde{\mathbf{A}}) = \sup \left\{ \|\widetilde{\mathbf{A}}\mathbf{x} - \mathbf{y}\| \mid \widetilde{\mathbf{A}} \in \mathcal{A} \right\} \tag{4.1}$$

which is always a convex function. The objective is to minimize this worst-case error as in (4.2).

$$\text{minimize} \quad e_{\mathrm{wc}}(\mathbf{x}, \widetilde{\mathbf{A}}) = \sup \left\{ \|\widetilde{\mathbf{A}}\mathbf{x} - \mathbf{y}\| \mid \widetilde{\mathbf{A}} \in \mathcal{A} \right\} \tag{4.2}$$

This is a worst case robust approximation problem where the variable is $\mathbf{x}$ and the problem data are $\mathbf{y}$ and the set $\mathcal{A}$. The robust approximation problem is always a convex optimization problem, but its tractability depends on the norm used and the description of the uncertainty set $\mathcal{A}$. The set $\mathcal{A}$ gives our uncertainty in the linear relation between $\mathbf{x}$ and $\mathbf{y}$. The robust approximation problem can be interpreted as the robust design problem: find design variables $\mathbf{x}$ that minimize the worst possible mismatch between $\widetilde{\mathbf{A}}\mathbf{x}$ and $\mathbf{y}$ over all possible values of $\widetilde{\mathbf{A}}$. For sparse signal recovery

we should consider the worst-case error (4.3) over all $\mathbf{x} \in \Sigma_s$.

$$e_{\text{wc}}(\mathbf{x}, \widetilde{\mathbf{A}}) = \inf_{\hat{\mathbf{x}}} \sup_{\mathbf{x} \in \Sigma_s} \mathbb{E} \left\{ \frac{1}{n} ||\hat{\mathbf{x}} - \mathbf{x}||_2^2 \right\} \tag{4.3}$$

4.1. Perturbation Analysis of Additive Noise

The noise term that is added to the ideal measurement vector in the signal model is called additive noise. For communication systems additive noise may arise from electronic components and amplifiers at the receiver or from inteference encountered in radio signal transmission. This type of noise can also be characterized as thermal noise. A best case and a worst case analysis of the signal reconstruction error under additive noise is presented.

4.1.1. Best Case Analysis

An oracle estimator is constructed using the knowledge of the support of $S(\mathbf{x})$. Since it depends on an unknown information that is the support of the signal in the actual problem, the oracle estimator is infeasible. However, the oracle gives a reference performance quality to compare against and simple closed form expressions can be obtained for the mean squared error [20]. Since that the support, $S(\mathbf{x})$, is known in advance, the values of the s nonzero entries is related to the solution of the least squares problem

$$\hat{\mathbf{x}}_S^\star = \arg\min_{\hat{\mathbf{x}}_S} \|\mathbf{A}_S \hat{\mathbf{x}}_S - \tilde{\mathbf{y}}\|_2 \tag{4.4}$$

where $\hat{\mathbf{x}}_S^\star$ is the restriction of $\hat{\mathbf{x}}^\star$ to the set $S(\mathbf{x})$. An ideal estimator $\hat{\mathbf{x}}^\star$ can be constructed by using the LS projection (4.5).

$$\hat{\mathbf{x}}_S^\star = (\mathbf{A}_S^* \mathbf{A}_S)^{-1} \mathbf{A}_S^* \tilde{\mathbf{y}} \tag{4.5}$$

The ideal estimator $\hat{\mathbf{x}}^\star$ is constructed by setting its entries indexed by the complement of $S(\mathbf{x})$ to zero. Assuming that the measurement noise is from zero mean Gaussian distribution, $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \sigma_z^2 \mathbf{I})$, the expected mean squared error (MSE) can be deduced from the following equations:

$$\begin{aligned}
\mathbb{E}\left\{\|\hat{\mathbf{x}}^\star - \mathbf{x}\|_2^2\right\} &= \mathbb{E}\left\{\|(\mathbf{A}_S^*\mathbf{A}_S)^{-1}\mathbf{A}_S^*\mathbf{z}\|_2^2\right\} \\
&= \mathbb{E}\left\{\left\langle(\mathbf{A}_S^*\mathbf{A}_S)^{-1}\mathbf{A}_S^*\mathbf{z}, (\mathbf{A}_S^*\mathbf{A}_S)^{-1}\mathbf{A}_S^*\mathbf{z}\right\rangle\right\} \\
&= \mathrm{Tr}\left(\mathbb{E}\left\{\left[(\mathbf{A}_S^*\mathbf{A}_S)^{-1}\mathbf{A}_S^*\mathbf{z}\mathbf{z}^*\mathbf{A}_S(\mathbf{A}_S^*\mathbf{A}_S)^{-1}\right]\right\}\right) \\
&= \mathrm{Tr}\left((\mathbf{A}_S^*\mathbf{A}_S)^{-1}\mathbf{A}_S^*\,\mathbb{E}\left\{\mathbf{z}\mathbf{z}^*\right\}\mathbf{A}_S(\mathbf{A}_S^*\mathbf{A}_S)^{-1}\right) \\
&= \sigma_z^2\,\mathrm{Tr}\left((\mathbf{A}_S^*\mathbf{A}_S)^{-1}\right) \qquad (4.6)
\end{aligned}$$

If the sensing matrix $\mathbf{A}$ obeys the RIP (2.8) for sparsity level s, then all the eigenvalues of the Grammian matrix $\mathbf{A}_S^*\mathbf{A}_S$ must belong to the interval $[1-\delta_s, 1+\delta_s]$. In this case the ideal expected MSE can be bounded as in (4.7).

$$\mathbb{E}\left\{\|\hat{\mathbf{x}}^\star - \mathbf{x}\|_2^2\right\} \geq \frac{s\sigma_z^2}{1+\delta_s} \qquad (4.7)$$

Suppose for simplicity that $\mathbf{A}$ is the identity matrix so that the measurements under Gaussian noise is also Gaussian distributed with $\tilde{\mathbf{y}} \sim \mathcal{N}(\mathbf{x}, \sigma_z^2\mathbf{I})$. In this case the oracle estimator yields the coordinates, locations of $\mathbf{x}$ which are significant, i.e. the set of indices where $|\mathbf{x}[i]| > \sigma$. Equipped with this knowledge, an ideal estimation can be done by setting $\hat{\mathbf{x}}^\star[i] = \tilde{\mathbf{y}}[i]$ for each index in the significant set, and $\hat{\mathbf{x}}^\star[i] = 0$ otherwise. The expected MSE of this ideal estimator [60] is given in (4.8).

$$\mathbb{E}\left\{\|\hat{\mathbf{x}}^\star - \mathbf{x}\|_2^2\right\} = \sum_{i=1}^{n} \min\left((\mathbf{x}[i])^2, \sigma_z^2\right) \qquad (4.8)$$

The ideal LS estimator, $\hat{\mathbf{x}}^\star$, can be found by minimizing the expected MSE

$$\hat{\mathbf{x}}^\star = \underset{\Lambda \subset \{1,\ldots,n\}}{\arg\min}\, \mathbb{E}\left\{\|\hat{\mathbf{x}}_\Lambda - \mathbf{x}\|_2^2\right\} \qquad (4.9)$$

where Λ is a subset of indices and $\tilde{\mathbf{y}}$ is regressed onto this subset, $\hat{\mathbf{x}}_\Lambda = (\mathbf{A}_\Lambda^* \mathbf{A}_\Lambda)^{-1} \mathbf{A}_\Lambda^* \tilde{\mathbf{y}}$. It can be asked whether any real estimator would obey

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_2^2 = \mathrm{O}(\log n)\, \mathbb{E}\left\{\|\mathbf{x}^\star - \mathbf{x}\|_2^2\right\} \tag{4.10}$$

with large probability. Equation (4.8) can be intepreted as a proxy for the ideal risk, $\mathbb{E}\{\|\mathbf{x}^\star - \mathbf{x}\|_2^2\}$. The error of this estimator can be written as in (4.11).

$$\|\hat{\mathbf{x}}_\Lambda - \mathbf{x}\|_2^2 = \|\hat{\mathbf{x}}_\Lambda - \mathbf{x}_\Lambda\|_2^2 + \|\mathbf{x}_\Lambda - \mathbf{x}\|_2^2 \tag{4.11}$$

The expected MSE of the first term, $\hat{\mathbf{x}}_\Lambda - \mathbf{x}_\Lambda = (\mathbf{A}_\Lambda^* \mathbf{A}_\Lambda)^{-1} \mathbf{A}_\Lambda^* \mathbf{A} \mathbf{x}_{\Lambda^c} + (\mathbf{A}_\Lambda^* \mathbf{A}_\Lambda)^{-1} \mathbf{A}_\Lambda^* \mathbf{z}$, is given in (4.12).

$$\mathbb{E}\left\{\|\hat{\mathbf{x}}_\Lambda - \mathbf{x}_\Lambda\|_2^2\right\} = \|(\mathbf{A}_\Lambda^* \mathbf{A}_\Lambda)^{-1} \mathbf{A}_\Lambda^* \mathbf{A} \mathbf{x}_{\Lambda^c}\|_2^2 + \sigma_z^2 \mathrm{Tr}((\mathbf{A}_\Lambda^* \mathbf{A}_\Lambda)^{-1}) \tag{4.12}$$

Using (4.7) as $\mathbb{E}\{\|\hat{\mathbf{x}}_\Lambda - \mathbf{x}_\Lambda\|_2^2\} \geq [\sigma_z^2 \,\mathrm{card}(\Lambda)]/[1 + \delta_{\mathrm{card}(\Lambda)}]$, this term can be written as

$$\mathbb{E}\left\{\|\hat{\mathbf{x}}_\Lambda - \mathbf{x}\|_2^2\right\} \geq \frac{1}{2}\left[\sum_{i \in \Lambda^c} \mathbf{x}_i^2 + \sigma_z^2 \mathrm{card}(\Lambda)\right] \tag{4.13}$$

for all sets Λ with $\mathrm{card}(\Lambda) \leq s$, and $\delta_s < 1$. Then the proxy (4.10) can be used to bound the ideal MSE (4.14).

$$\mathbb{E}\left\{\|\mathbf{x}^\star - \mathbf{x}\|_2^2\right\} \geq \frac{1}{2} \min_\Lambda \left(\sum_{i \in \Lambda^c} (\mathbf{x}[i])^2 + \sigma_z^2 \mathrm{card}(\Lambda)\right) = \frac{1}{2} \sum_i \min\left((\mathbf{x}[i])^2, \sigma_z^2\right) \tag{4.14}$$

4.1.2. Worst Case Analysis

Worst case analysis provides the worst performance that can be attained by any CS recovery algorithm for all possible signals of interest with a given sparsity level and a sensing matrix under additive noise. In worst case analysis there is no prior knowledge about the signal of interest.

Theorem 4.1. Suppose that we observe $\tilde{\mathbf{y}} = \mathbf{A}\mathbf{x} + \mathbf{z}$ where $\mathbf{A}$ is an $m \times n$ matrix with $m \geq s$, and $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \sigma_z^2 \mathbf{I})$. Then there exists a constant $C > 0$ such that for all $\mathbf{A} \in \mathcal{A}$, (4.15) holds [61].

$$e_{\text{wc}}(\mathbf{x}, \mathbf{A}) \geq C \frac{s\sigma_z^2}{\|\mathbf{A}\|_F^2} \log\left(\frac{n}{s}\right) \tag{4.15}$$

We also have that (4.16) is satisfied for all $\mathbf{A} \in \mathcal{A}$ [61].

$$e_{\text{wc}}(\mathbf{x}, \mathbf{A}) \geq \frac{s\sigma_z^2}{\|\mathbf{A}\|_F^2} \tag{4.16}$$

This theorem states a fundamental limit on the minimax risk which holds for any matrix $\mathbf{A} \in \mathcal{A} \subseteq \mathbb{R}^{m \times n}$ and any possible recovery algorithm. The constant C can asymptotically be taken as 1/128 [61]. This theorem says that there is no $\mathbf{A}$ and no recovery algorithm that does fundamentally better.

Theorem 4.2. If there exist a RIP constant $\delta_{2s} < \sqrt{2} - 1$ for $\mathbf{A}$, for partial perturbations $\|\mathbf{z}\|_2 \leq \varepsilon$ the solution of the BP problem (3.11) obeys

$$\|\mathbf{x}^\star - \mathbf{x}\|_2 \leq \frac{C_0}{\sqrt{s}} \|\mathbf{x} - \mathbf{x}^{(s)}\|_1 + C_1 \varepsilon \tag{4.17}$$

where $\mathbf{x}_s$ contains the s largest entries of $\mathbf{x}$ with rest set to zero.

4.2. Perturbation Anaysis of Multiplicative Noise

The variation or pertubation in the sensing matrix is called multiplicative error since it multiplies the variable $\mathbf{x}$. Multiplicative noise arise in many applications of compressed sensing including discrete radar/sonar imaging [62–64], sensor array processing [65], and channel estimation [66–68]. In all of these applications imaging targets are taken to be on a regular grid in delay, Doppler, and wavenumber. However no matter how large the size of the grid is, the sources of the actual field cannot be placed on the center of the grid points in frequency or wavenumber. The grid misalignment

problem results in a mismatch between the actual basis and the assumed sparsity basis and yields a perturbed sensing matrix that differs from the actual but unknown sensing matrix.

In the completely perturbed setup where $\mathbf{z} \neq 0$ and $\mathbf{Z} \neq 0$, the BP problem is given in (4.18).

$$\mathbf{x}^{\star} = \arg\min_{\hat{\mathbf{x}}} \|\hat{\mathbf{x}}\|_1 \qquad \text{subject to} \quad \|\widetilde{\mathbf{A}}\hat{\mathbf{x}} - \tilde{\mathbf{y}}\| \leq \varepsilon_{\text{tot}} \tag{4.18}$$

The perturbed sensing matrix can be written in terms of the sensing matrix (4.19).

$$\widetilde{\mathbf{A}} = \mathbf{A} + \mathbf{Z} \tag{4.19}$$

When (4.19) is used, the received signal can be written as

$$\tilde{\mathbf{y}} = \widetilde{\mathbf{A}}\mathbf{x} + \mathbf{z} = (\mathbf{A} + \mathbf{Z})\mathbf{x} + \mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{Z}\mathbf{x} + \mathbf{z} \tag{4.20}$$

the extra noise term, $\mathbf{Zx}$, can be intepreted as multiplicative noise. The perturbations of the completely perturbed BP problem (4.18) can be quantified with the following relative bounds

$$\frac{\|\mathbf{z}\|_2}{\|\mathbf{y}\|_2} \leq \varepsilon_{\mathbf{y}}, \ \frac{\|\mathbf{Z}\|_2}{\|\mathbf{A}\|_2} \leq \varepsilon_{\mathbf{A}}, \ \frac{\|\mathbf{Z}\|_2^{(l)}}{\|\mathbf{A}\|_2^{(l)}} \leq \varepsilon_{\mathbf{A}}^{(l)}, \ \varepsilon = \max_{l=1,\ldots,s} \varepsilon_{\mathbf{A}}^{(l)} \tag{4.21}$$

where $\|\cdot\|_2$ denotes the spectral norm for a matrix and $\|\cdot\|_2^{(l)}$ denote the largest spectral norm taken over all l-column submatrices. With the help of relative perturbations (4.21) the total perturbation term, $\mathbf{Zx} + \mathbf{z}$, can be bounded using the following

inequalities:

$$\begin{aligned}
\|\mathbf{Zx}\|_2 + \|\mathbf{z}\|_2 = \frac{\|\mathbf{Zx}\|_2 + \|\mathbf{z}\|_2}{\|\mathbf{y}\|_2}\,\|\mathbf{y}\|_2 &= \left(\frac{\|\mathbf{Zx}\|_2}{\|\mathbf{Ax}\|_2} + \frac{\|\mathbf{z}\|_2}{\|\mathbf{y}\|_2}\right)\,\|\mathbf{y}\|_2 \\
&\leq \left(\frac{\|\mathbf{Z}\|_2^{(s)}}{\sqrt{1-\delta_s}}\,\frac{\|\mathbf{x}\|_2}{\|\mathbf{x}\|_2} + \varepsilon_{\mathbf{y}}\right) + \|\mathbf{y}\|_2 \\
&\leq \left(\frac{\|\mathbf{Z}\|_2^{(s)}}{\sqrt{1+\delta_s}}\,\frac{\sqrt{1+\delta_s}}{\sqrt{1-\delta_s}} + \varepsilon_{\mathbf{y}}\right) + \|\mathbf{y}\|_2 \\
&\leq \left(\frac{\|\mathbf{Z}\|_2^{(s)}}{\|\mathbf{A}\|_2^{(s)}}\,\kappa_{\mathbf{A}}^{(s)} + \varepsilon_{\mathbf{y}}\right) + \|\mathbf{y}\|_2 \\
&\leq \left(\varepsilon_{\mathbf{A}}^{(s)}\kappa_{\mathbf{A}}^{(s)} + \varepsilon_{\mathbf{y}}\right) + \|\mathbf{y}\|_2 = \varepsilon_{\text{tot}}
\end{aligned} \tag{4.22}$$

where $\mathbf{A}$ is assumed to satisfy RIP (2.8), $\|\mathbf{A}\|_2^{(s)} \leq \sqrt{1+\delta_s}$ [32]. $\kappa_{\mathbf{A}}^{(s)}$ is defined as the ratio of $\kappa_{\mathbf{A}}^{(s)} = \sqrt{(1+\delta_s)/(1-\delta_s)}$. For sensing matrices with very small δ_s which means that every s column submatrice is approximately orthonormal, then $\kappa_{\mathbf{A}}^{(s)} \approx 1$.

Mean relative recovery error using the OMP algorithm is given for varying relative perturbation, $\varepsilon_{\mathbf{A}} = 0, 0.01, 0.05, 0.1$ in Figure 4.1. While the size of the signal vector is $n = 512$, the dimension of the measurement vector is $n = 128$ resulting in an undersampling ratio of 0.5. In order to focus on the effect of multiplicative noise, the measurements are not perturbed by additive noise. The sensing matrix is chosen from the Gaussian ensemble and the relative perturbation is computed as $\varepsilon_{\mathbf{A}} = \|\mathbf{Z}\|_2/\|\mathbf{A}\|_2$. The perturbed BP problem is solved by using the interior point method (Section 3.2.1) implementation of the CVX package [54, 55] in MATLAB. MOSEK is chosen as the solver of the CVX package. When there is no multiplicative noise $\varepsilon_{\mathbf{A}} = 0$, it is seen that the sparse recovery algorithms can yield solutions of this undersampled system of equations with relative error less than 0.1 for sparse signals with support size less than 40. As the relative bound of the multiplicative noise increases, it is observed that the compressibility of the signal decreases.

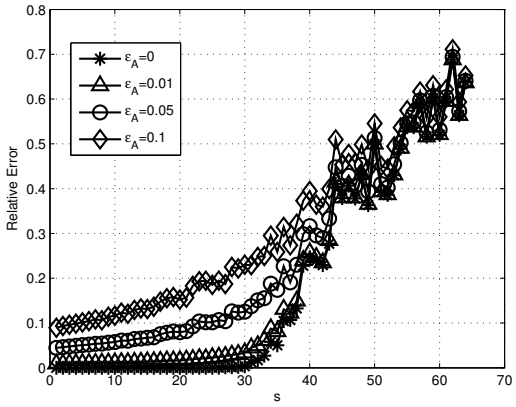


Figure 4.1. Mean relative error for varying relative perturbations.

4.2.1. Best Case Analysis

A best of the worst-case analysis can also be applied to the completely perturbed case where the sensing matrix is perturbed as $\widetilde{\mathbf{A}} = \mathbf{A} + \mathbf{Z}$. In this case the oracle estimator gives the LS solution (4.23).

$$\mathbf{x}_S^{\#} = \arg\min \|\widetilde{\mathbf{A}}_S \mathbf{x}_S - \tilde{\mathbf{y}}\|_2 \tag{4.23}$$

Golub and Van Loan's model in Theorem 5.3.1 [69] formulates the perturbed LS problem by subtituting $\widetilde{\mathbf{A}}_S = \mathbf{A}_S + (t/\varepsilon)\mathbf{Z}_S$ and $\tilde{\mathbf{y}} = \mathbf{y} + (t/\varepsilon)\mathbf{z}$ as

$$\left(\mathbf{A}_S + \frac{t}{\varepsilon}\mathbf{Z}_S\right)^* \left(\mathbf{A}_S + \frac{t}{\varepsilon}\mathbf{Z}_S\right) \mathbf{x}_S(t) = \left(\mathbf{A}_S + \frac{t}{\varepsilon}\mathbf{Z}_S\right)^* \left(\mathbf{y} + \frac{t}{\varepsilon}\mathbf{z}\right) \tag{4.24}$$

where the solution is assumed to be continuously differentiable for all $t \in [0, \varepsilon]$. With the solution given as $\mathbf{x}_S^{\#} = \mathbf{x}_S + \varepsilon \dot{\mathbf{x}}_S(0) + O(\varepsilon^2)$ [69], the error can be bounded as in (4.25).

$$\|\mathbf{x}_S^{\#} - \mathbf{x}_S\|_2 = \varepsilon \|\dot{\mathbf{x}}_S(0)\|_2 + O(\varepsilon^2) \tag{4.25}$$

Differentiating Equation (4.24) and setting $t = 0$ yields (4.26).

$$\dot{\mathbf{x}}_S(0) = \frac{1}{\varepsilon}(\mathbf{A}_S^*\mathbf{A}_S)^{-1}\mathbf{A}_S^*(\mathbf{z} - \mathbf{Z}_S\mathbf{x}_S) \tag{4.26}$$

The solution is bounded above by inserting (4.26) into (4.25)

$$\|\mathbf{x}^{\#} - \mathbf{x}\|_2 \leq \|\mathbf{A}_S^{\dagger}\|_2 \left(\frac{\|\mathbf{Z}_S\mathbf{x}_S\|_2}{\|\mathbf{A}\mathbf{x}\|_2} + \varepsilon_{\mathbf{y}}\right) \|\mathbf{y}\|_2 \tag{4.27}$$

where the complete solution, $\mathbf{x}^{\#}$, is obtained by extending $\mathbf{x}_S^{\#}$ with zero-padding on the complement of the support set S and $\mathbf{A}_S^{\dagger} = (\mathbf{A}_S^*\mathbf{A}_S)^{-1}\mathbf{A}_S^*$ is the pseudoinverse of $\mathbf{A}_S$ [32].

