Recovery guarantee and reconstruction algorithms for 1-bit compressive sensing

Amin Movahed

Communication Systems and Information Theory
Department of Signals and Systems
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden

EX030/2012
Recovery guarantee and reconstruction algorithms for 1-bit compressive sensing

by

Amin Movahed
Recovery guarantee and reconstruction algorithms for 1-bit compressive sensing

Amin Movahed

This thesis has been prepared using \LaTeX.

Copyright © Amin Movahed, 2012.
All rights reserved.

EX030/2012.

Department of Signals and Systems
Division of Communication systems and Information Theory
Chalmers University of Technology
SE-412 96 Gothenburg, Sweden
email: Movahed@student.chalmers.se

Cover:
The big picture of 1-bit compressive sensing. Binary measurements vector $\mathbf{b}$ is obtained by 1-bit quantization of measurements vector $\mathbf{y}$ which is generated by compressive sensing (measurement matrix $\Phi \in \mathbb{R}^{M \times N}$) over signal vector $\mathbf{x}$. 
To my dear and loving family
Recovery guarantee and reconstruction algorithms for 1-bit compressive sensing
Amin Movahed
Department of Signals and Systems
Division of Communication Systems and Information Theory
Chalmers University of Technology

Abstract

Compressive sensing is an emerging method for signal acquisition in which the number of samples ensuring exact reconstruction of the signal to be acquired is far less than the one in the conventional Nyquist sampling approach. In compressive sensing, the signal is acquired by means of few linear non-adaptive measurements, and then reconstructed by finding the sparsest solution via an $\ell_1$-minimization.

In the classic compressive sensing setup, each measurement outcome is described by a real value. In practice, for further processing and storage purposes, often the real-valued measurements need to be converted to finite-precision numbers. 1-bit compressive sensing refers to the extreme case where the quantizer is a simple sign comparator and each measurement is represented using one bit only, i.e., +1 or −1.

Several algorithms have been introduced in the literature for solving efficiently the reconstruction problem in the 1-bit compressive sensing setting, e.g., renormalized fixed point iteration (RFPI) and binary iterative hard thresholding (BIHT). However, these algorithms can not reconstruct the signal accurately when there is noise, i.e., bit flips, in the binary measurements. Adaptive outlier pursuit (AOP) is an algorithm which reconstructs the signal robustly against bit flips in the binary measurements. AOP requires the sparsity level of the signal to be reconstructed as an input. In many practical cases, however, the sparsity level of the signal is unknown and time variant.

In this thesis, we address reconstruction problem in 1-bit compressive sensing. We introduce a new algorithm for 1-bit compressive sensing which reconstructs the signal robustly from the noisy binary measurements. This new reconstruction algorithm does not require the sparsity level of the signal as an input. Therefore, our algorithm can be applied in the most practical scenarios in which the sparsity level of the signal is unknown.

Keywords: 1-bit quantization, compressive sensing, iterative algorithms, $\ell_1$-minimization.
I am heartily thankful to my supervisor, Giuseppe Durisi, whose encouragement, guidance and support from the initial to the final level enabled me to develop an understanding of the subject. Moreover, I would like to show my gratitude to my dear friend, Ashkan Panahi, from whom I learned many valuable things during this work.

Amin Movahed
Contents

Abstract

Acknowledgments

Contents

List of figures

List of tables

Abbreviations, Acronyms and Notations

1 Introduction
  1.1 Introduction to 1-bit compressive sensing
  1.2 The key contribution of the thesis
  1.3 The structure of the thesis

2 Compressive sensing
  2.1 Recovery via \( \ell_1 \)-minimization
  2.1.1 Sparse Recovery
  2.1.2 Null space property and restricted isometry property
  2.2 Random matrices and RIP
  2.2.1 Gaussian random matrices
  2.2.2 Sub-Gaussian random matrices
  2.3 Iterative hard thresholding

3 1-bit compressive sensing
  3.1 1-bit compressive sensing problem
  3.2 1-bit compressive sensing reconstruction
### 3.3 Iterative reconstruction algorithms for 1-bit compressive sensing

- 3.3.1 BIHT
- 3.3.2 BIHT-$\ell_2$
- 3.3.3 RFPI

### 4 1-bit compressive sensing in the presence of noise

- 4.1 Noise modelling in 1-bit compressive sensing
  - 4.1.1 Measurement noise
  - 4.1.2 Channel noise
  - 4.1.3 A combined model for binary noise
- 4.2 Iterative reconstruction algorithms for 1-bit compressive sensing in the presence of noise
  - 4.2.1 Adaptive outlier pursuit
  - 4.2.2 Noise-adaptive renormalized fixed point iterative

### 5 Simulations and numerical results

- 5.1 Algorithms designed for noiseless 1-bit compressive sensing
- 5.2 Algorithms designed for noisy 1-bit compressive sensing
  - 5.2.1 Signals with fixed sparsity level
  - 5.2.2 Signals with random sparsity level

### 6 Conclusion

- 6.1 Suggestion for future work

### Appendices

- A Proof of Theorems in Chapter
  - A.1 Proof of Theorem
  - A.2 Proof of Theorem
  - A.3 Proof of Theorem
  - A.4 Proof of Theorem

- B A model for the measurement noise in 1-bit compressive sensing

### References
List of Figures

2.1 The intersection point between $\ell_1$-ball and $\Phi x = y$ in 2-dimensional space. ........................................... 6
2.2 The rate of successful signal reconstruction via IHT ........... 11
3.1 1-bit compressive sensing and reconstruction block diagram .. 15
4.1 Binary symmetric channel model ................................. 22
4.2 1-bit compressive sensing in the presence of the noise ......... 23
5.1 The performance of BIHT, BIHT-$\ell_2$ and RFPI when $P_f = 0$ . 32
5.2 The performance of BIHT, BIHT-$\ell_2$ and RFPI when $P_f = 1\%$ 33
5.3 The performance of BIHT, BIHT-$\ell_2$ and RFPI when $P_f = 3\%$ 33
5.4 The performance of AOP-f, AOP-f-$\ell_2$ and NARFPI when $P_f = 3\%$ ................................................................. 34
5.5 The PMF of truncated triangular distribution of $s$ .............. 35
5.6 The performance of RFPI, AOP-f, AOP-f-$\ell_2$ and NARFPI when $P = 3\%$ and $s$ is random ................................. 36
5.7 The estimation error of NARFPI as a function of the number of binary measurements $M$ and of the sparsity level $s$ of $x$ .... 36
5.8 The estimation error of AOP-f as a function of the number of binary measurements $M$ and of the sparsity level $s$ of $x$ ....... 37
5.9 The estimation error of AOP-f-$\ell_2$ as a function of the number of binary measurements $M$ and of the sparsity level $s$ of $x$ ... 37
List of Tables

3.1 Three reconstruction algorithms for 1-bit compressive sensing in the noiseless scenario. ........................................ 20

4.1 Five reconstruction algorithms for 1-bit compressive sensing in the presence of the binary noise. ............................ 30
## Abbreviations and Acronyms