4.2.2. Worst Case Analysis

The objective of the worst case analysis under multiplicative noise is to find $\mathbf{x}$ that minimize the worst possible mismatch between $\widetilde{\mathbf{A}}\mathbf{x}$ and $\mathbf{y}$ over all possible values of $\widetilde{\mathbf{A}}$. Unlike the best case analysis there is no prior knowledge about the support of $\mathbf{x}$.

Theorem 4.3. Let $\mathbf{x}^{(s)}$ be the best s-term approximation to the signal $\mathbf{x}$, the tail of the signal be $\mathbf{x}^{(t)} = \mathbf{x} - \mathbf{x}^{(s)}$ and

$$\varepsilon_h = \frac{1.23}{(1 - \varepsilon_{\mathbf{A}})} \left[\varepsilon_{\mathbf{A}} + \varepsilon_{\mathbf{y}} + (1 + \varepsilon_{\mathbf{y}})(\beta + \gamma)\right] \|\mathbf{x}^{(s)}\|_2 \tag{4.28}$$

where $\beta = \|\mathbf{x}^{(t)}\|_2 / \|\mathbf{x}^{(s)}\|_2$ and $\gamma = \|\mathbf{x}^{(t)}\|_1 / \sqrt{s}\|\mathbf{x}^{(s)}\|_2$. If $\mathbf{A}$ satisfies RIP (2.8) of order $s + 1$ with RIC less than

$$\tilde{\delta}_{s+1} < \frac{1}{(\sqrt{s} + 1)} - \frac{3}{(\sqrt{s} + 1)} \frac{\varepsilon_h}{t_0} \tag{4.29}$$

where $t_0 = \min_{j \in S(\mathbf{x}^{(s)})} |\mathbf{x}_j|$, then for any signal OMP will recover the support set $S(\mathbf{x})$

exactly from $\tilde{\mathbf{y}}$ and $\widetilde{\mathbf{A}}$ in s iterations and the error is given in (4.30) [70].

$$\|\mathbf{x}^{\star} - \mathbf{x}\|_2 \leq \frac{\varepsilon_h}{\sqrt{1-\tilde{\delta}_s}} \tag{4.30}$$

Theorem 4.4. For the fixed relative perturbations $\varepsilon_{\mathbf{A}}$, $\varepsilon_{\mathbf{A}}^{(s)}$, $\varepsilon_{\mathbf{A}}^{(2s)}$, and $\varepsilon_{\mathbf{y}}$ in (4.21) the solution of the completely perturbed problem (4.18) is bounded by the total perturbation (4.22) as

$$\|\mathbf{x}^{\star} - \mathbf{x}\|_2 \leq \frac{\widetilde{C}_0}{\sqrt{s}}\|\mathbf{x} - \mathbf{x}^{(s)}\|_1 + \widetilde{C}_1\varepsilon_{\text{tot}} \tag{4.31}$$

provided that the sensing matrix $\mathbf{A}$ satisfies the RIP property (2.8) for

$$\delta_{2s} < \frac{\sqrt{2}}{\left[1+\varepsilon_{\mathbf{A}}^{(2s)}\right]^2} - 1 \tag{4.32}$$

and the signal $\mathbf{x}$ satisfies (4.33) [32].

$$\beta + \gamma < \frac{1}{\kappa_{\mathbf{A}}^{(s)}} \tag{4.33}$$

The constants, $\widetilde{C}_1$ and $\widetilde{C}_2$, are defined as in (4.34) and (4.35) respectively.

$$\widetilde{C}_0 = \frac{2\left(1 + (\sqrt{2}-1)\left[(1+\delta_{2s})\left(1+\varepsilon_{\mathbf{A}}^{(2s)}\right)^2 - 1\right]\right)}{1 - (\sqrt{2}+1)\left[(1+\delta_{2s})\left(1+\varepsilon_{\mathbf{A}}^{(2s)}\right)^2 - 1\right]} \tag{4.34}$$

$$\widetilde{C}_1 = \frac{4\sqrt{1+\delta_{2s}}\left(1+\varepsilon_{\mathbf{A}}^{(2s)}\right)}{1 - (\sqrt{2}+1)\left[(1+\delta_{2s})\left(1+\varepsilon_{\mathbf{A}}^{(2s)}\right)^2 - 1\right]} \tag{4.35}$$

In order for $\mathbf{A}$ to satisfy Equation (4.32) the RIC must be nonnegative $0 < \delta_{2s}$ and so the relative perturbation must be less than $\varepsilon_{\mathbf{A}}^{(2s)} < \sqrt[4]{2} - 1 \approx 0.1892$. The upper bound given in (4.31) can be considered the worst instance of the worst case scenarios.

CS recovery methods are known to achieve much better performance in practice. When the signal $\mathbf{x}$ is strictly sparse, the tail of the signal is zero, $\mathbf{x}^{(t)} = \mathbf{0}$, and so the terms related to the tail disappear. Then the accuracy of the solution becomes (4.36).

$$\|\mathbf{x}^\star - \mathbf{x}\|_2 \leq \widetilde{C}_1 \, \varepsilon_{\text{tot}} \leq \widetilde{C}_1 \left(\kappa_{\mathbf{A}}^{(s)} \varepsilon_{\mathbf{A}}^{(s)} + \varepsilon_{\mathbf{y}} \right) \|\mathbf{y}\|_2 \tag{4.36}$$

The bounds covered in this chapter shows the limitations of the CS recovery methods. Two types of perturbations, additive noise and multiplicative noise, are investigated. For each type of noise, a best-case and a worst-case analysis is presented. Best-case analysis assume that the sparsity of the signal of interest is known beforehand. Although it is not feasible to implement the best-case estimator in practice, the best-case bounds can be used as a benchmark to compare the performances of different sparse recovery methods. In worst-case analysis no knowledge about the signal of interest is known. They represent worst-case performances under which the CS methods in practice perform. The worst-case analysis can be used in designing sensing matrices for the applications of the sparse signal recovery since the parameters of the worst-case bounds can be related to the physical attributes of the applied system which can better demonstrate the limitations of the application at hand.

5. ELECTROMAGNETIC IMAGING

The reconstruction of the electromagnetic properties of the unknown targets from measured scattered field is called electromagnetic inverse scattering problem and it constitutes a nonlinear and ill-posed problem. The CS framework can provide a way to overcome the issues of such ill-posed inversion problems such as non-uniqueness and numerical stability. The efficient sparsity promoting methods of the CS framework have already found many applications in electromagnetic inverse source problems. Convex relaxation methods such as the BP method (Section 3.2) has been applied in the reconstruction of objects for microwave imaging [6–8]. Another sparse recovery method that has gained much recognition in imaging is the BCS method (Section 3.3). Through-the-wall radar imaging and microwave imaging of sparse scatterers has been recast into the BCS framework [9–11]. The MT-BCS method (Section 3.3) which improves upon the BCS method by jointly considering multiple correlated measurements has also been applied to time-harmonic two-dimensional inverse scattering problems with both sparse and nonsparse scatterers [12–14].

5.1. Inverse Scattering Problem

Resolving the positions of closely spaced scatterers is an important problem of electromagnetic imaging. The CS framework may allow us to accomplish this aim but in order to recast this problem within the CS framework, the nonlinear problem must first be linearized. Thus Born approximation (BA) is considered to linearize the nonlinear inverse problem [71–73]. Let us consider a configuration where two scattering objects are closely placed in homogeneous medium. Transverse-magnetic (TM) polarized incident electric field E^{inc} is considered for illumination. N_{tx} Transmitting and N_{rx} receiving antennas are located on a closed line around the object Γ. For each source, $i = 1, 2, \ldots, N_{\text{tx}}$, the scattered field E^{sc} can be formulated as

$$E_i^{\text{sc}}(\mathbf{r}) = k_0^2 \int_D G(\mathbf{r}, \mathbf{r}')\chi(\mathbf{r}')E_i^{\text{tot}}(\mathbf{r}')\mathrm{d}\mathbf{r}' \tag{5.1}$$

where $\mathbf{r} \in \Gamma$, s scatterers, $B_1, \ldots, B_s$, are located in the investigation domain D and the total electric field composing of the incident and scattered fields is given as $E_i^{\text{tot}}(\mathbf{r}')$. The wavenumber of free space is denoted by k_0. The object function $\chi(\mathbf{r}')$ defines the characteristics of the scatterer as $\chi = \epsilon_r - 1$ where relative permittivity ϵ_r is a complex number in general. $G(\mathbf{r}, \mathbf{r}')$ is the Green function of homogeneous medium and defined as

$$G(\mathbf{r}, \mathbf{r}') = \frac{j}{4} H_0 \left(k_0 \left| \mathbf{r} - \mathbf{r}' \right| \right) \tag{5.2}$$

where H_0 is the zero order Hankel function of the first kind. Inverse scattering algorithms aims to determine the unknown object function which assumed to be located in some investigation domain from the measured scattered field using (5.1). In general, it is a difficult task to obtain a solution due to nonlinearity because of the total field term in the right hand side of (5.1). BA can be used in linearization by replacing the unknown total field inside the investigation domain with the known incident field for weak scatterers. The pulse expansion function is used to approximate the object function in discretized investigation domain which results in (5.3).

$$E_i^{\text{sc}}(\mathbf{r}) = \sum_{n=1}^{N_{\text{c}}} \widetilde{\chi}_n \, k_0^2 \int_{\text{cell } n} G(\mathbf{r}, \mathbf{r}') E_i^{\text{inc}}(\mathbf{r}') \mathrm{d}\mathbf{r}' \tag{5.3}$$

N_{c} is the total number of square cells in the investigation domain and $\widetilde{\chi}_n$ is a constant value of object function for the n-th cell. For N_{tx} sources, (5.3) can be written in CS framework as

$$\mathbf{y} = \begin{bmatrix} E_1^{\text{sc}} \\ \vdots \\ E_{N_{\text{tx}}}^{\text{sc}} \end{bmatrix} = \mathbf{A}\mathbf{x} \tag{5.4}$$

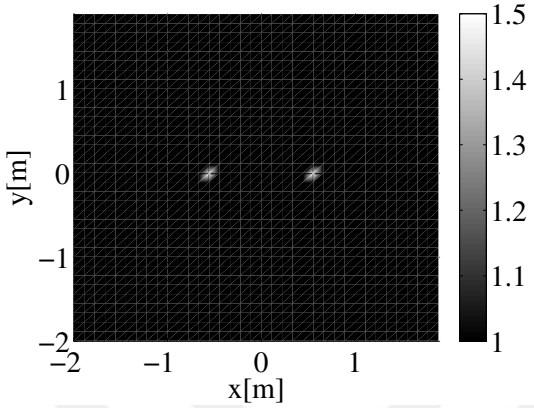
where $\mathbf{y}$ is the measurement vector of size $N_{\text{tx}} N_{\text{rx}} \times 1$, $\mathbf{A}$ is the sensing matrix of size $N_{\text{tx}} N_{\text{rx}} \times N_{\text{c}}$ and the unknown sparse vector of object functions is denoted as the vector $\mathbf{x} = [\widetilde{\chi}_1 \ldots \widetilde{\chi}_{N_{\text{c}}}]^T$ of size $N_{\text{c}} \times 1$.

5.2. Numerical Results

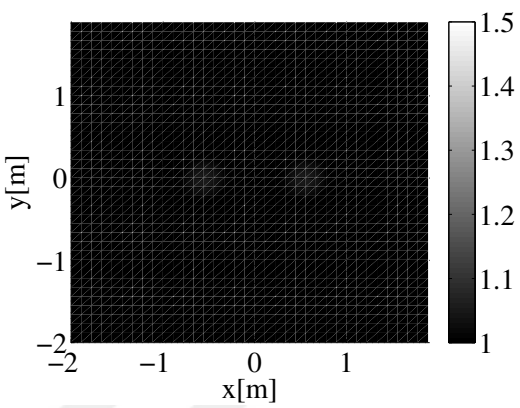
Numerical experiments are conducted to compare the performances of different CS algorithms as well as classical BA method for reconstruction of point-like scatterers. Source frequency is chosen as 300MHz, and the wavelength is ξ =1m. The scatterer objects are squares with 0.1m side length and complex permittivity $\epsilon_r = 1.5$. 30 line sources and 50 receivers are placed on a circular line of radius 3m. The size of the investigation domain is 4m $\times$ 4m and it is discretized into 37 $\times$ 37 sections. This setup is investigated for two cases: two scatterers placed in far proximity and four scatterers placed in close proximity.

5.2.1. Two Scatterers In Far Proximity

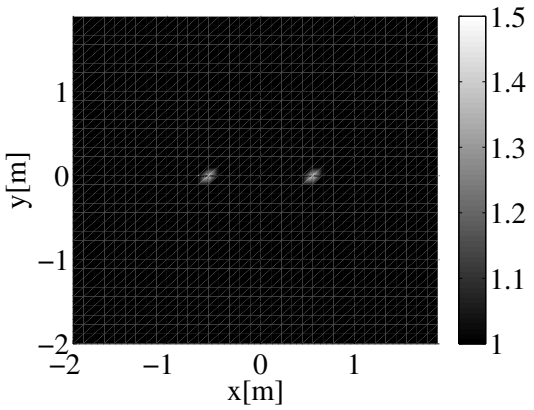
At first, the objects are separated by ξ which can be resolved by the classical BA method (Figure 5.1(b)). Sparse recovery methods of OMP, NESTA (Section 3.2.2) and MT-BCS (Section 3.3) algorithms are applied in the reconstruction of the objects. The SNR is set to 50 dB. The maximum number of iterations is assigned to 3000 for the NESTA algorithm in order to obtain a fair comparison. The reconstructed permittivity profiles of BA, OMP, NESTA and MT-BCS methods are shown in Figure 5.1. As it can be seen from Figure 5.1, all of the applied sparse recovery methods are able to resolve the sufficiently separated objects and provide better reconstruction than the classical BA method. Amongst the applied sparse recovery methods the OMP algorithm may be preferable for such high SNR and far proximity cases since the performance of the OMP algorithm matches to those of more complex methods of NESTA and MT-BCS with the least computational complexity. The total reconstruction error (Figure 5.2) of the methods are computed in order to compare the performance of the algorithms against varying levels of noise quantitatively. The SNR is set between 5dB and 50dB. As it can be observed from Figure 5.2, in low SNR region (5dB and 10dB) the NESTA method yields the best performance followed by the OMP algorithm. As SNR increases beyond 15dB, the OMP algorithm outperforms the NESTA method and becomes the best performing algorithm in high SNR region. The worst reconstruction performance belongs to the classical BA method as expected.



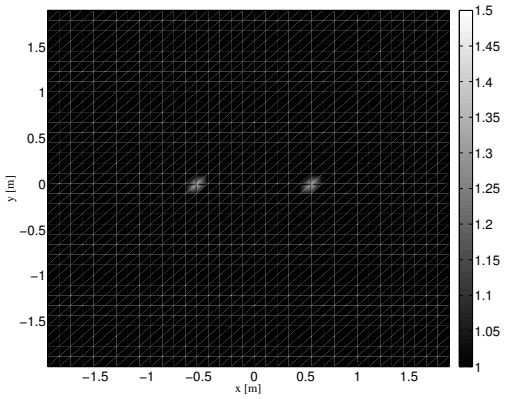
(a) Exact profile



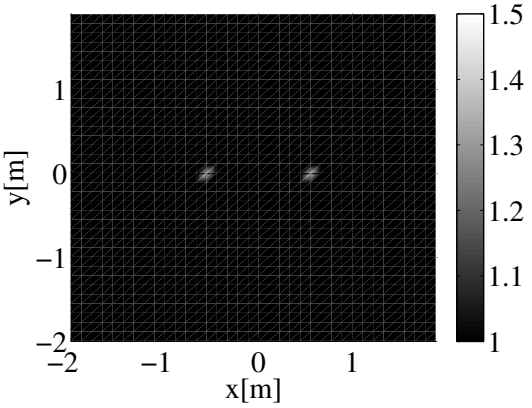
(b) BA



(c) OMP



(d) NESTA



(e) MT-BCS

Figure 5.1. Reconstructed permittivity profiles for SNR= 50dB and ξ.

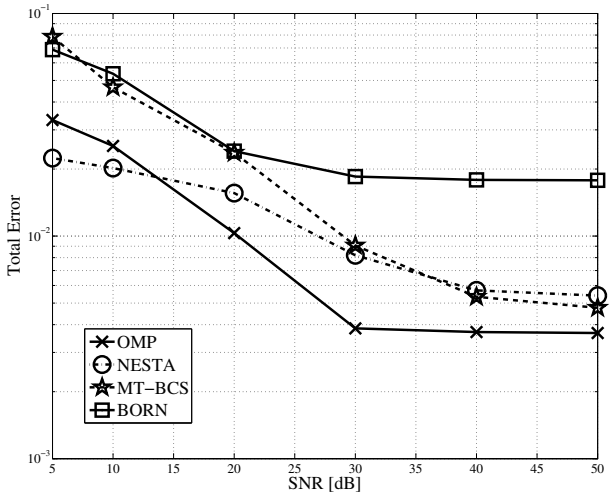
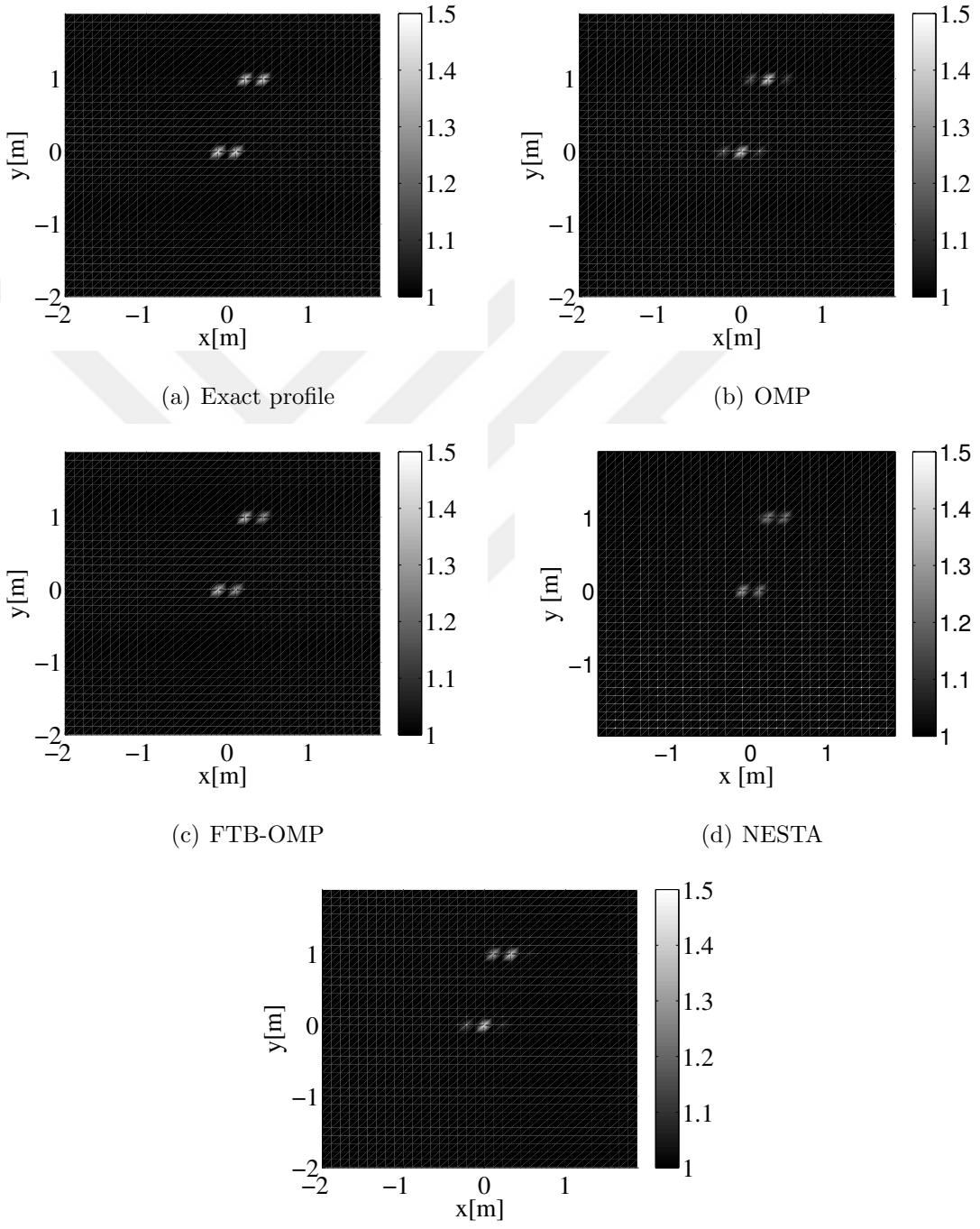


Figure 5.2. Total error for two objects separated with ξ.