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AOP</td>
<td>adaptive outlier pursuit</td>
</tr>
<tr>
<td>AOP-$\ell_2$</td>
<td>adaptive outlier pursuit using $\ell_2$-norm minimization</td>
</tr>
<tr>
<td>AOP-f</td>
<td>adaptive outlier pursuit with sign flips</td>
</tr>
<tr>
<td>AOP-f-$\ell_2$</td>
<td>adaptive outlier pursuit with sign flips using $\ell_2$-norm minimization</td>
</tr>
<tr>
<td>BIHT</td>
<td>binary iterative hard thresholding</td>
</tr>
<tr>
<td>BIHT-$\ell_2$</td>
<td>binary iterative hard thresholding using $\ell_2$-norm minimization</td>
</tr>
<tr>
<td>BSC</td>
<td>binary symmetric channel</td>
</tr>
<tr>
<td>IHT</td>
<td>iterative hard thresholding</td>
</tr>
<tr>
<td>NARFPI</td>
<td>noise adaptive renormalized fixed point iteration</td>
</tr>
<tr>
<td>RFPI</td>
<td>renormalized fixed point iteration</td>
</tr>
<tr>
<td>RIP</td>
<td>restricted isometry property</td>
</tr>
</tbody>
</table>
Notations

$\|\cdot\|_0$  \quad \ell_0$-norm (Number of non-zero elements)

$\|\cdot\|_1$  \quad \ell_1$-norm

$\|\cdot\|_2$  \quad \ell_2$-norm (Euclidean norm)

$\|\cdot\|_F$  \quad$\text{Frobenius-norm}$

$\odot$  \quad$\text{element-wise product}$

$(\cdot)^-$  \quad$\text{negative function}$

$\emptyset$  \quad$\text{empty set}$

$|T|$  \quad number of elements in set $T$

$\langle\cdot,\cdot\rangle$  \quad$\text{inner product}$

\  \quad$\text{set minus}$

$\succeq$  \quad$\text{element-wise inequality}$

$[\cdot]_i$  \quad $i$th element of the argument

$b$  \quad binary measurements vector

$\tilde{b}$  \quad noisy binary measurements vector

$\mathbb{C}$  \quad set of complex numbers

$\text{diag}(b)$  \quad a diagonal matrix whose diagonal is vector $b$

$\mathbb{E}(\cdot)$  \quad$\text{expected value of the argument}$

$\text{exp}(\cdot)$  \quad$\text{exponential function}$

$\Phi$  \quad measurement matrix

$\Phi^\dagger$  \quad$\text{conjugate transpose of } \Phi$

$\Phi^T$  \quad$\text{transpose of } \Phi$

$\Phi^{-1}$  \quad$\text{pseudo-inverse of } \Phi$

$\Phi_{[\cdot],T}$  \quad a sub-matrix of $\Phi$ spanned to columns of $\Phi$ in set $T$
Φ_{[V,:]} a sub-matrix of Φ spanned to rows of Φ in set V
φ_i,: the $i$th row of Φ
φ_{i,j} the element of Φ in row $i$ and column $j$
$\mathcal{H}_s(\cdot)$ a non-linear operator that keeps the $s$ largest elements of the argument and set the other elements to zero
ker{Φ} null-space of matrix Φ
$\lambda_{\text{max}}(\Phi)$ maximum eigenvalue of matrix Φ
$\lambda_{\text{min}}(\Phi)$ minimum eigenvalue of matrix Φ
Λ a binary vector containing 0s and 1s
$M$ dimension of the measurements vector
$N$ dimension of the signal vector
$[N]$ set of numbers from 1 to $N$
$N(\mu, \sigma^2)$ normal distribution with mean $\mu$ and variance $\sigma^2$
Ω a binary vector containing $-1$s and 1s
$O(\cdot)$ order of the argument
$P_f$ probability of sign flips
$P(\cdot)$ probability of the argument
$\mathbb{R}$ set of real numbers
$s$ sparsity level of the signal
Supp(x) set of indices of non-zero elements in x
Sub$(c^2)$ sub-Gaussian distribution with constant $c$
SSub$(c^2)$ strictly Sub-Gaussian distribution with constant $c$
$\sigma_{\text{max}}(\Phi)$ maximum singular value of matrix Φ
$\sigma_{\text{min}}(\Phi)$ minimum singular value of matrix Φ
tr(Φ) trace of matrix Φ
\( T^c \) complement of set \( T \)
\( \mathbf{x} \) signal vector
\( \hat{\mathbf{x}} \) estimation of signal vector
\( \mathbf{x}^{[n]} \) vector \( \mathbf{x} \) in \( n \)th iteration
\( \mathbf{x}_T \) a vector containing elements of \( \mathbf{x} \) spanned to the indices in set \( T \)
\( \mathbf{y} \) measurements vector
\( \tilde{\mathbf{y}} \) noisy measurements vector
1.1 Introduction to 1-bit compressive sensing

The Nyquist sampling theorem specifies that to avoid information loss when measuring a signal, one must sample it at least two times faster than the signal bandwidth \cite{1}. In many applications, like digital image and video cameras, the Nyquist rate is so high that the samples need to be compressed before storage or transmission. Furthermore, increasing the sampling rate is very expensive and impractical in applications like medical image scanners, radars and high-speed analog-to-digital converters.

Compressive sensing provides a completely new approach to data acquisition by suggesting that it is possible to improve the traditional Nyquist limits of sampling theory \cite{2}. Compressive sensing predicts that certain signals can be recovered from low rate measurements which are not sufficient according to Nyquist sampling theorem. Compressive sensing is based on the empirical observation that many types of signals or images can be well-approximated by a sparse expansion in terms of a suitable basis \cite{3, 4}. For example, in digital imaging systems, sparse signals can be obtained by applying wavelet transformation over original captured images. In compressive sensing, each measurement is obtained through an inner product between the vector of the sparse signal and the vector containing measuring elements. Then, the sparse signal can be reconstructed from few number of these measurements via an \( \ell_1 \)-minimization. In recent years, compressive sensing has attracted considerable attention in many areas of applied mathematics, computer science, and electrical engineering.

There are two main research topics in compressive sensing. The first
topic is about establishing mathematically rigorous recovery guarantees relating the signal dimension, the sparsity level of the signal and the number of measurements. The task of the second topic is to find algorithms that reconstruct the signals accurately from the measurements obtained by compressive sensing. E. Candès and T. Tao [5] showed that sparse signals can be reconstructed through an $\ell_1$-minimization when the measurement system is satisfying restricted isometry property. There are various reconstruction algorithms for compressive sensing. We single out the paper by J. A. Tropp and S. J. Wright [6] in which various types of algorithms for sparse reconstruction from compressive sensing are explained.

In practice, the measured signal should pass through a quantizer, which is located between the measuring and the recovering layers, provides recognizable data for further storage or transmission [7]. The quantizer affects the signal reconstruction accuracy. Performance analysis of the effects of the quantization and the noise of the measurement on various data reconstruction methods in compressive sensing is an emerging research topic. In this work, we focus on the extreme class of quantizers that have only two quantization levels, i.e., each quantized value is characterized by only one bit. 1-bit compressive sensing is the combination of compressive sensing with a 1-bit quantizer, which was first introduced by P. Boufounos and R. Baraniuk [8].

The main focus of this thesis is on the iterative reconstruction algorithms in 1-bit compressive sensing. As reconstruction solutions for 1-bit compressive sensing, two different iterative algorithms were introduced by P. Boufounos and R. Baraniuk [8] and L. Jacques et al. [9] for signal reconstruction in noiseless 1-bit compressive sensing. However, the reconstruction process becomes more challenging when there is noise in the quantized measurements. An iterative algorithm that is robust against noise in 1-bit compressive sensing has been proposed by M. Yan et al. [10]. This algorithm reconstructs the signal from the noisy binary measurements based on the following a priori information: 1) the sparsity level of the signal and 2) the amount of noise in the binary measurements.

1.2 The key contribution of the thesis

In many practical applications, the sparsity level of the signal is not known and, therefore, the algorithm proposed in [10] can not be fed by an exact input of the signal sparsity level. As the main contribution of this thesis,
we propose an algorithm which works robustly against noise in 1-bit compressive sensing and does not need the sparsity level of the signal to be reconstructed as an input. Hence, this new algorithm can be applied in many practical scenarios in which the sparsity level of the signal is unknown and time-variant. Another contribution of this thesis is to review and put the theoretical results, relating reconstruction guarantees in compressive sensing and 1-bit compressive sensing, in a single framework.