5.2.2. Four Scatterers In Close Proximity

In order to make the reconstruction a nontrivial task for the second case, the number of scatterers is increased to four and they are pairwise placed in close proximity of $\xi/8$ (Figure 5.3). First less noisy measurements are considered and the reconstructed permittivity profiles for this case (SNR=50dB) are shown in Figure 5.3. The parameters of the FTB-OMP algorithm are chosen as $\iota = 2, \vartheta = 16$. It is seen that both OMP and MT-BCS algorithms cannot differentiate all of the unknown objects (Figure 5.3(b) and Figure 5.3(e)). On the other hand, the FTB-OMP and the NESTA methods are clearly able to resolve all of the four objects, as shown in Figure 5.3(c) and Figure 5.3(d) respectively. The FTB-OMP algorithm aided with the tree based search outperforms the OMP algorithm as expected. However, it should be noted that this performance improvement comes at the expense of increased computational complexity. The reconstruction results with 30 dB SNR is also presented in order to obtain further insight about the robustness of these algorithms (Figure 5.4). While the OMP, MT-BCS and FTB-OMP methods all suffer from the higher noise level, the performance of the NESTA method proves to be more robust. The highest performance degradation due to noise belongs to the MT-BCS method. The NESTA method outperforms all of the other algorithms in this high noise case.



(a) Exact profile

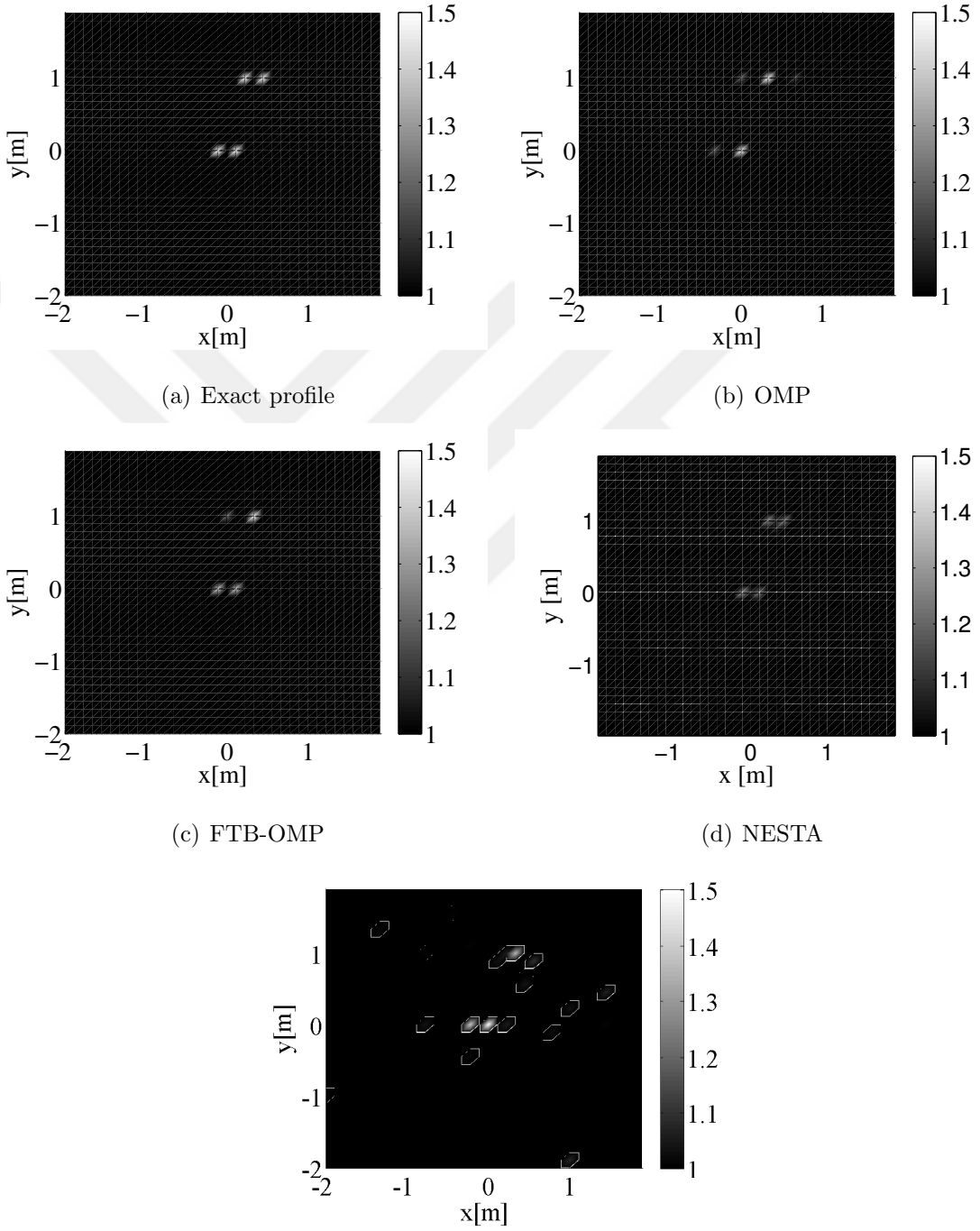
(b) OMP

(c) FTB-OMP

(d) NESTA

(e) MT-BCS

Figure 5.3. Reconstructed permittivity profiles for SNR= 50dB.



(a) Exact profile

(b) OMP

(c) FTB-OMP

(d) NESTA

(e) MT-BCS

Figure 5.4. Reconstructed permittivity profiles for SNR= 30dB.

The reconstruction error of the close proximity setup against varying SNR levels are presented in Figure 5.5. As it can be seen from Figure 5.5, the best performance for the noisy region belongs to the NESTA method. The performances of both greedy methods, OMP and FTB-OMP, are very close to each other in the noisy region but once the SNR reaches 20dB, the performance of the FTB-OMP method begin to scale better with the increasing SNR level and differentiate itself from the OMP method. The OMP method, on the other hand, appears to reach an error floor at 20dB SNR. When the SNR reaches 30dB, the performance of the FTB-OMP method catches up the NESTA method and then begin to outperform the NESTA method past 30dB SNR. The MT-BCS method yields the worst performance in the high noise region (5dB and 10dB) but then catches up the OMP method and hit an error floor at 20dB SNR just like the OMP method.

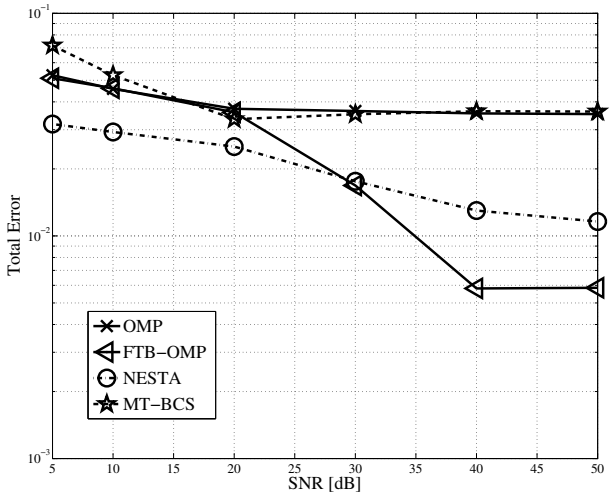


Figure 5.5. Total error for two objects separated with $\xi/8$.

6. JOINT FREQUENCY OFFSET AND CHANNEL ESTIMATION FOR OFDM

Orthogonal frequency division multiplexing (OFDM) has become a standard multi-carrier modulation technique for broadband wireless communication networks due to its resistance to interblock interference (IBI) caused by frequency selective multipath fading channels. The success of OFDM systems relies on the orthogonality property of its chosen subcarriers which transforms the frequency selective channel into a set of frequency flat fading channels simplifying the equalization task. However it is known that OFDM is very sensitive to both frequency synchronization and channel estimation errors. The oscillator mismatches and/or Doppler shifts introduce carrier frequency offset (CFO) which destroys the orthogonality amongst the subcarriers. As a result, the signal constellation is rotated and inter-carrier interference (ICI) occurs. The quality of the channel estimates, which is vital for coherent data detection, is also affected negatively under CFO. All of these effects reduce the effective SNR at the receiver and degrade the performance of the OFDM system. Thus, accurate CFO and channel estimation is essential for exploiting the full potential of OFDM systems.

CFO and channel estimation can be done separately by first estimating the CFO and then performing channel estimation in the second step. Both training symbol based and blind CFO estimators have been proposed in the literature [74–78]. The estimator proposed by Moose uses repeated data symbols [74]. Classen's method inserts pilot subcarriers in OFDM blocks for CFO estimation and assume that the channel does not change for two consecutive OFDM blocks [75]. Both methods assume a sufficiently small CFO and also a not high SNR so that the ICI is much smaller than the noise and can be ignored. They are sensitive to the value of the CFO and are valid for small CFO values. The blind estimator developed by Beek exploits the redundancy in the cyclic prefix (CP) [76]. A blind estimator based on MUSIC subspace method is developed in [77] using the virtual subcarriers (VC). The p-algorithm in [78] considers both VC and pilot carriers in each OFDM block but assumes that the channel remains

constant for two OFDM blocks like the Classen method. Channel estimation can be performed after applying these well established CFO synchronization methods. However, since perfect synchronization is not possible, the residual CFO will degrade the performance of the channel estimate significantly. Thus, better performance can be obtained when CFO and channel are estimated jointly. There exists a number of joint CFO and channel estimators that use pilot symbols [79–82]. Both [79] and [80] use the framework of the expectation-maximization (EM) algorithm. While [79] directly computes the channel parameters, [80] estimates the parameters of a basis expansion based parametric channel model assuming the channel delays are known. The initialization of [79] requires a coarse CFO estimate such as provided by the Beek's method [76]. A joint maximum likelihood estimate (MLE) of CFO and channel impulse response (CIR) using a training symbol is given in [81]. The estimator in [82] is an approximate MLE since the received signal samples of the OFDM system are assumed to be Gaussian. Also both [79] and [82] require the second-order statistics of the channel and noise. For [82], an initial estimate of both the channel and CFO is necessary to initialize the joint MLE.

6.1. FFT Implementation of OFDM with Carrier Frequency Offset

Information is transmitted in blocks of symbols in an OFDM system. An OFDM block consists of N_{s} transmitted symbols $X[k], k = 0, ..., N_{\mathrm{s}} - 1$, each chosen from a signal constellation. Time domain samples of the OFDM block is obtained after performing inverse discrete Fourier transform (IDFT) on the transmitted symbols

$$x[i] = \frac{1}{N_{\mathrm{s}}} \sum_{k=0}^{N_{\mathrm{s}}-1} X[k] e^{j2\pi k(i-N_{\mathrm{g}})/N_{\mathrm{s}}} \tag{6.1}$$

where $i = N_{\mathrm{g}}, ..., N_{\mathrm{t}} - 1$, N_{t} is the total number of samples in a OFDM block including the cyclic-prefix (CP) samples, $N_{\mathrm{t}} = N_{\mathrm{s}} + N_{\mathrm{g}}$. CP samples required to ensure that no IBI occurs and so the first N_{g} samples of the OFDM block, $\{x[0], ..., x[N_{\mathrm{g}} - 1]\}$, are taken same as the last N_{g} samples of the OFDM block $\{x[N_{\mathrm{t}} - N_{\mathrm{g}}], ..., x[N_t - 1]\}$.

The baseband signal is converted up to the passband by carrier modulation and then the receiver converts the passband signal down to baseband by using a local carrier signal with the same carrier frequency. Distortions occur when the carrier signal of the transmitter is not perfectly matched to the carrier signal of the receiver. There are two types of carrier signal distortions between the transmitter and the receiver. The instability of the carrier signal generators used at the transmitter and the receiver is called phase noise. The other type of distortion is called CFO, Δf_{c}. CFOs are typically introduced by a small frequency mismatch in the local oscillators of the transmitter and the receiver. Doppler shifts can also induce a slight frequency change of the carrier frequency which leads to CFO. CFO can be normalized as $\epsilon_{\mathrm{n}} = \Delta f_{\mathrm{c}} N_{\mathrm{s}} T$ where $1/(N_{\mathrm{s}}T)$ is the subcarrier bandwidth and T is the sampling period. Some examples of the maximum Doppler frequency and the normalized CFO for different communication systems such as DMB, 3GPP and mobile WiMAX, at the mobile speed of 120 km/h are shown in Table 6.1 [83].

Table 6.1. Doppler frequency and normalized CFO.

System	f_{c}	$1/(N_{\mathrm{s}}T)$	**Doppler frequency**	ϵ_{n}
DMB	375MHz	1kHz	41.67Hz	0.042
3GPP	2GHz	15kHz	222.22Hz	0.0148
Mobile WiMAX	2.3GHz	9.765kHz	255.55Hz	0.0263

The received signal after being transmitting $x[i]$ through a channel $h[i]$ can be written in terms of the normalized CFO as

$$y[i] = N_{\mathrm{s}}\, c_{\epsilon_{\mathrm{n}}}[i](h[i] * x[i]) + z[i] \tag{6.2}$$

where $*$ is the convolution operator, i.e. $h[i] * x[i] = \sum_{r=0}^{N_{\mathrm{g}}} h[r]x[i-r]$, and z is zero mean complex additive white Gaussian noise with variance σ_z^2, and $c_{\epsilon_{\mathrm{n}}}[i]$ is given in (6.3).

$$c_{\epsilon_{\mathrm{n}}}[i] = \frac{1}{N_{\mathrm{s}}} e^{j2\pi\epsilon_{\mathrm{n}} i/N_{\mathrm{s}}} \tag{6.3}$$

The relative frequency offset can be divided into an integer and non-integer/fractional part, $\epsilon_{\mathrm{n}} = \epsilon_{\mathrm{i}} + \epsilon_{\mathrm{f}}$. If only the integer part of the CFO is present, a CFO of ϵ_{i} causes a phase offset $2\pi i\epsilon_{\mathrm{i}}/N_{\mathrm{s}}$ in the time domain, $e^{j2\pi\epsilon_{\mathrm{i}}i/N_{\mathrm{s}}}x[i]$ ignoring the effect of the channel. This is equivalent to cyclic-shifted frequency domain signals, $X[k-\epsilon_{\mathrm{i}}]$ which causes a significant degradation in the BER performance if not compensated. However, since there is no fractional CFO, the orthogonality amongts the carriers is maintained and so no ICI occurs. Assuming that the integer part is zero, $\epsilon_{\mathrm{i}} = 0$, the DFT applied received samples can be written as:

$$\begin{aligned}
Y[k] &= \sum_{i=0}^{N_{\mathrm{s}}-1} y[i]e^{-j2\pi ki/N_s} \\
&= \sum_{i=0}^{N_{\mathrm{s}}-1} \frac{1}{N_{\mathrm{s}}} \sum_{l=0}^{N_{\mathrm{s}}-1} H[l]X[l]e^{j2\pi(l+\epsilon_{\mathrm{f}})i/N_{\mathrm{s}}}e^{-j2\pi ki/N_{\mathrm{s}}} + \sum_{i=0}^{N_{\mathrm{s}}-1} z[i]e^{-j2\pi ki/N_{\mathrm{s}}} \\
&= \frac{1}{N_{\mathrm{s}}} \sum_{l=0}^{N_{\mathrm{s}}-1} H[l]X[l] \sum_{i=0}^{N_{\mathrm{s}}-1} e^{j2\pi(l-k+\epsilon_{\mathrm{f}})i/N_{\mathrm{s}}} + Z[k] \\
&= \frac{1}{N_{\mathrm{s}}} H[k]X[k] \sum_{i=0}^{N_{\mathrm{s}}-1} e^{j2\pi\epsilon_{\mathrm{f}}i/N_{\mathrm{s}}} + \frac{1}{N_{\mathrm{s}}} \sum_{l=0,l\neq k}^{N_{\mathrm{s}}-1} H[l]X[l] \sum_{i=0}^{N_{\mathrm{s}}-1} e^{j2\pi(l-k+\epsilon_f)i/N_{\mathrm{s}}} + Z[k] \\
&= \frac{1}{N_{\mathrm{s}}} \frac{1-e^{j2\pi\epsilon_{\mathrm{f}}}}{1-e^{j2\pi\epsilon_{\mathrm{f}}/N_{\mathrm{s}}}} H[k]X[k] + \frac{1}{N_{\mathrm{s}}} \sum_{l=0,l\neq k}^{N_{\mathrm{s}}-1} H[l]X[l] \frac{1-e^{j2\pi(l-k+\epsilon_{\mathrm{f}})}}{1-e^{j2\pi(l-k+\epsilon_{\mathrm{f}})/N_{\mathrm{s}}}} + Z[k] \\
&= C_{\epsilon_{\mathrm{f}}}[0]H[k]X[k] + I[k] + Z[k] \qquad (6.4)
\end{aligned}$$

where $H[k]$ and $X[k]$ are the DFT applied $h[i]$ and $x[i]$ respectively. The ICI term, $I[k]$, is given in (6.5)

$$I[k] = \sum_{r=1}^{N_{\mathrm{s}}-1} C_{\epsilon_{\mathrm{f}}}[r]H[k-r]X[k-r] \qquad (6.5)$$

$C_{\epsilon_{\mathrm{f}}}[r]$ can be expressed as in (6.6).

$$C_{\epsilon_{\mathrm{f}}}[r] = \left[\frac{\sin(\pi(\epsilon_{\mathrm{f}}-r))}{N_{\mathrm{s}}\sin(\pi(\epsilon_{\mathrm{f}}-r)/N_{\mathrm{s}})} e^{j\pi(\epsilon_{\mathrm{f}}-r)(N_{\mathrm{s}}-1)/N_{\mathrm{s}}}\right] \qquad (6.6)$$

When both the integer part and the fractional part are present ($|\epsilon_{\mathrm{n}}| \geq 0.5$), the received

sample at the k-th subcarrier is given in (6.7) [84].

$$Y[k] = C_{\epsilon_n}[0]H[k-\epsilon_i]X[k-\epsilon_i] + Z[k] + \sum_{r\neq\epsilon_i}^{N_s-1} C_{\epsilon_n}[r]H[k-r]X[k-r] \tag{6.7}$$

For this case the integer frequency offset correction algorithms such as [85–88] are employed. In this work the integer frequency offset is assumed to be zero and so $|\epsilon_n| < 0.5$. The received symbol at the k-th subcarrier can be written as in (6.8).

$$Y[k] = C_{\epsilon_n}[0]H[k]X[k] + I[k] + Z[k] \tag{6.8}$$

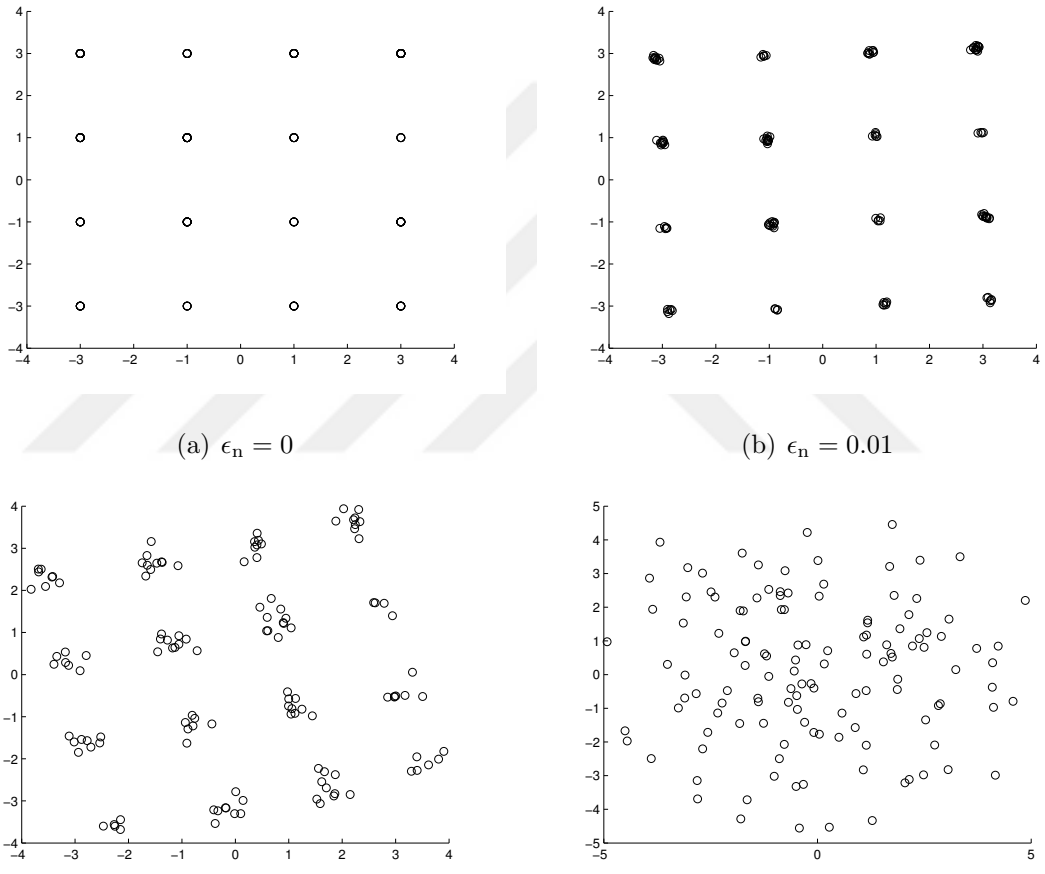
As it is seen from (6.8), the received signal is affected by the CFO. While the magnitude of the signal $H[k]X[k]$ is attenuated by $|[\sin(\pi\epsilon_n)]/[N_s \sin(\pi\epsilon_n/N_s)]|$, its phase is increased by $\pi\epsilon_n(1-1/N_s) + 2\pi\epsilon_n(N_g/N_s)$. Effect of CFO on the constellation can be seen in Figure 6.1. Ignoring the effects of the channel and the noise, it is clear that the amplitude and phase distortion for the 16-QAM rectangular constellation becomes severe as the CFO increases from 0 to 0.3.

6.2. Effect of Carrier Frequency Offset on SNR

Since the CFO attenuates the desired signal and introduces the ICI, the SNR decreases. The SNR degradation can be analyzed for $\epsilon_i = 0$ because the integer frequency offset does not affect the SNR. In order to decode the received symbol $Y[k]$ correctly, the receiver should be able to estimate the phase of $C_{\epsilon_n}[0]H[k]$. Let $C_{\epsilon_n}[0]H[k] = |C_{\epsilon_n}[0]H[k]|e^{j\theta[k]}$ and assume that the phase $\theta[k]$ is estimated perfectly via training preambles or pilot subcarriers. The received symbol $Y[k]$ can be multiplied by $e^{-j\theta[k]}$

$$e^{-j\theta[k]}Y[k] = |C_{\epsilon_n}[0]H[k]|X[k] + I'[k] + Z'[k] \tag{6.9}$$

where the ICI and the noise is multiplied by $e^{-j\theta[k]}$ given as $I'[k] = e^{-j\theta[k]}I[k]$ and $Z'[k] = e^{-j\theta[k]}Z[k]$ respectively [84]. The desired signal in the decision metric (6.9)



(a) $\epsilon_{\mathrm{n}} = 0$ (b) $\epsilon_{\mathrm{n}} = 0.01$

(c) $\epsilon_{\mathrm{n}} = 0.05$ (d) $\epsilon_{\mathrm{n}} = 0.3$

Figure 6.1. 16-QAM constellation for varying CFO.

is $|C_{\epsilon_\mathrm{n}}[0]H[k]|X[k]$ because knowledge of $|C_{\epsilon_\mathrm{n}}[0]H[k]|$ is required for demodulation. Taking the desired signal into consideration, the SNR for the k-th subcarrier can be written as in (6.10).

$$\mathrm{SNR}(\epsilon_\mathrm{n}, k) = \frac{\mathbb{E}\{|C_{\epsilon_\mathrm{n}}[0]H[k]X[k]|^2\}}{\mathbb{E}\{|I'[k] + Z'[k]|^2\}} \tag{6.10}$$

The power of the desired signal can be computed as

$$\begin{aligned}\mathbb{E}\{|C_{\epsilon_\mathrm{n}}[0]H[k]X[k]|^2\} &= |C_{\epsilon_\mathrm{n}}[0]|^2\ \mathbb{E}\{|H[k]|^2\}\ \mathbb{E}\{|X[k]|^2\} \\ &= \sigma_X^2 |C_{\epsilon_\mathrm{n}}[0]|^2\ \mathbb{E}\{|H[k]|^2\}\end{aligned} \tag{6.11}$$

since the channel is independent of the transmitted symbol and $\mathbb{E}\{|X[k]|^2\} = \sigma_X^2$. Assuming that the channel response at time p is uncorrelated with the channel response at time i, $\mathbb{E}\{|H[k]|^2\}$ can be given as

$$\begin{aligned}\mathbb{E}\{|H[k]|^2\} &= \mathbb{E}\left\{\sum_{i=0}^{N_\mathrm{g}}\sum_{p=0}^{N_\mathrm{g}} h[i]h^*[p]e^{-j2\pi(i-p)k/N_\mathrm{s}}\right\} \\ &= \sum_{i=0}^{N_\mathrm{g}} \mathbb{E}\{|h[i]|^2\} = P_h\end{aligned} \tag{6.12}$$

where P_h denotes the total average power of the channel response. Using (6.12) in (6.11), the power of the desired signal is given in (6.13).

$$\mathbb{E}\{|C_{\epsilon_\mathrm{n}}[0]H[k]X[k]|^2\} = \sigma_X^2 P_h |C_{\epsilon_\mathrm{n}}[0]|^2 \tag{6.13}$$

The power of the ICI-plus-noise signal is

$$\mathbb{E}\{|I^{'}[k] + Z^{'}[k]|^2\} = \mathbb{E}\{|I^{'}[k]|^2\} + \mathbb{E}\{|Z^{'}[k]|^2\} \tag{6.14}$$

since the ICI and the noise are independent. Assuming that data symbol in each subcarrier is independent of one another $\mathbb{E}\{X[k-r]X^*[k-p]\} = \sigma_X^2\delta[p-r]$, the ICI

power can be expressed as:

$$\begin{aligned}\mathbb{E}\{|I^{'}[k]|^2\} &= \mathbb{E}\left\{\sum_{r=1}^{N_\mathrm{s}-1}\sum_{p=1}^{N_\mathrm{s}-1} C_{\epsilon_\mathrm{n}}[r]C^{*}_{\epsilon_\mathrm{n}}[p]H[k-r]H^{*}[k-p]X[k-r]X^{*}[k-p]\right\} \\ &= \sigma_X^2 \sum_{r=1}^{N_\mathrm{s}-1} |C_{\epsilon_\mathrm{n}}[r]|^2 \ \mathbb{E}\{|H[k-r]|^2\} \\ &= \sigma_X^2 P_h \sum_{r=1}^{N_\mathrm{s}-1} |C_{\epsilon_\mathrm{n}}[r]|^2 = \sigma_X^2 P_h (1-|C_{\epsilon_\mathrm{n}}[0]|^2) \end{aligned} \tag{6.15}$$

The magnitude of $|C_{\epsilon_\mathrm{n}}[k]|$ is always less than or equal to 1 (6.16).

$$|C_{\epsilon_\mathrm{n}}[k]| = \left|\sum_{i=0}^{N_\mathrm{s}-1} c_{\epsilon_\mathrm{n}}[i]e^{-j2\pi ik/N_\mathrm{s}}\right| \leq \sum_{i=0}^{N_\mathrm{s}-1} \left|c_{\epsilon_\mathrm{n}}[i]e^{-j2\pi ik/N_s}\right| = \sum_{i=0}^{N_\mathrm{s}-1} \frac{1}{N_\mathrm{s}} = 1 \tag{6.16}$$

Due to the DFT property and (6.16), $\sum_{r=0}^{N_\mathrm{s}-1} |C_{\epsilon_\mathrm{n}}[r]|^2 = N_\mathrm{s} \sum_{i=0}^{N_\mathrm{s}-1} |c_{\epsilon_\mathrm{n}}[i]|^2 = 1$. Using (6.12) and (6.15) the power of the ICI-plus-noise signal is given in (6.17).

$$\mathbb{E}\{|I^{'}[k] + Z^{'}[k]|^2\} = (1-|C_{\epsilon_\mathrm{n}}[0]|^2)P_h\sigma_X^2 + \sigma_Z^2 \tag{6.17}$$

The SNR of the k-th subcarrier can be rewritten by putting (6.13) and (6.17) into (6.10) as

$$\mathrm{SNR}(\epsilon_\mathrm{n}) = \frac{\mathbb{E}\{|C_{\epsilon_\mathrm{n}}[0]H[k]X[k]|^2\}}{\mathbb{E}\{|I^{'}[k] + Z^{'}[k]|^2\}} = \frac{|C_{\epsilon_\mathrm{n}}[0]|^2 P_h \sigma_X^2}{(1-|C_{\epsilon_\mathrm{n}}[0]|^2)P_h\sigma_X^2 + \sigma_Z^2} \tag{6.18}$$

where $|C_{\epsilon_\mathrm{n}}[0]|^2 = [\sin(\pi\epsilon_\mathrm{n})]/[N_\mathrm{s}\sin(\pi\epsilon_\mathrm{n}/N_\mathrm{s})]$ due to (6.6) and the argument k of SNR is dropped since it is the same for all subcarriers. The SNR degradation caused by the carrier frequency offset can be found as

$$D(\epsilon_\mathrm{n}) = \frac{\mathrm{SNR}(0)}{\mathrm{SNR}(\epsilon_\mathrm{n})} = \frac{1+(1-|C_{\epsilon_\mathrm{n}}[0]|^2)}{|C_{\epsilon_\mathrm{n}}[0]|^2}\mathrm{SNR}(0) \tag{6.19}$$

where the nominal SNR in the absence of carrier frequency offset is given in (6.20).

$$\mathrm{SNR}(0) = \frac{P_h\sigma_X^2}{\sigma_Z^2} \tag{6.20}$$

As it can be seen from (6.19), the frequency offset decreases the signal power by $|C_l(0)|^2$ and increases interference power by the amount of the decreased power. The SNR degradation is plotted as a function of the relative frequency offset for $-0.5 < \epsilon_{\rm n} < 0.5$ in Figure 6.2. The nominal SNR is changing between -5dB to 20dB. As the frequency offset increases, the SNR degradation also increases. When the nominal SNR is high, the noise power is small and the effect of ICI becomes more visible.

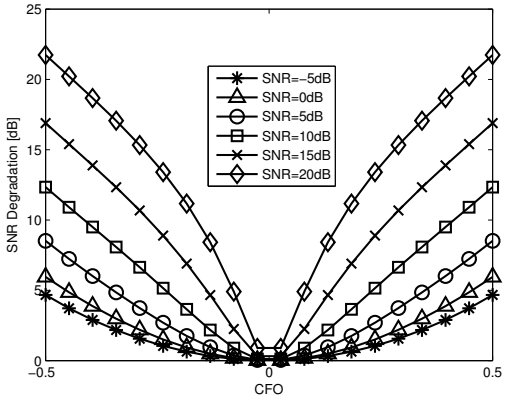


Figure 6.2. SNR degradation versus CFO.

6.3. Proposed Method

The DFT applied received signal samples can be represented in matrix form as

$$\tilde{\mathbf{y}} = \mathbf{C}_{\epsilon_{\rm n}}\mathbf{X}\mathbf{F}\begin{bmatrix} h[0] \\ \vdots \\ h[N_{\rm s}-1] \end{bmatrix} + \begin{bmatrix} Z[0] \\ \vdots \\ Z[N_{\rm s}-1] \end{bmatrix} \tag{6.21}$$

where $\mathbf{F}$ denotes the $N_{\rm s}$-point DFT matrix, the frequency domain received signal is $\tilde{\mathbf{y}} = [Y[0]\cdots Y[N_{\rm s}-1]]^T$, the pilot symbol matrix is $\mathbf{X} = \text{diag}\{X[0], ..., X[N_{\rm s}-1]\}$ and $\mathbf{C}_{\epsilon_{\rm n}}$ is given in (6.22).

$$\mathbf{C}_{\epsilon_{\rm n}} = \begin{bmatrix} C[0] & C[N_{\rm s}-1] & \cdots & C[1] \\ C[1] & C[0] & \cdots & C[2] \\ \vdots & \ddots & \ddots & \vdots \\ C[N_{\rm s}-1] & C[N_{\rm s}-2] & \cdots & C[0] \end{bmatrix} \tag{6.22}$$

Assuming that channel length is smaller than the number of subcarriers $L \ll N_{\mathrm{s}}$ and the number of cyclic prefix is chosen as $N_{\mathrm{g}} \geq L$, the channel coefficient vector is zero-padded as $\mathbf{h} = [h[0] \cdots h[N_{\mathrm{g}}-1], \mathbf{0}_{1\times N_{\mathrm{s}}-N_{\mathrm{g}}}]^T$. The zeros can be removed by using the first N_{g} columns of the N_{s}-point DFT matrix $\mathbf{F}$ that is $\mathbf{F}_{N_{\mathrm{g}}}$

$$\begin{aligned} \tilde{\mathbf{y}} &= \mathbf{C}_{\epsilon_{\mathrm{n}}}\mathbf{X}\mathbf{F}_{N_{\mathrm{g}}} \begin{bmatrix} h[0] \\ \vdots \\ h[N_{\mathrm{g}}-1] \end{bmatrix} + \mathbf{z} \\ &= \widetilde{\mathbf{A}}\mathbf{h} + \mathbf{z} \end{aligned} \tag{6.23}$$

where $\widetilde{\mathbf{A}} = \mathbf{C}_{\epsilon_{\mathrm{n}}}\mathbf{X}\mathbf{F}_{N_{\mathrm{g}}}$, $\mathbf{z} = [Z[0] \cdots Z[N_{\mathrm{s}}-1]]^T$, and $\mathbf{h} = [h[0] \cdots h[N_{\mathrm{g}}-1]]^T$.

A search space for the fractional frequency offset can be created by a set of d values $\{\tilde{\epsilon}_0, \tilde{\epsilon}_1, ..., \tilde{\epsilon}_{d-1}\} \in [-0.5, 0.5]$. Each value $\tilde{\epsilon}_i$ corresponds to a $\mathbf{C}_{\tilde{\epsilon}_i}$ and the received samples can be rewritten as

$$\begin{aligned} \tilde{\mathbf{y}} &= [\mathbf{C}_{\tilde{\epsilon}_0}\mathbf{X}\mathbf{F}_{N_{\mathrm{g}}} \cdots \mathbf{C}_{\tilde{\epsilon}_{d-1}}\mathbf{X}\mathbf{F}_{N_{\mathrm{g}}}]\,\mathbf{x} + \mathbf{z} \\ &= \mathbf{A}\mathbf{x} + \mathbf{z} \end{aligned} \tag{6.24}$$

where the vector $\mathbf{x}$ is built by stacking d blocks of $N_{\mathrm{g}} \times 1$ channel vectors, $\mathbf{x} = [\mathbf{h}_0 \ldots \mathbf{h}_{d-1}]$, and the matrix $\mathbf{A}$ is constructed by concatenating d matrices each corresponding to a $\mathbf{C}_{\tilde{\epsilon}_i}$. Each block of $\mathbf{x}$ is represented as $\mathbf{x}[i] = \mathbf{h}_i = [h[0], \ldots, h[N_{\mathrm{g}}-1]]^T$. In this representation $\mathbf{x}$ can be considered as a 1 block-sparse vector. Thus, a joint estimator of CFO and CIR using the CS framework can be constructed in this way. The proposed framework allows solving for both CFO and CIR simultaneously by using the CS recovery algorithms for block-sparse signals. There exists numerous implementations of the CS based block-sparse signal recovery methods for the estimation of the CIR of the OFDM systems [15–17]. In these works, block sparsity is achieved by either assuming that the several channel instantiations are group-sparse, locations of the nonzero channel coefficients are same [17] or concatenating multiple CIRs of different antennas with common support in a block sparse structure [15,16]. However, frequency

offset is not considered in these works, which assume perfect synchronization for the OFDM system and so only CIR can be estimated using these methods. The main advantages of the proposed method are the following:

- The proposed method only needs one OFDM block of training symbols unlike Moose, Classen and the p-algorithm. The use of multiple blocks makes the estimation susceptible to changes in the channel or the CFO. Also the use of more blocks means an increase in pilot overhead.
- The proposed method does not require any initialization or the second-order statistics of the channel and noise unlike EM based methods. Apriori knowledge about either the channel or the noise may not be available for every case.
- The performance of the proposed method does not depend on the value of the CFO.

6.4. Numerical Results

Numerical results are presented for three scenarios: time invariant sparse multipath channel, time varying sparse multipath channel and time varying sparse multipath with varying CFO. The number of carriers of the OFDM system is chosen as $N_{\mathrm{s}} = 128$. The length of the sparse multipath channel is fixed to $L = 20$ with $s = 6$ nonzero coefficients for all cases. The nonzero coefficients of the channel are generated from independent complex Gaussian with variances set according to an exponential decaying power delay profile. The CP length is set to $N_{\mathrm{g}} = 25$ to prevent IBI. The mean squared error of the frequency estimates (MSE CFO) and channel estimates (MSE CIR) are given for varying SNR $= 10\log_{10}(\|\mathbf{y}\|_2^2/\|\mathbf{z}\|_2^2)$ and 1000 Monte Carlo iterations. The pilots symbols are randomly chosen from 16-QAM constellation and then are fixed for each Monte Carlo iteration.

6.4.1. Time Invariant Sparse Multipath Channel

The sparse multipath channel is assumed to be fixed between two consecutive OFDM blocks. The magnitude of the generated time invariant sparse multipath chan-

nel is shown in Figure 6.3 and it can be seen that only $s = 6$ coefficients of the CIR are nonzero.

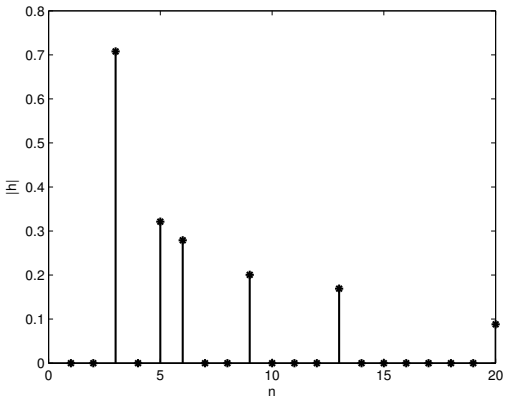
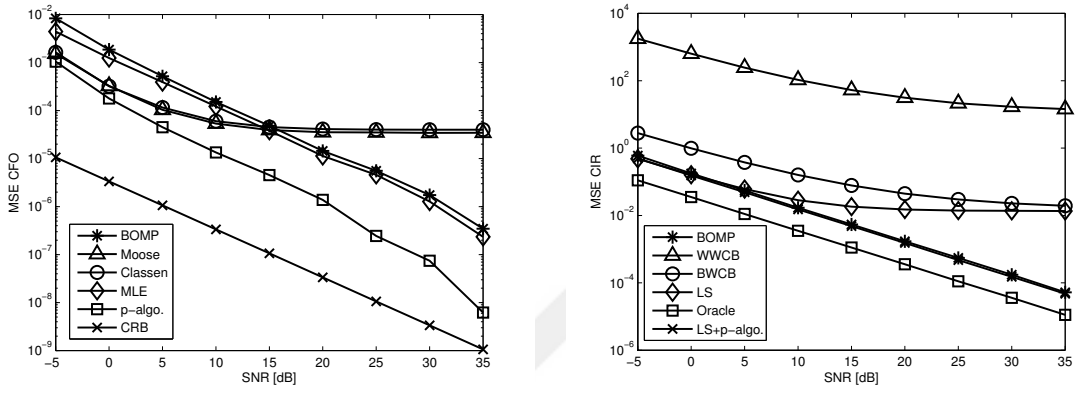


Figure 6.3. Impulse response of the time invariant sparse multipath channel.

MSE obtained using the BOMP algorithm is compared to pilot symbol based Moose [74], Classen [75] methods, the approximate MLE [82] and the p-algorithm [78] for CFO value $\epsilon_{\mathrm{n}} = 0.03$. The search space of the BOMP, the approximate MLE and the p-algorithm is assigned to the set of 500 uniformly sampled values within the interval $[-0.5, 0.5]$. The actual CFO value, $\epsilon_{\mathrm{n}} = 0.03$, is included in the search grid to prevent error floors in the MSE of the CFO estimates. The block coherence of the dictionary built using the search grid $\mu_B(\mathbf{A}) = 0.04$ satisfy the recovery condition given in Equation (3.8) for $N_{\mathrm{g}} = 25$. MSE of the CFO estimates are illustrated in Figure 6.4(a). Moose and Classen methods perform better than the BOMP and the MLE for SNR levels below 15dB. Moose and Classen methods ignore the ICI and so when the SNR is low, ignoring the ICI term does not degrade the performance significantly. However, as the SNR increases beyond 15dB the Moose and Classen methods reach an error floor. The performance of the BOMP algorithm is close to the performance of the MLE which is implemented by using the actual channel and noise statistics. Since in practice it will also be neccessary to estimate both the channel and noise statistics, the performance of the MLE given in Figure 6.4(a) can be considered an ideal case. It is seen that the best performance is achieved by the p-algorithm [78]. Although the p-algorithm makes use of two consecutive OFDM blocks like Moose and Classen methods, it is based on cost function minimization using a search space and so it performs better than Moose and Classen methods. The Cramer-Rao bound (CRB) (Figure 6.4(a))

given in [89] represent a lower bound for the maximum likelihood estimation using one OFDM block of pilot symbols. CRB is used as a benchmark for evaluating different CFO estimators and the closest performance to CRB belongs to the p-algorithm.



(a) MSE of CFO estimates for $\epsilon_{\mathrm{n}} = 0.03$ (b) MSE of CIR estimates for $\epsilon_{\mathrm{n}} = 0.03$

Figure 6.4. MSE of CFO and CIR estimates for time invariant sparse channel.