1.3 The structure of the thesis

- **Chapter 2** We start the first part of this chapter by defining the compressive sensing problem and the conditions in which linear measurements system in compressive sensing can guarantee a perfect signal reconstruction. In this thesis we mainly focus on iterative reconstruction algorithms, therefore, in the second part of this chapter, we introduce an iterative reconstruction algorithm for compressive sensing.

- **Chapter 3** We define the 1-bit compressive sensing problem. We also explain the conditions guaranteeing reconstruction in 1-bit compressive sensing. Due to time constraints, we do not provide details of the proof of reconstruction guarantees in 1-bit compressive sensing. We also introduce some iterative reconstruction algorithms for 1-bit compressive sensing, e.g. BIHT, BIHT-$\ell_2$ and RFPI, designed to work in the noiseless scenario.

- **Chapter 4** In the first part, we model the binary noise in 1-bit compressive sensing. In the second part, we discuss reconstruction algorithms designed to perform robustly in the presence of the noise, e.g. AOP, AOP-$\ell_2$, AOP-f and AOP-f-$\ell_2$. In the third part, we propose our contribution which is a new reconstruction algorithm, NARFPI, that is robust against noise and does not need any a priori knowledge of the signal sparsity level.

- **Chapter 5** We evaluate the algorithms through Matlab simulations. In the first part, we compare the performance of the iterative algorithms designed for noiseless case in the noiseless and the noisy scenarios. In the second part, we compare the performance of our new algorithm with the algorithms designed to work in the noisy scenario and in the case that sparsity level of the signal is unknown.
• **Chapter 6**: We summarize the results obtained from previous chapters and we discuss about possible future work to extend the proposed reconstruction algorithm.

• **Appendix A**: The theorems in Chapter 2 are proved in detail.

• **Appendix B**: We present a model for measurement noise to be used in Chapter 4.
Chapter 2

Compressive sensing

2.1 Recovery via $\ell_1$-minimization

In this chapter, we introduce the compressive sensing problem and then we show the conditions in which the recovery for compressive sensing is guaranteed. In addition, we explain an iterative reconstruction algorithm for compressive sensing.

2.1.1 Sparse Recovery

Signal vector $\mathbf{x}$ is called $s$-sparse if no more than $s$ of its elements have non-zero values, i.e., $\|\mathbf{x}\|_0 \leq s$. In compressive sensing, a $s$-sparse signal is measured through few non-adaptive linear measurements. In other words, let $\mathbf{x} \in \mathbb{C}^N$ be an $N$-dimensional vector that is $s$-sparse. $\mathbf{y} \in \mathbb{C}^M$, which is called measurements vector, is obtained by left multiplying $\mathbf{x}$ with the measurement matrix $\Phi \in \mathbb{C}^{M \times N}$ according to

$$\Phi \mathbf{x} = \mathbf{y}. \quad (2.1)$$

We are interested in the case when the dimension of the measurements vector is less than the dimension of the signal, that is, $M \ll N$. Since $\Phi$ is a fat matrix, by solving (2.1) for $\mathbf{x}$ we obtain infinitely many solutions. By imposing the additional requirement that $\mathbf{x}$ is $s$-sparse, the problem changes to searching for the sparsest solution of (2.1). Therefore, we need to solve the $\ell_0$-minimization problem