Figure 6.4(b) shows the MSE of CIR estimates. The performance of the least-squares (LS) CIR estimate can be improved by first applying a CFO synchronization method. LS method combined with the p-algorithm can be considered a two step joint CFO and CIR estimator. Figure 6.4(b) indicates that the two step LS estimator gives the best performance and the BOMP method performs closely to the two step LS estimator. The best possible channel estimate is represented by the oracle which is computed using the actual CFO value and the support of the channel. The closest performance to the oracle is achieved by the two step LS estimator and the BOMP method. Using the perturbation analysis (Chapter 4), two upper bounds for the channel estimation performance of the OFDM system with CFO $\epsilon_{\mathrm{n}} = 0.03$ can be obtained. These bounds are plotted as WWCB and BWCB (Figure 6.4(b)) which stands for worst worst-case upper bound given in Equation (4.36) and best worst-case upper bound given in Equation (4.27) respectively. The coefficients for the WWCB are computed as $\varepsilon_{\mathbf{A}}^{(s)} = 0.1654$, $\varepsilon_{\mathbf{A}}^{(2s)} = 0.1710$, $\widetilde{C}_1 = 45.08$ and $\kappa_{\mathbf{A}}^{(s)} = 1$. Since the exact computation of the RIC δ_s and δ_{2s} is not possible, $\sigma_{\max}^{(s)}(\mathbf{A})$ and $\sigma_{\min}^{(s)}(\mathbf{A})$ are computed to obtain a bound for the RIC. The bound obtained is on the order of 10^{-15} making the RIC practically zero. WWCB shows the worst possible channel estimation performance for a OFDM system with CFO $\epsilon_{\mathrm{n}} = 0.03$ and a sparse multipath channel with $s = 6$

nonzero coefficients. BWCB gives a much tighter bound than WWCB since it makes use of the channel tap values in its computations. The advantage of WWCB is that its computation only requires the sparsity level of the multipath channel.

In order to demonstrate that the performance of the proposed joint estimator does not depend on the CFO value, the SNR is fixed at 15dB and the CFO is varied in a set of 30 uniformly spaced values within the interval $[-0.5, 0.5]$. As the value of CFO increases, MSE of Moose and Classen methods increase due to their sensitivity to the value of the CFO while MSE of the BOMP method does not increase as seen in Figure 6.5. Classen's method assumes a sufficiently small CFO in order for ICI to be neglected so the performance of the Classen's method degrades as the CFO increases. Only for very small CFO values, $\epsilon_{\mathrm{n}} \in [-0.0172, 0.0172]$, the performances of the Moose and Classen methods are better than the BOMP method. The BOMP method and the p-algorithm do not exhibit any sensitivity to the value of the CFO and their performances do not depend on the value of the CFO. The p-algorithm gives the closest performance to the CRB but needs at least two OFDM blocks like the Moose and Classen methods.

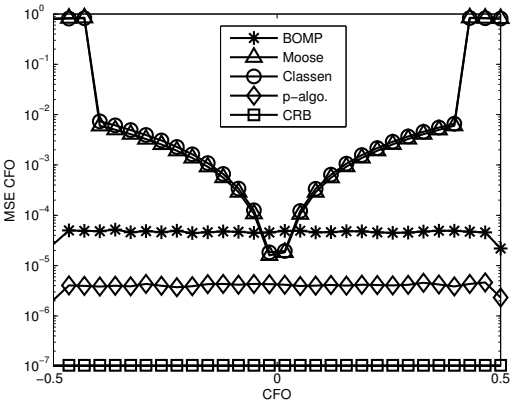
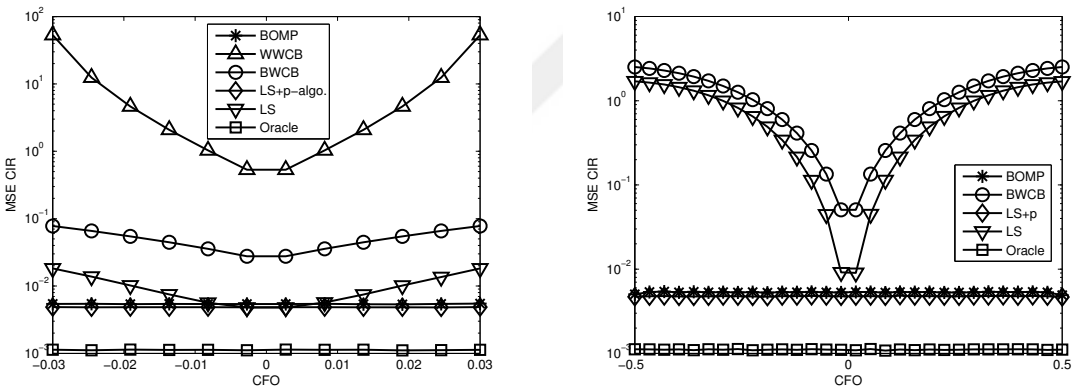


Figure 6.5. MSE of CFO estimates for time invariant channel and 15dB SNR.

MSE of CIR estimates are given first for a smaller CFO range of $[-0.03, 0.03]$ as seen in Figure 6.6(a). WWCB cannot be computed for $\epsilon_{\mathrm{n}} > 0.0306$ since the relative bound on the perturbation must satisfy $\varepsilon_{\mathbf{A}}^{(2s)} < 0.1892$ and so the CFO range is assigned to the set of 12 uniformly sampled values within the interval $[-0.03, 0.03]$. For this range of smaller CFO values, it is observed that the WWCB becomes much looser as the CFO increases. Thus the WWCB can be considered much more sensitive to CFO

value than the BWCB. The best performance is achieved by the two step LS estimator, while the performance of the BOMP method is close. As expected, the oracle bound is the lower bound for the sparse channel estimation and is not subject to any change by CFO value since it is provided with the actual CFO value and support of the sparse channel. Figure 6.6(b) gives the MSE of CIR estimates versus the extended range of CFO values, $[-0.5, 0.5]$. Of the two upper bounds only the BWCB can be shown for the CIR estimate within this range since it is not possible to calculate the WWCB for $\epsilon_{\mathrm{n}} > 0.0306$. Although the BWCB is a much tighter bound and can be computed for the full CFO range, the computation of the BWCB requires the prior knowledge of the channel. On the other hand, the computation of the WWCB does not require the channel information. Best performance is achieved by the two step LS estimator combined with the p-algorithm while the proposed BOMP is very close.



(a) $\epsilon_{\mathrm{n}} \in [-0.03, 0.03]$ (b) $\epsilon_{\mathrm{n}} \in [-0.5, 0.5]$

Figure 6.6. MSE of CIR estimates for time invariant channel and 15dB SNR.

6.4.2. Correlated Time Varying Sparse Multipath Channel

Depending on the mobility of the receiver, the channel may remain essentially constant over the duration of the block, or may be slowly time varying. For time varying sparse channels the CIR cannot be assumed to be static over two consecutive OFDM blocks. When the channel is time varying, the nonzero channel coefficients may vary slowly and are temporally correlated. The parameters of the channel may remain constant over several OFDM blocks. Consequently the locations of the nonzero channel coefficients coincide in successive channel instantiations. The temporal variations of

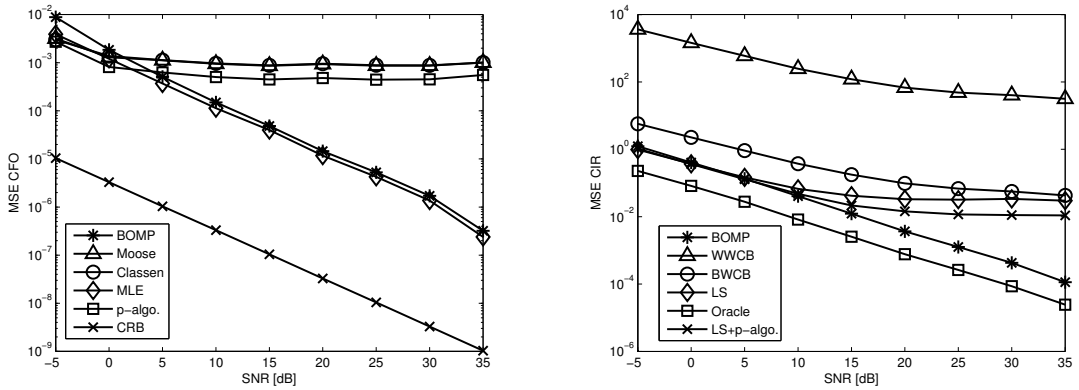
such a channel can be captured by an autoregressive (AR) state space model. It is shown in that the local behaviour of fading wireless channels is accurately represented by the first order AR model [90]. The temporally correlated time varying sparse channel is generated using a first order AR model given as

$$\mathbf{h}_{l+1} = \rho \mathbf{h}_l + \mathbf{u}_l \tag{6.25}$$

where $l = 1, 2, ...$ denotes the OFDM block, $\mathbf{u}_l$ denotes the driving noise and ρ is calculated using the zero order Bessel function of the first kind, $\rho = J_0(2\pi\Delta f_c T)$ [17]. Each entry of the driving noise is generated independently from zero mean complex white Gaussian distribution with variance $(1-\rho^2)\varsigma_i$ [17,91]. The sparse channel is initialized using multivariate complex white Gaussian distribution with the covariance matrix $\mathbf{\Omega} = \text{diag}\,(\varsigma_1, ..., \varsigma_L)$ where the variances are set according to the same exponential power decay profile used in Section 6.4.1.

MSE of CFO estimates for the time varying sparse channel are shown in Figure 6.7(a). For SNR=-5 dB the p-algorithm yields lower MSE than the MLE and the BOMP methods. As SNR increases beyond -5 dB, the performance of the the p-algorithm stays the same. The performances of the Moose, Classen and the p-algorithm cannot be improved with increasing SNR due to the time varying channel. The p-algorithm performs better than the Moose and Classen methods. The performances of the MLE and the BOMP methods are not affected by the time varying sparse channel. Figure 6.7(b) shows the MSE of CIR estimates for the time varying sparse channel. The pure LS estimator without any CFO synchronization performs closely to the BOMP method for SNR below 10 dB. As SNR increases beyond 10 dB, the MSE of the pure LS estimator reaches an error floor seen in Figure 6.7(b). If the results illustrated in Figure 6.7(b) are compared to the results of the time invariant sparse channel (Figure 6.4(b)), the BWCB is seen to be a much tighter bound as the SNR increases for the time varying sparse channel. The upper bounds WWCB and BWCB exhibit an error floor as the SNR increases for the time varying sparse channel. Degradation due to time varying channel is not observed for the BOMP method since the BOMP method

does not use a channel model.



(a) MSE of CFO estimates for $\epsilon_{\mathrm{n}} = 0.03$ (b) MSE of CIR estimates for $\epsilon_{\mathrm{n}} = 0.03$

Figure 6.7. MSE of CFO and CIR estimates for time varying sparse channel.

MSE of CFO estimates versus the CFO range of $[-0.5, 0.5]$ are presented in Figure 6.8. The p-algorithm no longer holds the performance advantage but manages to perform better than the Moose and Classen methods. As the CFO increases, the performance of the Moose and Classen methods degrades severely. Figure 6.8 shows that the Moose and Classen methods are prone to greater performance degradation for larger CFO values when the sparse channel is time varying. The p-algorithm does not show sensitivity to CFO value.

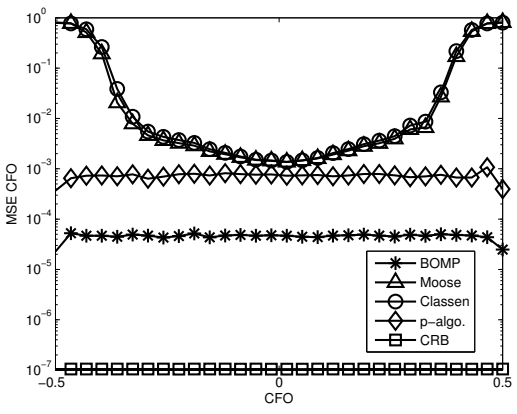
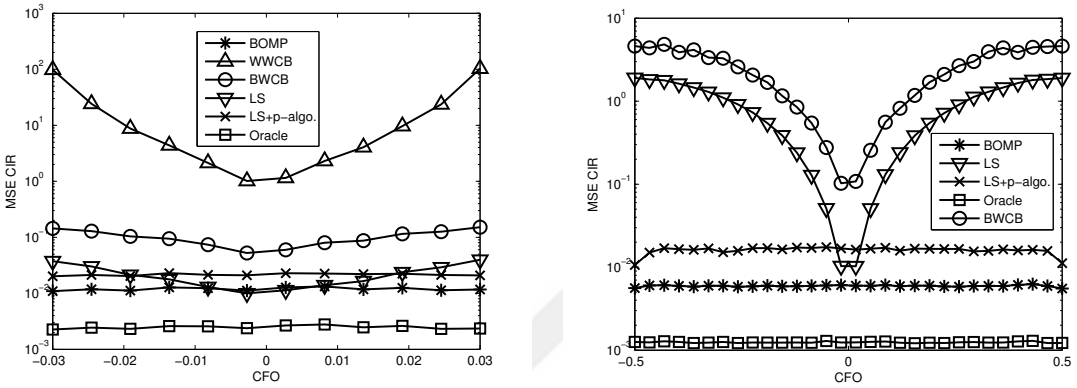


Figure 6.8. MSE of CFO estimates for time varying sparse channel and 15dB SNR.

WWCB for the time varying sparse channel is illustrated in Figure 6.9(a). It is seen that the performance the pure LS method is much better than the performance of the two step LS estimator combined with the p-algorithm. MSE of CIR estimates

versus the extended CFO range $[-0.5, 0.5]$ are presented in Figure 6.9(b). For smaller CFO values the pure LS channel estimate is better than the two step LS estimate using the p-algorithm synchronization.



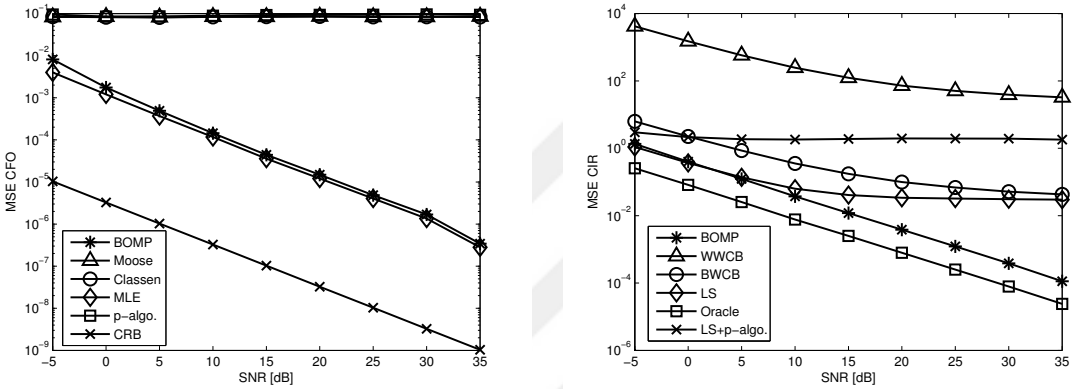
(a) $\epsilon_n \in [-0.03, 0.03]$ (b) $\epsilon_n \in [-0.5, 0.5]$

Figure 6.9. MSE of CIR estimates versus CFO for 15dB SNR.

6.4.3. Uncorrelated Time Varying Sparse Multipath Channel

The support, the nonzero taps, of the sparse multipath channel are generated from uniform distribution while the nonzero coefficients are generated from standard Gaussian distribution for each OFDM block. MSE of CFO estimates for this case are shown in Figure 6.10(a). The performances of the Moose, Classen and the p-algorithm does not improve as the SNR increases if the sparse channel changes between two consecutive OFDM blocks. All of the three methods which rely on consecutive OFDM blocks, yield the worst performance. The performances of the MLE and the BOMP methods stay consistent with the performances given for the time invariant channel since they use only one OFDM block and are not affected by the changing channel. Figure 6.10(b) shows the MSE of CIR estimates for the time varying channel. It is observed that the performance of the two step LS estimator using the p-algorithm is worse than the pure LS estimate which does not use any CFO synchronization. The performance of the two step LS estimator is observed to be even worse than the BWCB for SNR greater than 0 dB. The pure LS estimator without any CFO synchronization performs closely to the BOMP method for SNR below 10 dB. As SNR increases beyond 10 dB, the MSE of the pure LS estimator reaches an error floor as seen in Figure 6.10(b).

Figure 6.10(b) indicates that the upper bounds WWCB and BWCB exhibit an error floor as the SNR increases for the time varying sparse channel. Figure 6.10(b) shows that the performance of the BOMP method does not exhibit any degradation for the time varying channel as a consequence of not relying on a channel model. The BOMP method gives the closest performance to the oracle bound for time varying sparse channel.



(a) MSE of CFO estimates for $\epsilon_{\mathrm{n}} = 0.03$ (b) MSE of CIR estimates for $\epsilon_{\mathrm{n}} = 0.03$

Figure 6.10. MSE of CFO and CIR estimates for uncorrelated time varying sparse channel.

MSE of CFO estimates versus the CFO range of $[-0.5, 0.5]$ are presented in Figure 6.11. The Moose, Classen, and the p-algorithm gives the worst performance as seen in Figure 6.11. The closest performance to the CRB bound is achieved by the BOMP method.

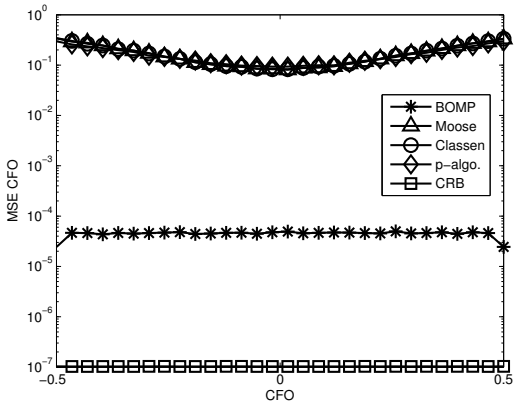
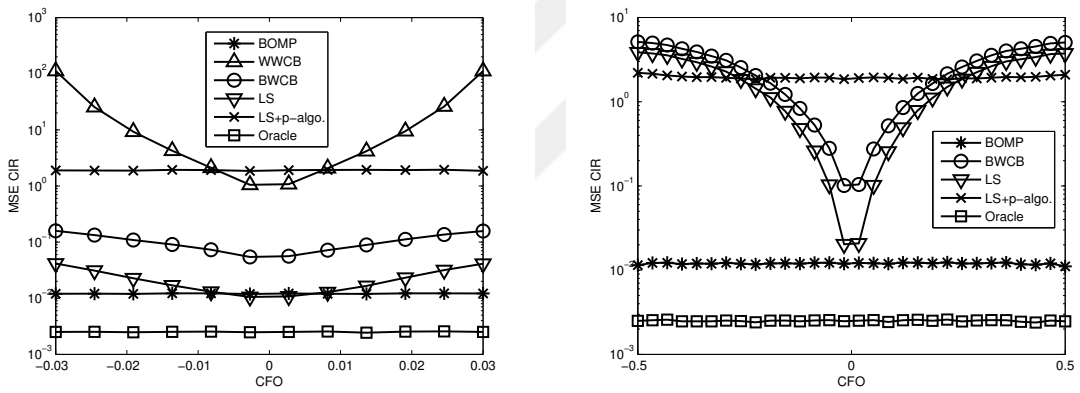


Figure 6.11. MSE of CFO estimates for time varying sparse channel and 15dB SNR.

WWCB for the time varying sparse channel is illustrated in Figure 6.12(a). When the sparse channel is time varying, it is seen that the performance the pure LS method is much better than the performance of the two step LS estimator combined with the p-algorithm. MSE of CIR estimates versus the extended CFO range $[-0.5, 0.5]$ are presented in Figure 6.12(b). For smaller CFO values the pure LS channel estimate yields lower MSE than the two step LS estimate using the p-algorithm synchronization. The closest performance to the oracle bound versus the whole CFO range is achieved by the BOMP method. For smaller CFO values the pure LS channel estimate yields lower MSE than the two step LS estimate using the p-algorithm synchronization. The closest performance to the oracle bound versus the whole CFO range belongs to the BOMP method.



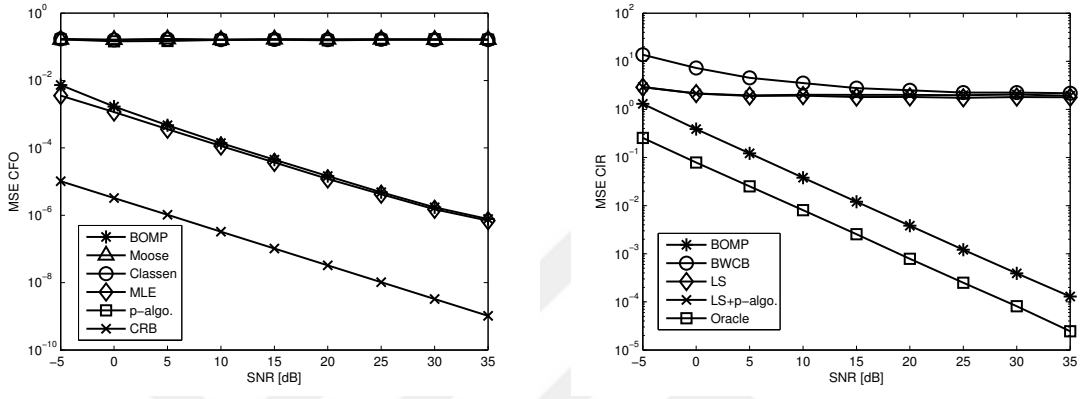
(a) $\epsilon_{\mathrm{n}} \in [-0.03, 0.03]$ (b) $\epsilon_{\mathrm{n}} \in [-0.5, 0.5]$

Figure 6.12. MSE of CIR estimates versus CFO for 15dB SNR.

6.4.4. Time Varying Sparse Multipath Channel with varying CFO

CFO may also be subject to change between each successive OFDM block. In order to give numerical results for this case, CFO is assumed to be uniformly distributed in the interval, $\epsilon_{\mathrm{n}} \sim U(-0.03, 0.03)$, for each successive OFDM block as in [89]. The sparse multipath channel is time varying as in Section 6.4.2. Figure 6.13(a) gives the MSE of CFO estimates versus SNR for varying channel and CFO. The Moose, Classen, and p-algorithm perform worse than the MLE and the BOMP methods since both CFO and CIR is changing for each OFDM block. The BOMP and the MLE methods give the same performance and are not affected by the changing CFO and CIR. Figure 6.13(b)

shows the MSE of CIR estimates. The pure LS method and the LS method combined with the CFO estimate provided by the p-algorithm produces the same MSE. They are subject to the same error floor as the BWCB for SNR levels above 25 dB. The closest performance to the oracle estimator is achieved by the BOMP method.



(a) $\epsilon_{\mathrm{n}} \sim U(-0.03, 0.03)$ (b) $\epsilon_{\mathrm{n}} \sim U(-0.03, 0.03)$

Figure 6.13. MSE of CFO and CIR estimates for time varying sparse channel.

7. CONCLUSION

In this thesis sparse signal recovery from incomplete and perturbed data is investigated. Sparse signal recovery allows the representation of signals using far less measurements than the tradional Nyquist/Shannon sampling theorem requires. For this reason there has been a great interest in this subject and it has found many applications. The focus of this thesis is on two applications of sparse signal recovery: electromagnetic inverse scattering and joint estimation of CFO and CIR for SISO OFDM systems. A brief review on the uniqueness and the stability analysis of the sparse solutions is given in Chapter 2. Spark, mutual coherence, restricted isometry property and null space property and their relations with each other are explained. The sparse recovery algorithms are presented in three categories: greedy algorithms, the convex relaxation by ℓ_1 minimization, and the Bayesian compressed sensing method (Chapter 3). While the covered greedy pursuits consist of the OMP, FTB-OMP and BOMP methods, interior point methods and the NESTA methods are explained under ℓ_1 minimization. The Bayesian compressed sensing approach is covered under its own section. The performance bounds of the sparse recovery methods are given for additive and multiplicative noise in Chapter 4. A best case and a worst case analysis is provided for each type of noise.

The electromagnetic inverse scattering problem is about the reconstruction of the electromagnetic properties of the unknown targets from measured scattered field and it is a nonlinear ill-posed problem. With the introduction of the CS framework, the sparse signal recovery methods such as NESTA and MT-BCS have quickly found applications in electromagnetic imaging problems. However, the application of greedy pursuit based sparse signal recovery methods has not been considered yet to the best of our knowledge. We propose the application of the OMP and the FTB-OMP algorithms to increase the reconstruction resolution of electromagnetic scattering problems. Our imaging setup is focused on closely placed point-like scatterers. The reconstruction performance of the proposed greedy based recovery methods are compared against those of the other sparse recovery methods including NESTA and MT-BCS that had

already been applied in the literature. The imaging setup is investigated under two cases: a far proximity case with two scatterers where the objects are separated by ξ and a close proximity case with four scatterers where the separation is decreased to $\xi/8$. For the far proximity case it is observed that the the performance of the OMP algorithm betters the performances of the NESTA and the MT-BCS methods in high SNR region where the SNR is greater than 15dB. Considering the low computational complexity of the OMP compared to NESTA and MT-BCS, the application of the OMP may be preferable for these scenarios. In order to present a more challanging setup the number of scatterers is increased to four and the separation distance is decreased to $\xi/8$ for the close proximity case. The numerical results of this case show that the FTB-OMP benefits from the implementation of the tree based search and increases the resolution of the reconstruction compared to the OMP method though at a cost of higher computational complexity. It is seen that the superior performance belongs to the FTB-OMP in the very low noise region. For both the far and the close promity cases, the NESTA method gives the best reconstruction performance for high noise region where the SNR is less than 15dB.

A joint estimator of CFO and CIR using the CS framework of block-sparse signals is proposed for OFDM systems. A dictionary is created by concatenating matrices of OFDM training symbols perturbed by CFO values taken from a search space and then this dictionary is used to represent the unknown CIR as a 1 block-sparse signal. The proposed framework allows solving for both CFO and CIR simultaneously using the CS recovery algorithms for block-sparse signals. There exists numerous implementations of the CS based block-sparse signal recovery methods for the estimation of the CIR of the OFDM systems [15–17]. In these works, the block sparsity is achieved by either assuming that the several channel instantiations are group-sparse, locations of the nonzero channel coefficients are same, [17] or concatenating multiple CIRs of different antennas with common support in a block sparse structure [15, 16]. However, CFO is not considered in these works, which assume perfect synchronization for the OFDM system and so only the CIR can be estimated using these methods. The proposed framework proves advantageous when compared to these other CS based methods since they cannot estimate the CFO and the CIR jointly. Comparisons against the classical

joint estimation methods can also be made. The proposed method does not require any initialization or the second-order statistics of the channel and noise unlike EM based methods or the MLE method. This is important since apriori knowledge about either the channel or the noise may not be available for every case. The proposed method only needs one OFDM block of training symbols while the methods such as the Moose, Classen and the p-algorithm needs more than one OFDM block of training symbols. The use of multiple blocks makes the estimation susceptible to changes in the channel or the CFO. Also the use of more blocks means an increase in pilot overhead. The performance of the proposed method does not depend on the value of the CFO like the MLE method. The BOMP algorithm is used to reconstruct the CIR coefficients. Our work make use of the worst-case bounds of the perturbed CS recovery. Worst case analysis using perturbation bounds from CS theory are applied to give bounds for sparse channel estimation. These bounds provide a way to observe how the performance of the sparse channel estimation methods scales with the perturbation due to frequency offset. To compare the advantages and disadvantages of the proposed framework three different cases including the time invariant sparse channel, time varying sparse channel and time varying sparse channel with varying CFO are investigated. For the simulation of the time varying sparse channel case, both a slowly changing AR based model and an agressively changing random channel model is used. For the time invariant sparse channel numerical results show that the proposed estimator gives the same CFO estimation performance as the MLE while the p-algorithm provides the best performance. When the multipath channel is time varying, the performances of the Moose, Classen and the p-algorithm all degrade as expected.

REFERENCES

1. Shannon, C. E., "Communication in the Presence of Noise", *Proceedings of the IRE*, Vol. 37, No. 1, pp. 10–21, Jan 1949.

2. Nyquist, H., "Certain Topics in Telegraph Transmission Theory", *Transactions of the American Institute of Electrical Engineers*, Vol. 47, No. 2, pp. 617–644, April 1928.

3. Whittaker, E. T., "XVIII.—On the Functions which are represented by the Expansions of the Interpolation-Theory", *Proceedings of the Royal Society of Edinburgh*, Vol. 35, pp. 181–194, 1 1915.

4. Donoho, D., "Compressed sensing", *IEEE Trans. Inf. Theory*, Vol. 52, No. 4, pp. 1289–1306, April 2006.

5. Candès, E. J., J. Romberg and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information", *IEEE Trans. Inf. Theory*, Vol. 52, No. 2, pp. 489–509, Feb 2006.

6. Marengo, E., "Inverse Scattering by Compressive Sensing and Signal Subspace Methods", *IEEE Int. Workshop on CAMPSAP*, pp. 109–112, Dec 2007.

7. Pan, L., X. Chen and S. P. Yeo, "A Compressive-Sensing-Based Phaseless Imaging Method for Point-Like Dielectric Objects", *IEEE Trans. Antennas. Propag.*, Vol. 60, No. 11, pp. 5472–5475, Nov 2012.

8. Bevacqua, M. T., L. Crocco, L. D. Donato and T. Isernia, "Microwave Imaging of Nonweak Targets via Compressive Sensing and Virtual Experiments", *IEEE Antennas Wireless Propag. Letters*, Vol. 14, pp. 1035–1038, 2015.

9. Wu, Q., Y. Zhang, F. Ahmad and M. Amin, "Compressive-Sensing-Based High-

Resolution Polarimetric Through-the-Wall Radar Imaging Exploiting Target Characteristics", *IEEE Trans. Antennas. Propag.*, Vol. 14, pp. 1043–1047, 2015.

10. Poli, L., G. Oliveri and A. Massa, "Microwave Imaging Within the First-Order Born Approximation by Means of the Contrast-Field Bayesian Compressive Sensing", *IEEE Trans. Antennas. Propag.*, Vol. 60, No. 6, pp. 2865–2879, June 2012.

11. Oliveri, G., P. Rocca and A. Massa, "A Bayesian-Compressive-Sampling-Based Inversion for Imaging Sparse Scatterers", *IEEE Trans. Geoscience and Remote Sensing*, Vol. 49, No. 10, pp. 3993–4006, Oct 2011.

12. Oliveri, G., N. Anselmi and A. Massa, "Compressive Sensing Imaging of Non-Sparse 2D Scatterers by a Total-Variation Approach Within the Born Approximation", *IEEE Trans. Antennas. Propag.*, Vol. 62, No. 10, pp. 5157–5170, Oct 2014.

13. Poli, L., G. Oliveri, P. Rocca and A. Massa, "Bayesian Compressive Sensing Approaches for the Reconstruction of Two-Dimensional Sparse Scatterers Under TE Illuminations", *IEEE Trans. Geoscience and Remote Sensing*, Vol. 51, No. 5, pp. 2920–2936, May 2013.

14. Poli, L., G. Oliveri, F. Viani and A. Massa, "MT-BCS-Based Microwave Imaging Approach Through Minimum-Norm Current Expansion", *IEEE Trans. Antennas. Propag.*, Vol. 61, No. 9, pp. 4722–4732, Sept 2013.

15. Barbotin, Y., A. Hormati, S. Rangan and M. Vetterli, "Estimation of Sparse MIMO Channels with Common Support", *IEEE Trans. Comm.*, Vol. 60, No. 12, pp. 3705–3716, December 2012.

16. Shen, W., L. Dai, Z. Gao and Z. Wang, "Spatially correlated channel estimation based on block iterative support detection for massive MIMO systems", *Electronics Letters*, Vol. 51, No. 7, pp. 587–588, 2015.

17. Prasad, R., C. Murthy and B. Rao, "Joint Approximately Sparse Channel Estimation and Data Detection in OFDM Systems Using Sparse Bayesian Learning", *IEEE Trans. Sig. Process.*, Vol. 62, No. 14, pp. 3591–3603, July 2014.

18. Donoho, D. and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization", *Proceedings of the National Academy of Sciences*, Vol. 100, No. 5, pp. 2197–2202, 2003.

19. Tillmann, A. and M. Pfetsch, "The Computational Complexity of the Restricted Isometry Property, the Nullspace Property, and Related Concepts in Compressed Sensing", *IEEE Trans. Inf. Theory*, Vol. 60, No. 2, pp. 1248–1259, Feb 2014.

20. Elad, M., *Sparse and Redundant Representations*, Springer, 2010.

21. Tropp, J., "Just relax: convex programming methods for identifying sparse signals in noise", *IEEE Trans. on Inf. Theory*, Vol. 52, No. 3, pp. 1030–1051, March 2006.

22. Donoho, D., M. Elad and V. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise", *IEEE Trans. Inf. Theory*, Vol. 52, No. 1, pp. 6–18, Jan 2006.

23. Donoho, D. and X. Huo, "Uncertainty principles and ideal atomic decomposition", *IEEE Trans. Inf. Theory*, Vol. 47, No. 7, pp. 2845–2862, Nov 2001.

24. Kovacevic, J. and A. Chebira, "Life Beyond Bases: The Advent of Frames (Part I)", *IEEE Signal Processing Magazine*, Vol. 24, No. 4, pp. 86–104, July 2007.

25. Kovacevic, J. and A. Chebira, "Life Beyond Bases: The Advent of Frames (Part II)", *IEEE Signal Processing Magazine*, Vol. 24, No. 5, pp. 115–125, Sept 2007.

26. Welch, L., "Lower bounds on the maximum cross correlation of signals (Corresp.)", *IEEE Trans. Inf. Theory*, Vol. 20, No. 3, pp. 397–399, May 1974.

27. Fornasier, M. and H. Rauhut, "Compressive Sensing", O. Scherzer (Editor), *Hand-*

book of Mathematical Methods in Imaging, pp. 187–228, Springer New York, 2011.

28. Qi, C. and L. Wu, "Optimized Pilot Placement for Sparse Channel Estimation in OFDM Systems", *IEEE Sig. Process. Letters*, Vol. 18, No. 12, pp. 749–752, 2011.

29. Applebaum, L., W. Bajwa, A. Calderbank, J. Haupt and R. Nowak, "Deterministic pilot sequences for sparse channel estimation in OFDM systems", *DSP 2011*, pp. 1–7, July 2011.

30. Qi, C., G. Yue, L. Wu and A. Nallanathan, "Pilot Design for Sparse Channel Estimation in OFDM-Based Cognitive Radio Systems", *IEEE Trans. Veh. Tech.*, Vol. 63, No. 2, pp. 982–987, Feb 2014.

31. Candès, E. and T. Tao, "Decoding by linear programming", *IEEE Trans. Inf. Theory*, Vol. 51, No. 12, pp. 4203–4215, Dec 2005.

32. Herman, M. A. and T. Strohmer, "General Deviants: An Analysis of Perturbations in Compressed Sensing", *IEEE J. Sel. Topics Sig. Process.*, Vol. 4, No. 2, pp. 342–349, April 2010.

33. Cohen, A., W. Dahmen and R. DeVore, "Compressed sensing and best k-term approximation", *AMS Journal*, Vol. 22, No. 1, pp. 211–231, 2009.

34. Mallat, S. G. and Z. Zhang, "Matching pursuits with time-frequency dictionaries", *IEEE Trans. Sig. Process.*, Vol. 41, No. 12, pp. 3397–3415, Dec 1993.

35. Pati, Y., R. Rezaiifar and P. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition", *Asilomar Conf. on Signals, Systems and Computers*, pp. 40–44 vol.1, Nov 1993.

36. Tropp, J. A. and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit", *IEEE Trans. Inf. Theory*, Vol. 53, No. 12, pp. 4655–4666, 2007.

37. Needell, D. and J. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples", *Applied and Computational Harmonic Analysis*, Vol. 26, No. 3, pp. 301 – 321, 2009.

38. Needell, D. and R. Vershynin, "Uniform Uncertainty Principle and Signal Recovery via Regularized Orthogonal Matching Pursuit", *Foundations of Computational Mathematics*, Vol. 9, No. 3, pp. 317–334, 2009.

39. Needell, D. and R. Vershynin, "Signal Recovery From Incomplete and Inaccurate Measurements Via Regularized Orthogonal Matching Pursuit", *IEEE J. Sel. Topics Sig. Process.*, Vol. 4, No. 2, pp. 310–316, April 2010.

40. Donoho, D., Y. Tsaig, I. Drori and J.-L. Starck, "Sparse Solution of Underdetermined Systems of Linear Equations by Stagewise Orthogonal Matching Pursuit", *IEEE Trans. Inf. Theory*, Vol. 58, No. 2, pp. 1094–1121, Feb 2012.

41. Cotter, S. and B. Rao, "Application of tree-based searches to matching pursuit", *IEEE ICASSP*, Vol. 6, pp. 3933–3936 vol.6, 2001.

42. Karabulut, G., L. Moura, D. Panario and A. Yongacoglu, "Integrating flexible tree searches to orthogonal matching pursuit algorithm", *IEE Proc.-Vis. Image, Sig. Process.*, Vol. 153, No. 5, pp. 538–548, Oct 2006.

43. Eldar, Y., P. Kuppinger and H. Bolcskei, "Block-Sparse Signals: Uncertainty Relations and Efficient Recovery", *IEEE Trans. Sig. Process.*, Vol. 58, No. 6, pp. 3042–3054, June 2010.

44. Elhamifar, E. and R. Vidal, "Block-Sparse Recovery via Convex Optimization", *IEEE Trans. Sig. Process.*, Vol. 60, No. 8, pp. 4094–4107, Aug 2012.

45. Yuan, M. and Y. Lin, "Model selection and estimation in regression with grouped variables", *J. R. Statist. Soc. B*, Vol. 68, No. 1, pp. 49–67, 2006.

46. Eldar, Y. and M. Mishali, "Robust Recovery of Signals From a Structured Union of Subspaces", *IEEE Trans. Inf. Theory*, Vol. 55, No. 11, pp. 5302–5316, Nov 2009.

47. Zhang, Z. and B. Rao, "Recovery of block sparse signals using the framework of block sparse Bayesian learning", *IEEE ICASSP*, pp. 3345–3348, March 2012.

48. Li, S., Q. Li, G. Li, X. He and L. Chang, "Iteratively reweighted least squares for block-sparse recovery", *IEEE ICIEA*, pp. 1061–1066, June 2014.

49. Tibshirani, R., "Regression Shrinkage and Selection via the Lasso", *Journal of the Royal Statistical Society. Series B*, Vol. 58, No. 1, pp. 267–288, 1996.

50. Chen, S. S., D. L. Donoho and M. A. Saunders, "Atomic Decomposition by Basis Pursuit", *SIAM Journal on Scientific Computing*, Vol. 20, No. 1, pp. 33–61, 1998.

51. Kim, S. J., K. Koh, M. Lustig, S. Boyd and D. Gorinevsky, "An Interior-Point Method for Large-Scale l1-Regularized Least Squares", *IEEE J. Sel. Topics Sig. Process.*, Vol. 1, No. 4, pp. 606–617, Dec 2007.

52. Lobo, M. S., L. Vandenberghe, S. Boyd and H. Lebret, "Applications of second-order cone programming", *Linear Algebra and its Applications*, Vol. 284, No. 1–3, pp. 193–228, 1998, ILAS Symposium on Fast Algorithms for Control, Signals and Image Processing.

53. Boyd, S. and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

54. CVX Research, I., "CVX: Matlab Software for Disciplined Convex Programming, version 2.1", `http://cvxr.com/cvx`, accessed at April 2016.

55. Grant, M. and S. Boyd, "Graph implementations for nonsmooth convex programs", V. Blondel, S. Boyd and H. Kimura (Editors), *Recent Advances in Learning and Control*, Lecture Notes in Control and Information Sciences, pp. 95–110, Springer-

Verlag Limited, 2008.

56. Becker, S., J. Bobin and E. Candes, "NESTA: A Fast and Accurate First-Order Method for Sparse Recovery", *SIAM Journal on Imaging Sciences*, Vol. 4, No. 1, pp. 1–39, 2011.

57. Becker, S. R., E. J. Candès and M. C. Grant, "Templates for convex cone problems with applications to sparse signal recovery", *Mathematical Programming Computation*, Vol. 3, No. 3, pp. 165–218, 2011.

58. Ji, S., Y. Xue and L. Carin, "Bayesian Compressive Sensing", *IEEE Trans. Sig. Process.*, Vol. 56, No. 6, pp. 2346–2356, June 2008.

59. Ji, S., D. Dunson and L. Carin, "Multitask Compressive Sensing", *IEEE Trans. Sig. Process.*, Vol. 57, No. 1, pp. 92–106, Jan 2009.

60. Emmanuel Candès, T. T., "The Dantzig Selector: Statistical Estimation When p Is Much Larger than n", *The Annals of Statistics*, Vol. 35, No. 6, pp. 2313–2351, 2007.

61. Candès, E. J. and M. A. Davenport, "How well can we estimate a sparse vector?", *Applied and Computational Harmonic Analysis*, Vol. 34, No. 2, pp. 317 – 323, 2013.

62. Baraniuk, R. and P. Steeghs, "Compressive Radar Imaging", *IEEE Radar Conf.*, pp. 128–133, April 2007.

63. Herman, M. A. and T. Strohmer, "High-Resolution Radar via Compressed Sensing", *IEEE Trans. Sig. Process.*, Vol. 57, No. 6, pp. 2275–2284, June 2009.

64. Gurbuz, A. C., J. H. McClellan and V. Cevher, "A compressive beamforming method", *IEEE ICASSP*, pp. 2617–2620, March 2008.

65. Chi, Y., L. L. Scharf, A. Pezeshki and A. R. Calderbank, "Sensitivity to Basis Mismatch in Compressed Sensing", *IEEE Trans. Sig. Process.*, Vol. 59, No. 5, pp.

2182–2195, May 2011.

66. Bajwa, W., J. Haupt, A. Sayeed and R. Nowak, "Compressed Channel Sensing: A New Approach to Estimating Sparse Multipath Channels", *IEEE Proc.*, Vol. 98, No. 6, pp. 1058–1076, June 2010.

67. Haupt, J., W. U. Bajwa, G. Raz and R. Nowak, "Toeplitz Compressed Sensing Matrices With Applications to Sparse Channel Estimation", *IEEE Trans. Inf. Theory*, Vol. 56, No. 11, pp. 5862–5875, Nov 2010.

68. Bajwa, W. U., A. M. Sayeed and R. Nowak, "Learning sparse doubly-selective channels", *46th Annual Allerton Conference on Communication, Control, and Computing*, pp. 575–582, Sept 2008.

69. Golub, G. H. and C. F. Van Loan, *Matrix Computations (3rd Ed.)*, Johns Hopkins University Press, Baltimore, MD, USA, 1996.