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{x}\|_0$$

subject to $\Phi \mathbf{x} = \mathbf{y} \quad (2.2)$
where $\hat{x}$ denotes estimation of $x$. Unfortunately, (2.2) is an NP-hard problem [11] and is not easy to solve. One approach to solve (2.2) is to use convex relaxation. The $\ell_1$-minimization problem

$$\hat{x} = \arg \min_x \|x\|_1$$

subject to $\Phi x = y$ (2.3)

is a convex relaxation of (2.2) [12]. In Figure 2.1 the mechanism of solving (2.3) in 2-dimensional space is shown. In words, the intersection point between the smallest possible $\ell_1$-ball and $\Phi x = y$ is the sparsest solution satisfying $\Phi x = y$. Note that the sparse $\hat{x}$ in 2-dimensional space is a 2-dimensional vector with a zero in one of its two elements. (we selected the 2-dimensional space for the sake of simplicity, however, this mechanism holds for every $N$-dimensional space). Therefore, the problem of compressive sensing converts to solving (2.3).

### 2.1.2 Null space property and restricted isometry property

In order to make sure that through solving (2.3) we obtain a unique solution coinciding with the solution of (2.2), the measurement matrix $\Phi$ should fulfill some requirements. First, we define null space property and show that in the
case this property holds, the solution of $\ell_1$-minimization is unique. Then, we define another property called restricted isometry property (RIP) and we show that RIP guarantees null space property.

**Null space property**

**Definition 1.** A matrix $\Phi \in \mathbb{C}^{M \times N}$ satisfies the null space property of order $s$ if for all $S \subset [N]$, $|S| = s$ we have that
\[
\|v_S\|_1 < \|v_{S^c}\|_1 \quad \text{for all } v \in \ker\{\Phi\} \setminus \{0\}
\] (2.4)

where $[N] = \{1, 2, \ldots, N\}$, $\ker\{\Phi\} = \{x \in \mathbb{C}^N, \Phi x = 0\}$, $v_S = (v_j)_{j \in S}$ and $S^c = [N] \setminus S$.

The next theorem shows the importance of null space property in compressive sensing reconstruction via $\ell_1$-minimization.

**Theorem 1** ([12, Theorem 2.3]). Let $\Phi \in \mathbb{C}^{M \times N}$. Every $s$-sparse vector $x \in \mathbb{C}^N$ is the unique solution of $\ell_1$-minimization problem (2.3) with $\Phi x = y$ if and only if $\Phi$ satisfies the null space property of order $s$.

Theorem 1 is proved in Appendix A.1.

**Restricted isometry property**

**Definition 2.** For each $s = 1, 2, \ldots$ and for all $s$-sparse $x \in \mathbb{C}^N$, we define $\delta_s$ to be the smallest scalar such that
\[
(1 - \delta_s) \|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1 + \delta_s) \|x\|_2^2 \quad \text{for all } s\text{-sparse } x.
\] (2.5)

Then, the matrix $\Phi$ is said to satisfy the restricted isometry property (RIP) with restricted isometry constant $\delta_s$.

The following theorem shows in which conditions on restricted isometry constant, null space property holds.

**Theorem 2** ([12, Theorem 2.3]). Let $\Phi \in \mathbb{C}^{M \times N}$ has restricted isometry constant $\delta_{2s} < 1/3$. Then $\Phi$ satisfies the null space property of order $s$.

Theorem 2 is proved in Appendix A.2.

### 2.2 Random matrices and RIP

Up to here, we showed how the signal recovery through $\ell_1$-minimization can be guaranteed by RIP. In this section, we explain which matrices are obeying this property and are suitable to be applied in compressive sensing.
2.2.1 Gaussian random matrices

Definition 3 ([13 Definition 2.1]). A standard real/complex Gaussian $M \times N$ matrix $\Phi$ has i.i.d. real/complex zero-mean Gaussian entries with identical variance $\sigma^2 = 1/M$. The probability density function of a complex Gaussian matrix with i.i.d. zero-mean Gaussian entries with variance $\sigma^2$ is

$$(\pi\sigma^2)^{-MN} \exp \left[-\frac{\text{tr}(\Phi\Phi^\dagger)}{\sigma^2}\right].$$