70. Ding, J., L. Chen and Y. Gu, "Perturbation Analysis of Orthogonal Matching Pursuit", *IEEE Trans. Sig. Process.*, Vol. 61, No. 2, pp. 398–410, Jan 2013.

71. Slaney, M., A. Kak and L. Larsen, "Limitations of Imaging with First-Order Diffraction Tomography", *IEEE Trans. Microwave Theory Tech.*, Vol. 32, No. 8, pp. 860–874, Aug 1984.

72. Chew, W. and Y. Wang, "Reconstruction of two-dimensional permittivity distribution using the distorted Born iterative method", *IEEE Trans. Medical Imag.*, Vol. 9, No. 2, pp. 218–225, Jun 1990.

73. van den Berg, P. M. and R. E. Kleinman, "A contrast source inversion method", *Inverse Problems*, Vol. 13, pp. 1607 – 1620, 1997.

74. Moose, P. H., "A technique for orthogonal frequency division multiplexing frequency offset correction", *IEEE Trans. Comm.*, Vol. 42, No. 10, pp. 2908–2914,

Oct 1994.

75. Classen, F. and H. Meyr, "Frequency synchronization algorithms for OFDM systems suitable for communication over frequency selective fading channels", *IEEE VTC*, pp. 1655–1659 vol.3, Jun 1994.

76. van de Beek, J. J., M. Sandell and P. O. Borjesson, "ML estimation of time and frequency offset in OFDM systems", *IEEE Trans. Sig. Process.*, Vol. 45, No. 7, pp. 1800–1805, Jul 1997.

77. Liu, H. and U. Tureli, "A high-efficiency carrier estimator for OFDM communications", *IEEE Comm. Letters*, Vol. 2, No. 4, pp. 104–106, April 1998.

78. Gao, F., T. Cui and A. Nallanathan, "Scattered Pilots and Virtual Carriers Based Frequency Offset Tracking for OFDM Systems: Algorithms, Identifiability, and Performance Analysis", *IEEE Trans. Comm.*, Vol. 56, No. 4, pp. 619–629, April 2008.

79. Lee, J.-H., J. C. Han and S.-C. Kim, "Joint carrier frequency synchronization and channel estimation for OFDM systems via the EM algorithm", *IEEE Trans. Veh. Tech.*, Vol. 55, No. 1, pp. 167–172, Jan 2006.

80. Simon, E. P., L. Ros, H. Hijazi and M. Ghogho, "Joint Carrier Frequency Offset and Channel Estimation for OFDM Systems via the EM Algorithm in the Presence of Very High Mobility", *IEEE Trans. Sig. Process.*, Vol. 60, No. 2, pp. 754–765, Feb 2012.

81. Morelli, M. and U. Mengali, "Carrier-frequency estimation for transmissions over selective channels", *IEEE Trans. Comm.*, Vol. 48, No. 9, pp. 1580–1589, Sep 2000.

82. Cui, T. and C. Tellambura, "Joint Frequency Offset and Channel Estimation for OFDM Systems Using Pilot Symbols and Virtual Carriers", *IEEE Trans. Wireless Comm.*, Vol. 6, No. 4, pp. 1193–1202, April 2007.

83. Cho, Y., J. Kim, W. Yang and C. Kang, *MIMO-OFDM Wireless Communications with MATLAB*, Wiley, 2010.

84. Lee, J., H.-L. Lou, D. Toumpakaris and J. Cioffi, "SNR Analysis of OFDM Systems in the Presence of Carrier Frequency Offset for Fading Channels", *IEEE Trans. Wireless Comm.*, Vol. 5, No. 12, pp. 3360–3364, December 2006.

85. Schmidl, T. and D. Cox, "Robust frequency and timing synchronization for OFDM", *IEEE Trans. Comm.*, Vol. 45, No. 12, pp. 1613–1621, Dec 1997.

86. Morelli, M., A. D'Andrea and U. Mengali, "Frequency ambiguity resolution in OFDM systems", *IEEE Comm. Letters*, Vol. 4, No. 4, pp. 134–136, April 2000.

87. Lee, J., H.-L. Lou and D. Toumpakaris, "Joint Maximum Likelihood Estimation of Integer Carrier Frequency Offset and Channel in OFDM Systems", *IEEE ICC*, Vol. 11, pp. 5074–5079, June 2006.

88. Toumpakaris, D., J. Lee and H.-L. Lou, "Estimation of Integer Carrier Frequency Offset in OFDM Systems Based on the Maximum Likelihood Principle", *IEEE Trans. Broadcasting*, Vol. 55, No. 1, pp. 95–108, March 2009.

89. Bai, L. and Q. Yin, "CRB for Carrier Frequency Offset Estimation with Pilot and Virtual Subcarriers", *IEEE Comm. Letters*, Vol. 16, No. 4, pp. 522–525, April 2012.

90. Feng, T., T. Field and S. Haykin, "Stochastic Differential Equation Theory Applied to Wireless Channels", *IEEE Trans. Comm.*, Vol. 55, No. 8, pp. 1478–1483, Aug 2007.

91. Chen, W. and R. Zhang, "Kalman-filter channel estimator for OFDM systems in time and frequency-selective fading environment", *IEEE ICASSP*, Vol. 4, pp. iv–377–iv–380 vol.4, May 2004.

92. Candès, E. J., "The restricted isometry property and its implications for compressed sensing", *Comptes Rendus Mathematique*, Vol. 346, No. 9–10, pp. 589 – 592, 2008.

93. Davenport, M. A. and M. B. Wakin, "Analysis of Orthogonal Matching Pursuit Using the Restricted Isometry Property", *IEEE Trans. Inf. Theory*, Vol. 56, No. 9, pp. 4395–4401, Sept 2010.

94. Stoica, P. and O. Besson, "Training sequence design for frequency offset and frequency-selective channel estimation", *IEEE Trans. Comm.*, Vol. 51, No. 11, pp. 1910–1917, Nov 2003.

95. Kay, S., *Fundamentals of Statistical Signal Processing: Estimation theory*, No. v. 2 in Fundamentals of Statistical Signal Processing, Prentice-Hall PTR, 1993.

96. Stoica, P. and R. L. Moses, *Introduction to spectral analysis*, Vol. 1, Prentice hall Upper Saddle River, 1997.

APPENDIX A: PROOFS IN CHAPTER 4

A.1. Proof of Theorem 4.1

The proof is given for $\sigma = 1$ and the proof for arbitrary σ can be obtained by simple scaling. The signal $\mathbf{x}$ is assumed to be distributed uniformly on a finite set of points $\mathcal{X} \subset \Sigma_s$, where the elements of $\mathcal{X}$ are well separated. From the convexity of the Kullback-Leibler divergence, mutual information can be written as

$$I(\mathbf{x}, \mathbf{y}) \leq \frac{1}{|\mathcal{X}|^2} \sum_{i,l=1}^{|\mathcal{X}|} \mathrm{D}_{\mathrm{KL}}(\mathcal{P}_i, \mathcal{P}_l) \tag{A.1}$$

where $\mathrm{D}_{\mathrm{KL}}(\mathcal{P}_i, \mathcal{P}_l)$ represents the Kullback-Leibler divergence from $\mathcal{P}_i$ to $\mathcal{P}_l$. $\mathcal{P}_i$ denotes the distribution $\mathbf{y}$ conditioned on $\mathbf{x} = \mathbf{x}_i$. Since the additive noise is distributed as $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, $\mathcal{P}_i$ is given as $\mathcal{N}(\mathbf{A}\mathbf{x}_i, \mathbf{I})$. Thus $\mathrm{D}_{\mathrm{KL}}(\mathcal{P}_i, \mathcal{P}_l) = (1/2)\|\mathbf{A}(\mathbf{x}_i - \mathbf{x}_l)\|_2^2$. Mutual information can be written in terms of the entropy as

$$I(\mathbf{x}, \mathbf{y}) = H(\mathbf{x}) - H(\mathbf{x}|\mathbf{y}) \tag{A.2}$$

where $H(\mathbf{x}) = \log |\mathcal{X}|$ since $\mathbf{x}$ is assumed to be uniformly distributed. The conditioned entropy can be bounded using the weakened Fano's inequality

$$H(\mathbf{x}|\mathbf{y}) \leq 1 + P_e \log |\mathcal{X}| \tag{A.3}$$

where $P_e = \mathbb{P}[T(\hat{\mathbf{x}}) \neq \mathbf{x}]$. For any $\mathbf{x} = \mathbf{x}_i \in \mathcal{X}$ and any $\hat{\mathbf{x}}$, $T(\hat{\mathbf{x}} \neq \mathbf{x}_i)$ if and only if there exists an $\mathbf{x}_l \in \mathcal{X}$ with $l \neq i$ satisfying (A.4).

$$\|\hat{\mathbf{x}} - \mathbf{x}_i\|_2 \geq \|\hat{\mathbf{x}} - \mathbf{x}_l\|_2 \geq \|\mathbf{x}_i - \mathbf{x}_l\|_2 - \|\hat{\mathbf{x}} - \mathbf{x}_i\|_2 \tag{A.4}$$

This would imply (A.5).

$$2\|\hat{\mathbf{x}} - \mathbf{x}_i\|_2 \geq \|\mathbf{x}_i - \mathbf{x}_l\|_2 \geq \sqrt{8ne_{\text{wc}}(\mathbf{x}, \mathbf{A})} \tag{A.5}$$

It is assumed that the Euclidean distance between any pair of points $\mathbf{x}_i, \mathbf{x}_l \in \mathcal{X}$ obeys $\|\mathbf{x}_i - \mathbf{x}_l\|_2^2 \geq 8ne_{\text{wc}}(\mathbf{x}, \mathbf{A})$. Using the Markov's inequality, P_e can be bounded as in (A.6).

$$P_e \leq \mathbb{P}[\|\hat{\mathbf{x}} - \mathbf{x}_i\|_2^2 \geq 2ne_{\text{wc}}(\mathbf{x}, \mathbf{A})] \leq \frac{\mathbb{E}\left\{||\hat{\mathbf{x}} - \mathbf{x}_i||_2^2\right\}}{(n/2)e_{\text{wc}}(\mathbf{x}, \mathbf{A})} \leq \frac{ne_{\text{wc}}}{(n/2)e_{\text{wc}}} = \frac{1}{2} \tag{A.6}$$

Combining (A.6) with (A.3), (A.1) can be rewritten as in (A.7).

$$\frac{1}{2}\log|\mathcal{X}| - 1 \leq \frac{1}{2|\mathcal{X}|^2}\sum_{i,l=1}^{|\mathcal{X}|}\|\mathbf{A}(\mathbf{x}_i - \mathbf{x}_l)\|_2^2 \tag{A.7}$$

Suppose that there exists a set $\mathcal{X}$ with cardinality $|\mathcal{X}| = (n/s)^{s/4}$, then (A.7) can be further simplified using the following equations:

$$\frac{s}{4}\log(n/s) - 2 \leq \frac{1}{|\mathcal{X}|^2}\sum_{i,l=1}^{|\mathcal{X}|}\|\mathbf{A}(\mathbf{x}_i - \mathbf{x}_l)\|_2^2 \tag{A.8}$$

$$= \text{Tr}\left\{\mathbf{A}^*\mathbf{A}\left[\frac{1}{|\mathcal{X}|^2}\sum_{i,l=1}^{|\mathcal{X}|}(\mathbf{x}_i - \mathbf{x}_l)(\mathbf{x}_i - \mathbf{x}_l)^*\right]\right\} \tag{A.9}$$

It can be shown that

$$\frac{1}{|\mathcal{X}|^2}\sum_{i,l=1}^{|\mathcal{X}|}(\mathbf{x}_i - \mathbf{x}_l)(\mathbf{x}_i - \mathbf{x}_l)^* = 2(\mathbf{K}_{\mathbf{xx}} - \mathbf{m}_{\mathbf{x}}\mathbf{m}_{\mathbf{x}}^*) \tag{A.10}$$

where $\mathbf{m}_{\mathbf{x}}$ and $\mathbf{K}_{\mathbf{xx}}$ are given as $\mathbf{m}_{\mathbf{x}} = (1/|\mathcal{X}|)\sum_{i=1}^{|\mathcal{X}|}\mathbf{x}_i$ and $\mathbf{K}_{\mathbf{xx}} = (1/|\mathcal{X}|)\sum_{i=1}^{|\mathcal{X}|}\mathbf{x}\mathbf{x}_i^*$ respectively [61]. Thus using (A.10), (A.9) can be bounded as

$$\text{Tr}\left\{\mathbf{A}^*\mathbf{A}(\mathbf{K}_{\mathbf{xx}} - \mathbf{m}_{\mathbf{x}}\mathbf{m}_{\mathbf{x}}^*)\right\} \leq \text{Tr}\{\mathbf{A}^*\mathbf{A}\mathbf{K}_{\mathbf{xx}}\} \tag{A.11}$$

where the inequality is due to $\mathrm{Tr}(\mathbf{A}^*\mathbf{A}\mathbf{m}_\mathbf{x}\mathbf{m}_\mathbf{x}^*) = \|\mathbf{A}\mathbf{m}_\mathbf{x}\|_2^2 \geq 0$. Since $\mathbf{A}^*\mathbf{A}$ and $\mathbf{K}_{\mathbf{xx}}$ are positive semidefinite

$$\mathrm{Tr}(\mathbf{A}^*\mathbf{A}\mathbf{K}_{\mathbf{xx}}) \leq \mathrm{Tr}(\mathbf{A}^*\mathbf{A})\|\mathbf{K}_{\mathbf{xx}}\|_2 = \|\mathbf{A}\|_F^2\|\mathbf{K}_{\mathbf{xx}}\|_2 \tag{A.12}$$

The norm of $\mathbf{K}_{\mathbf{xx}}$ can be bounded as $\|\mathbf{K}_{\mathbf{xx}}\|_2 \leq (1+\beta)/n$

$$\begin{aligned} \frac{32(1+\beta)\|\mathbf{A}\|_F^2}{n} e_{\mathrm{wc}} &\geq \frac{s}{4}\log(n/s) - 2 \\ e_{\mathrm{wc}} &\geq \frac{ns\log(n/s)}{128(1+\beta)\|\mathbf{A}\|_F^2} = \frac{Cs\log(n/s)}{\|\mathbf{A}\|_F^2} \end{aligned} \tag{A.13}$$

where β is a constant that can be arbitrarily small as n gets bigger.

A.2. Proof of Theorem 4.2

Let $\mathbf{e}$ denote the perturbation from the true solution $\mathbf{x}$ of the BP problem (3.11), then the minimizer can be given as $\mathbf{x}^\star = \mathbf{x}+\mathbf{e}$. The perturbation from the true solution can be decomposed into a sum of vectors $\mathbf{e}_{\Lambda_0}, \mathbf{e}_{\Lambda_1}, \ldots$ each of sparsity at most s. Λ_0 denotes the locations of the s largest coefficients of $\mathbf{x}$, Λ_1 to the locations of the s largest coefficients of $\mathbf{e}_{\Lambda_0^c}$, Λ_2 to the locations of the next s largest coefficients of $\mathbf{e}_{\Lambda_0^c}$, and so on [92]. Using $\mathbf{e}_{\Lambda_0}$ and $\mathbf{e}_{\Lambda_1}$, $\mathbf{e}$ can be decomposed into

$$\|\mathbf{x}^\star - \mathbf{x}\|_2 = \|\mathbf{e}\|_2 \leq \|\mathbf{e}_{(\Lambda_0\cup\Lambda_1)}\|_2 + \|\mathbf{e}_{(\Lambda_0\cup\Lambda_1)^c}\|_2 \tag{A.14}$$

as given in [31, 92]. The second term on the right hand side of (A.14) can be written as in (A.15).

$$\|\mathbf{e}_{(\Lambda_0\cup\Lambda_1)^c}\|_2 = \|\sum_{i\geq 2}\mathbf{e}_{\Lambda_i}\|_2 \leq \sum_{i\geq 2}\|\mathbf{e}_{\Lambda_i}\|_2 \leq s^{(-1/2)}\sum_{i\geq 2}\|\mathbf{e}_{\Lambda_{i-1}}\|_1 \leq s^{(-1/2)}\|\mathbf{e}_{\Lambda_0^c}\|_1 \tag{A.15}$$

The error restricted to Λ_0^c cannot be very large because $\|\mathbf{x}+\mathbf{e}\|_1$ should be minumum due to $\mathbf{x}^\star$ being the solution and so (A.16) holds.

$$\|\mathbf{x}\|_1 \geq \|\mathbf{x}+\mathbf{h}\|_1 \geq \|\mathbf{x}_{\Lambda_0}\|_1 - \|\mathbf{e}_{\Lambda_0}\|_1 + \|\mathbf{e}_{\Lambda_0^c}\|_1 - \|\mathbf{x}_{\Lambda_0^c}\|_1 \tag{A.16}$$

Since $\|\mathbf{x}\|_1 - \|\mathbf{x}_{\Lambda_0}\|_1 = \|\mathbf{x}_{\Lambda_0^c}\|_1$, (A.16) can be used to bound $\|\mathbf{e}_{\Lambda_0^c}\|_1$ (A.17).

$$\|\mathbf{e}_{\Lambda_0^c}\|_1 \leq 2\|\mathbf{x}_{\Lambda_0^c}\|_1 + \|\mathbf{e}_{\Lambda_0}\|_1 \tag{A.17}$$

Using (A.17) in (A.15) yields the following equations:

$$\begin{aligned}
\|\mathbf{e}_{(\Lambda_0\cup\Lambda_1)^c}\|_2 &\leq s^{(-1/2)}\|\mathbf{e}_{\Lambda_0^c}\|_1 \leq s^{(-1/2)}(2\|\mathbf{x}_{\Lambda_0^c}\|_1 + \|\mathbf{e}_{\Lambda_0}\|_1) \\
&\leq s^{(-1/2)}(2\|\mathbf{x}_{\Lambda_0^c}\|_1 + s^{(1/2)}\|\mathbf{e}_{\Lambda_0}\|_2) \qquad \text{(A.18)} \\
&\leq 2s^{(-1/2)}\|\mathbf{x}-\mathbf{x}^{(s)}\|_1 + \|\mathbf{e}_{\Lambda_0}\|_2 \qquad \text{(A.19)}
\end{aligned}$$

The second step is to bound the first term on the right hand side of (A.14). In order to accomplish this, $\|\mathbf{A}\mathbf{e}_{\Lambda_0\cup\Lambda_1}\|_2^2$ is written as

$$\begin{aligned}
\|\mathbf{A}\mathbf{e}_{\Lambda_0\cup\Lambda_1}\|_2^2 &= \langle \mathbf{A}\mathbf{e}_{\Lambda_0\cup\Lambda_1}, \mathbf{A}\mathbf{e}_{\Lambda_0\cup\Lambda_1}\rangle \\
&= \langle \mathbf{A}\mathbf{e}_{\Lambda_0\cup\Lambda_1}, \mathbf{A}\mathbf{e}\rangle - \langle \mathbf{A}\mathbf{e}_{\Lambda_0\cup\Lambda_1}, \sum_{i\geq 2}\mathbf{A}\mathbf{e}_{\Lambda_i}\rangle \qquad \text{(A.20)}
\end{aligned}$$

where $\mathbf{A}\mathbf{e}_{\Lambda_0\cup\Lambda_1} = \mathbf{A}\mathbf{e} - \sum_{i\geq 2}\mathbf{A}\mathbf{e}_{\Lambda_i}$. The first term on the right hand side of (A.20) can bounded as

$$|\langle \mathbf{A}\mathbf{e}_{\Lambda_0\cup\Lambda_1}, \mathbf{A}\mathbf{e}\rangle| \leq \|\mathbf{A}\mathbf{e}_{\Lambda_0\cup\Lambda_1}\|_2\|\mathbf{A}\mathbf{e}\|_2 \leq 2\varepsilon\sqrt{1+\delta_{2s}}\|\mathbf{e}_{\Lambda_0\cup\Lambda_1}\|_2 \tag{A.21}$$

where the Cauchy-Schwarz inequality is applied first and then the RIP is used for $\|\mathbf{A}\mathbf{e}_{\Lambda_0\cup\Lambda_1}\|_2$. The term $\|\mathbf{A}\mathbf{e}\|_2$ is bounded using

$$\|\mathbf{A}\mathbf{e}\|_2 = \|\mathbf{A}(\mathbf{x}^\star - \mathbf{x})\|_2 \leq \|\mathbf{A}\mathbf{x}^\star - \mathbf{y}\|_2 + \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \leq 2\varepsilon \tag{A.22}$$

where both $\mathbf{x}^\star$ and $\mathbf{x}$ satisfy the constraint, $\|\mathbf{A}\hat{\mathbf{x}}-\mathbf{y}\|_2 \leq \varepsilon$, in order to be feasible. The second term on the right hand side of (A.20) can be bounded using the Cauchy-Schwarz inequality and the parallelogram identity combined with the RIP (A.23).

$$\begin{aligned}\langle \mathbf{A}\mathbf{e}_{\Lambda_0\cup\Lambda_1}, \sum_{i\geq 2} \mathbf{A}\mathbf{e}_{\Lambda_i}\rangle &\leq |\langle \mathbf{A}\mathbf{e}_{\Lambda_0}, \sum_{i\geq 2} \mathbf{A}\mathbf{e}_{\Lambda_i}\rangle| + |\langle \mathbf{A}\mathbf{e}_{\Lambda_1}, \sum_{i\geq 2} \mathbf{A}\mathbf{e}_{\Lambda_i}\rangle| \\ &\leq \delta_{2s}\|\mathbf{e}_{\Lambda_0}\|_2 \sum_{i\geq 2} \|\mathbf{e}_{\Lambda_i}\|_2 + \delta_{2s}\|\mathbf{e}_{\Lambda_1}\|_2 \sum_{i\geq 2} \|\mathbf{e}_{\Lambda_i}\|_2 \\ &\leq \delta_{2s}(\|\mathbf{e}_{\Lambda_0}\|_2 + \|\mathbf{e}_{\Lambda_1}\|_2) \sum_{i\geq 2} \|\mathbf{e}_{\Lambda_i}\|_2 \end{aligned} \tag{A.23}$$

Using $\|\mathbf{e}_{\Lambda_0}\|_2 + \|\mathbf{e}_{\Lambda_1}\|_2 \leq \sqrt{2}\|\mathbf{e}_{\Lambda_0\cup\Lambda_1}\|_2$ and $\|\sum_{i\geq 2} \mathbf{e}_{\Lambda_i}\|_2 \leq s^{-1/2}\|\mathbf{e}_{\Lambda_0^c}\|_1$, the bound (A.23) can be further simplified as in (A.24).

$$\begin{aligned}\langle \mathbf{A}\mathbf{e}_{\Lambda_0\cup\Lambda_1}, \sum_{i\geq 2} \mathbf{A}\mathbf{e}_{\Lambda_i}\rangle &\leq \delta_{2s}(\sqrt{2}\|\mathbf{e}_{\Lambda_0\cup\Lambda_1}\|_2) \sum_{i\geq 2} \|\mathbf{e}_{\Lambda_i}\|_2 \\ &\leq \sqrt{2}\delta_{2s}\|\mathbf{e}_{\Lambda_0\cup\Lambda_1}\|_2(s^{-1/2}\|\mathbf{e}_{\Lambda_0^c}\|_1) \end{aligned} \tag{A.24}$$

Combining (A.21) and (A.24) on the lower RIP bound for the error restricted to $\Lambda_0\cup\Lambda_1$ gives (A.25).

$$\begin{aligned}(1-\delta_{2s})\|\mathbf{e}_{\Lambda_0\cup\Lambda_1}\|_2^2 &\leq \|\mathbf{A}\mathbf{e}_{\Lambda_0\cup\Lambda_1}\|_2^2 \\ &\leq 2\varepsilon\sqrt{1+\delta_{2s}}\|\mathbf{e}_{\Lambda_0\cup\Lambda_1}\|_2 + s^{-1/2}\sqrt{2}\delta_{2s}\|\mathbf{e}_{\Lambda_0\cup\Lambda_1}\|_2\|\mathbf{e}_{\Lambda_0^c}\|_1 \end{aligned} \tag{A.25}$$

Dividing both sides of (A.25) by $\|\mathbf{e}_{\Lambda_0\cup\Lambda_1}\|_2$ and $(1-\delta_{2s})$ results in:

$$\begin{aligned}\|\mathbf{e}_{\Lambda_0\cup\Lambda_1}\|_2 &\leq \frac{2\sqrt{1+\delta_{2s}}}{1-\delta_{2s}}\varepsilon + \frac{\sqrt{2}\delta_{2s}}{1-\delta_{2s}}s^{-1/2}\|\mathbf{e}_{\Lambda_0^c}\|_1 \\ &\leq \alpha\varepsilon + \rho s^{-1/2}\|\mathbf{e}_{\Lambda_0}\|_1 + 2\rho s^{-1/2}\|\mathbf{x}-\mathbf{x}^{(s)}\|_1 \\ &\leq \alpha\varepsilon + \rho\|\mathbf{e}_{\Lambda_0\cup\Lambda_1}\|_2 + 2\rho s^{-1/2}\|\mathbf{x}-\mathbf{x}^{(s)}\|_1 \\ &\leq (1-\rho)^{-1}(\alpha\varepsilon + 2\rho s^{-1/2}\|\mathbf{x}-\mathbf{x}^{(s)}\|_1) \end{aligned} \tag{A.26}$$

where $\alpha = (2\sqrt{1+\delta_{2s}})/(1-\delta_{2s})$, and $\rho = (\sqrt{2}\delta_{2s})/(1-\delta_{2s})$. Finally the error (A.14) can be bounded as:

$$\begin{aligned}\|\mathbf{e}\|_2 \leq \|\mathbf{e}_{(\Lambda_0\cup\Lambda_1)}\|_2 + \|\mathbf{e}_{(\Lambda_0\cup\Lambda_1)^c}\|_2 &\leq 2s^{-1/2}\|\mathbf{x}-\mathbf{x}^{(s)}\|_1 + 2\|\mathbf{e}_{\Lambda_0\cup\Lambda_1}\|_2 \\ &\leq 2\frac{(1+\rho)}{(1-\rho)}s^{-1/2}\|\mathbf{x}-\mathbf{x}^{(s)}\|_1 + 2\frac{\alpha}{(1-\rho)}\varepsilon \\ &\leq \frac{C_0}{\sqrt{s}}\|\mathbf{x}-\mathbf{x}^{(s)}\|_1 + C_1\varepsilon \end{aligned} \tag{A.27}$$

where $C_0 = 2(1+\rho)(1-\rho)^{-1}$ and $C_1 = 2\alpha(1-\rho)^{-1}$ as in [92].

A.3. Proof of Theorem 4.3

Assuming that the support is exactly recovered

$$\begin{aligned}\|\mathbf{x}^\star - \mathbf{x}\|_2 \leq \|\widetilde{\mathbf{A}}_S^\dagger\tilde{\mathbf{y}} - \mathbf{x}\|_2 &= \left\|\widetilde{\mathbf{A}}_S^\dagger\left[\widetilde{\mathbf{A}}(\mathbf{x}^{(s)}+\mathbf{x}^{(t)}) - \mathbf{Z}\mathbf{x} + \mathbf{z}\right] - \mathbf{x}\right\|_2 \\ &= \|\widetilde{\mathbf{A}}_S^\dagger\|_2\|\mathbf{r}\|_2 \end{aligned} \tag{A.28}$$

where $\mathbf{r} = \widetilde{\mathbf{A}}\mathbf{x}^{(t)} - \mathbf{Z}\mathbf{x} + \mathbf{z}$. $\mathbf{r}$ can be bounded using the following inequalities:

$$\begin{aligned}\|\mathbf{r}\|_2 &\leq \|\widetilde{\mathbf{A}}\mathbf{x}^{(t)}\|_2 + \|\mathbf{Z}\mathbf{x}\|_2 + \|\mathbf{z}\|_2 \\ &\leq \|\widetilde{\mathbf{A}}\mathbf{x}^{(t)}\|_2 + \|\mathbf{Z}\mathbf{x}\|_2 + \varepsilon_{\mathbf{y}}\|\mathbf{A}\mathbf{x}\|_2 \\ &\leq (\|\widetilde{\mathbf{A}}\mathbf{x}^{(t)}\|_2 + \|\mathbf{Z}\mathbf{x}\|_2)(1+\varepsilon_{\mathbf{y}}) + \varepsilon_{\mathbf{y}}\|\widetilde{\mathbf{A}}\mathbf{x}^{(s)}\|_2 \end{aligned} \tag{A.29}$$

$\|\widetilde{\mathbf{A}}\mathbf{x}^{(t)}\|_2$ and $\|\mathbf{Z}\mathbf{x}\|_2$ can be written as

$$\|\widetilde{\mathbf{A}}\mathbf{x}^{(t)}\|_2 \leq \sqrt{1+\tilde{\delta}_s}\left(\|\mathbf{x}^{(t)}\|_2 + \frac{\|\mathbf{x}^{(t)}\|_1}{\sqrt{s}}\right) = \sqrt{1+\tilde{\delta}_s}\,(\beta+\gamma)\|\mathbf{x}^{(s)}\|_2 \tag{A.30}$$

and

$$\|\mathbf{Z}\mathbf{x}\|_2 \leq \|\mathbf{Z}\mathbf{x}^{(t)}\|_2 + \|\mathbf{Z}\mathbf{x}^{(s)}\|_2 \leq \|\mathbf{Z}\|_2^{(s)} \left(\|\mathbf{x}^{(s)}\|_2 + \|\mathbf{x}^{(t)}\|_2 + \frac{\|\mathbf{x}^{(t)}\|_1}{\sqrt{s}} \right)$$
$$\leq \frac{\varepsilon_{\mathbf{A}}}{1-\varepsilon_{\mathbf{A}}} \sqrt{1+\tilde{\delta}_s}(1+\beta+\gamma)\|\mathbf{x}^{(s)}\|_2 \tag{A.31}$$

respectively [37].

$$\|\mathbf{r}\|_2 \leq \frac{\sqrt{1+\tilde{\delta}_s}}{1-\varepsilon_{\mathbf{A}}} \left[\varepsilon_{\mathbf{A}} + \varepsilon_{\mathbf{y}} + (1+\varepsilon_{\mathbf{y}})(\beta+\gamma)\right] \|\mathbf{x}^{(s)}\|_2 \tag{A.32}$$
$$\leq \frac{1.23}{1-\varepsilon_{\mathbf{A}}} \left[\varepsilon_{\mathbf{A}} + \varepsilon_{\mathbf{y}} + (1+\varepsilon_{\mathbf{y}})(\beta+\gamma)\right] \|\mathbf{x}^{(s)}\|_2 = \varepsilon_h \tag{A.33}$$

Finally using that $\tilde{\delta}_s \leq 1/(\sqrt{s}+1) < 0.5$ [93], (A.32) becomes (A.33).

A.4. Proof of Theorem 4.4

For the perturbed sensing matrix, the RIC is modified as $\tilde{\delta}_s$ and the relation of the RIC $\hat{\delta}_s$ for $\widetilde{\mathbf{A}}$ to the RIC δ_s for $\mathbf{A}$ is given in (2.14). Let $\tilde{\mathbf{e}}$ denote the perturbation from the true solution $\mathbf{x}$ of the BP problem (4.18) induced by $\mathbf{Z}$ and $\mathbf{z}$. Then the minimizer of the BP problem can be given as $\mathbf{x}^\star = \mathbf{x} + \tilde{\mathbf{e}}$. Using the same steps given in the proof of Theorem 4.2, the perturbation from the true solution can be decomposed into a sum of vectors as in

$$\|\tilde{\mathbf{e}}\|_2 \leq \left\|\tilde{\mathbf{e}}_{(\Lambda_0 \cup \Lambda_1)}\right\|_2 + \left\|\tilde{\mathbf{e}}_{(\Lambda_0 \cup \Lambda_1)^c}\right\|_2 \leq \frac{2(1+\tilde{\rho})}{(1-\tilde{\rho})} \frac{\|\mathbf{x}-\mathbf{x}^{(s)}\|_1}{\sqrt{s}} + \frac{2\tilde{\alpha}}{(1-\tilde{\rho})}\varepsilon_{\text{tot}} \tag{A.34}$$
$$\leq \widetilde{C}_0 \frac{\|\mathbf{x}-\mathbf{x}^{(s)}\|_1}{\sqrt{s}} + \widetilde{C}_1 \varepsilon_{\text{tot}} \tag{A.35}$$

where $\tilde{\alpha} = (2\sqrt{1+\tilde{\delta}_{2s}})/(1-\tilde{\delta}_{2s})$ and $\tilde{\rho} = \sqrt{2}\tilde{\delta}_{2s}/(1-\tilde{\delta}_{2s})$ are modified versions of α and ρ from (A.27) [32]. Using (2.14) and substituting $2s$ in place of s allows $\widetilde{C}_0$ and

$\widetilde{C}_1$ to be written in terms of the original RIC δ_{2s} as in (A.36) and (A.37) respectively.

$$\widetilde{C}_0 = \frac{2(1+\tilde{\rho})}{(1-\tilde{\rho})} = \frac{2\left(1+(\sqrt{2}-1)\left[(1+\delta_{2s})\left(1+\varepsilon_{\mathbf{A}}^{(2s)}\right)^2-1\right]\right)}{1-(\sqrt{2}+1)\left[(1+\delta_{2s})\left(1+\varepsilon_{\mathbf{A}}^{(2s)}\right)^2-1\right]} \tag{A.36}$$

$$\widetilde{C}_1 = \frac{2\tilde{\alpha}}{(1-\tilde{\rho})} = \frac{4\sqrt{1+\delta_{2s}}\left(1+\varepsilon_{\mathbf{A}}^{(2s)}\right)}{1-(\sqrt{2}+1)\left[(1+\delta_{2s})(1+\varepsilon_{\mathbf{A}}^{(2s)})^2-1\right]} \tag{A.37}$$

Finally the solution of the completely perturbed problem (4.18) can be bounded as in (A.38) [32].

$$\|\mathbf{x}^\star - \mathbf{x}\|_2 \leq \widetilde{C}_0 \frac{\|\mathbf{x}-\mathbf{x}^{(s)}\|_1}{\sqrt{s}} + \widetilde{C}_1\, \varepsilon_{\text{tot}} \tag{A.38}$$

APPENDIX B: CRAMER RAO BOUND

In this appendix the closed form expression for the Cramer-Rao bound (CRB) is derived for the data-aided estimation of frequency offset and channel impulse response. The received signal sampled at the symbol rate through a frequency-selective channel can be written in vector form as

$$\tilde{\mathbf{y}} = \mathbf{\Lambda}(\omega_0)\mathbf{A}\mathbf{h} + \mathbf{z} \tag{B.1}$$

where $\tilde{\mathbf{y}} = [y[0] \ \dots \ y[N-1]]^T$, $\mathbf{z} = [z[0] \ \dots \ z[N-1]]^T$

$$\mathbf{\Lambda}(\omega_0) = \text{diag}\left(1, e^{j\omega_0}, \dots, e^{j(N-1)\omega_0}\right) \tag{B.2}$$

$$\mathbf{A}[i,l] = \mathbf{x}[i-l], \quad 1 \le i \le N, 1 \le l \le L \tag{B.3}$$

The angular frequency offset is denoted as $\omega_0 = 2\pi\epsilon$, and the channel impulse response of length L is shown as $\mathbf{h} = [h[0] \ \dots \ h[L-1]]^T$. The noise vector, $\mathbf{z}$, is assumed to be zero mean circularly symmetric complex valued Gaussian with covariance matrix (B.4).

$$\mathbf{K}_{\mathbf{zz}} = \mathbb{E}\{\mathbf{z}\mathbf{z}^H\} = \sigma_z^2\mathbf{I} \tag{B.4}$$

The parameter vector to be estimated is given as

$$\boldsymbol{\eta} = \begin{bmatrix} \omega_0 \\ \mathbf{h}_{\text{Re}} \\ \mathbf{h}_{\text{Im}} \end{bmatrix} \tag{B.5}$$

where $\mathbf{h}_{\text{Re}}$ and $\mathbf{h}_{\text{Im}}$ stand for the real and imaginary part of $\mathbf{h} = \mathbf{h}_{\text{Re}} + j\mathbf{h}_{\text{Im}}$ respectively [94]. The received signal $\tilde{\mathbf{y}}$ (B.1) is also complex valued circularly symmetric Gaussian vector with mean vector $\mathbf{m}_{\tilde{\mathbf{y}}} = \mathbf{\Lambda}(\omega_0)\mathbf{A}\mathbf{h}$ and covariance matrix $\mathbf{K}_{\tilde{\mathbf{y}}\tilde{\mathbf{y}}} = \sigma_z^2\mathbf{I}$. The Fisher information matrix (FIM) for the estimation of $[\boldsymbol{\eta}^T \ \sigma_z^2]^T$ is block diagonal and

so the estimation of σ_z^2 is decoupled from that of $\boldsymbol{\eta}$ [95,96]. The FIM for $\boldsymbol{\eta}$ is in (B.6).

$$\frac{2}{\sigma_z^2}\Re\left\{\frac{\partial \mathbf{m}_{\tilde{\mathbf{y}}}^H}{\partial \boldsymbol{\eta}}\frac{\partial \mathbf{m}_{\tilde{\mathbf{y}}}}{\partial \boldsymbol{\eta}^T}\right\} \tag{B.6}$$

The partial derivatives with respect to parameters (B.5) can be written as:

$$\begin{aligned}
\frac{\partial \mathbf{m}_{\tilde{\mathbf{y}}}}{\partial \omega_0} &= \frac{\partial}{\partial \omega_0}\boldsymbol{\Lambda}\mathbf{A}\mathbf{h} = j\mathbf{D}\boldsymbol{\Lambda}\mathbf{A}\mathbf{h} \\
\frac{\partial \mathbf{m}_{\tilde{\mathbf{y}}}^*}{\partial \mathbf{h}_{\mathrm{Re}}} &= \frac{\partial}{\partial \mathbf{h}_{\mathrm{Re}}}(\boldsymbol{\Lambda}\mathbf{A}\mathbf{h})^* = \mathbf{A}^*\boldsymbol{\Lambda}^* \\
\frac{\partial \mathbf{m}_{\tilde{\mathbf{y}}}^*}{\partial \mathbf{h}_{\mathrm{Im}}} &= \frac{\partial}{\partial \mathbf{h}_{\mathrm{Im}}}(\boldsymbol{\Lambda}\mathbf{A}\mathbf{h})^* = -j\mathbf{A}^*\boldsymbol{\Lambda}^*
\end{aligned} \tag{B.7}$$

where $\mathbf{D} = \mathrm{diag}(0, \ldots, N-1)$. When (B.7) is put into (B.6), the FIM can be expressed as in (B.8).

$$\frac{2}{\sigma_z^2}\begin{bmatrix}
\mathbf{h}^*\mathbf{A}^*\mathbf{D}^2\mathbf{A}\mathbf{h} & \Im\{\mathbf{h}^*\mathbf{A}^*\mathbf{D}\mathbf{A}\} & \Re\{\mathbf{h}^*\mathbf{A}^*\mathbf{D}\mathbf{A}\} \\
-\Im\{\mathbf{A}^*\mathbf{D}\mathbf{A}\mathbf{h}\} & \Re\{\mathbf{A}^*\mathbf{A}\} & -\Im\{\mathbf{A}^*\mathbf{A}\} \\
\Re\{\mathbf{A}^*\mathbf{D}\mathbf{A}\mathbf{h}\} & \Im\{\mathbf{A}^*\mathbf{A}\} & \Re\{\mathbf{A}^*\mathbf{A}\}
\end{bmatrix} \tag{B.8}$$

The inverse of the FIM (B.8) yields the Cramer-Rao bound (CRB). Using a lemma on the inverse of partitioned matrices, the CRB can be obtained as

$$\mathbf{CRB} = \frac{\sigma_z^2}{2}\begin{bmatrix}
\mathbf{B}^{-1} & \mathbf{B}^{-1}\Im\left\{\mathbf{b}^T\right\} & -\mathbf{B}^{-1}\Re\{\mathbf{b}^T\} \\
\mathbf{B}^{-1}\Im\{\mathbf{b}\} & \Re\{\mathbf{S}\} + \mathbf{B}^{-1}\Im\{\mathbf{b}\}\Im\{\mathbf{b}^T\} & -\Im\{\mathbf{S}\} - \mathbf{B}^{-1}\Im\{\mathbf{b}\}\Re\{\mathbf{b}^T\} \\
-\mathbf{B}^{-1}\Re\{\mathbf{b}\} & \Im\{\mathbf{S}\} - \mathbf{B}^{-1}\Re\{\mathbf{b}\}\Im\{\mathbf{b}^T\} & \Re\{\mathbf{S}\} + \mathbf{B}^{-1}\Re\{\mathbf{b}\}\Re\{\mathbf{b}^T\}
\end{bmatrix} \tag{B.9}$$

where $\mathbf{b}$, $\mathbf{S}$ and $\mathbf{B}$ are given in (B.10).

$$\begin{aligned}
\mathbf{b} &= (\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*\mathbf{D}\mathbf{A}\mathbf{h} \\
\mathbf{S} &= (\mathbf{A}^*\mathbf{A})^{-1} \\
\mathbf{B} &= \mathbf{h}^*\mathbf{A}^*\mathbf{D}(\mathbf{I} - \mathbf{P}_{\mathbf{A}})\mathbf{D}\mathbf{A}\mathbf{h}
\end{aligned} \tag{B.10}$$

$\mathbf{P_A}$ is the projection matrix for $\mathbf{A}$, $\mathbf{P_A} = \mathbf{A}(\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*$. The CRB for ω_0 [81,89,94] is given as in (B.11).

$$\mathbf{CRB}(\omega_0) = \frac{\sigma_z^2}{2}\mathbf{B}^{-1} = \frac{\sigma_z^2}{2}\left[\mathbf{h}^*\mathbf{A}^*\mathbf{D}(\mathbf{I}-\mathbf{P_A})\mathbf{DAh}\right]^{-1} \tag{B.11}$$

(B.9) shows that the CRB depends on $\mathbf{h}$ but not on ω_0. The CRB estimation errors for the real and imaginary parts of $\mathbf{h}$ are correlated to one another. This implies that the channel estimation errors are not circular. The CRB of $\mathbf{h}$ for any unbiased estimate $\hat{\mathbf{h}}$ is given in (B.12).

$$\mathbb{E}\left\{\left(\hat{\mathbf{h}}-\mathbf{h}\right)\left(\hat{\mathbf{h}}-\mathbf{h}\right)^*\right\} \geq \mathbf{CRB}(\mathbf{h}) = \frac{\sigma_z^2}{2}\left(2\mathbf{S}+\mathbf{B}^{-1}\mathbf{b}\mathbf{b}^*\right) \tag{B.12}$$