(2.6)

Theorem 3 ([14]). Let $\Phi \in \mathbb{C}^{M\times N}$ be a random Gaussian matrix with zero-mean and $\sigma^2 = 1/M$. If

$$M \geq \frac{4}{k^2} s \log (N/s)$$

(2.7)

$\Phi$ satisfies RIP for $\delta_{2s} < 1/3$ with probability greater than

$$1 - 2 \exp \left(2s \log (N/2s) - Mk^2/2\right)$$

(2.8)

for every $0 < k < 0.15 - \sqrt{2s/M}$.

Theorem 3 is proved in Appendix A.3. The main result of Theorem 3 is that any random Gaussian matrix, whose number of rows is (2.7), satisfies null space property with high probability and, therefore, $x$ can be reconstructed accurately through solving (2.3).

2.2.2 Sub-Gaussian random matrices

In the following, we introduce sub-Gaussian matrices and we show in which conditions on this class of matrices RIP is satisfied.

Definition 4. A random variable $X$ is called sub-Gaussian if there exists a constant $c > 0$ such that

$$\mathbb{E}(\exp(Xt)) \leq \exp\left(c^2t^2/2\right),$$

(2.9)

where $t \in \mathbb{R}$. We use the notation $X \sim \text{Sub}(c^2)$ to denote that $X$ satisfies (2.9).

Lemma 1 ([15 Lemma 1.2]). If $X \sim \text{Sub}(c^2)$ then $\mathbb{E}(X) = 0$ and $\mathbb{E}(X^2) \leq c^2$.

This lemma shows that if $X \sim \text{Sub}(c^2)$ then $\sigma^2 \leq c^2$ where $\sigma^2$ is the variance of $X$. In the case that $\sigma^2 = c^2$, we define more specific class of distributions.
Definition 5. If \( X \sim \text{Sub}(\sigma^2) \) where \( \sigma^2 = \mathbb{E}(X^2) \) then \( X \) is called **strictly sub-Gaussian** and we use the notation \( X \sim \text{SSub}(\sigma^2) \).

**Theorem 4** ([16, Theorem 7.3]). Let \( \Phi \in \mathbb{C}^{M \times N} \) whose entries \( \phi_{ij} \) are i.i.d. with \( \phi_{ij} \sim \text{SSub}(1/M) \) and \( \delta_s \in (0,1) \). If

\[
M \geq \alpha s \log(\frac{N}{s}),
\]

then \( \Phi \) satisfies RIP of order \( s \) with probability greater than \( 1 - 2e^{-\beta M} \), where \( \alpha \) is arbitrary, \( \beta = \frac{\delta^2}{2\mu} - \log(\frac{42e}{\delta s})/\alpha \) and \( \mu = \frac{2}{1 - \log(2)} \).

Theorem 4 is proved in Appendix A.4. Theorem 4 proves that any matrix whose elements are strictly sub-Gaussian fulfils RIP with restricted isometry constant \( \delta_s \) when the number of rows is \((2.10)\).

### 2.3 Iterative hard thresholding

In the previous sections, we discussed about the characteristic of the measuring matrix which guarantees the reconstruction. In this section and next chapters, we focus on the reconstruction algorithms. Throughout this work, we assume that \( \Phi \) is a real random Gaussian matrix. Gaussian random matrices are practically easy to generate and, therefore, no prior storage is needed. One natural variation of compressive sensing problem is to relax constraint in \((2.2)\) and allow some error tolerance \( \varepsilon \geq 0 \) [6]. Therefore, we obtain

\[
\hat{x} = \arg \min_x \|x\|_0 \quad \text{subject to } \|y - \Phi x\|_2 \leq \varepsilon
\]

The measurement error can be defined in different ways. In \((2.11)\), the predicted measurement error is measured by euclidean norm. In the case that the sparsity level of \( x \), \( s \), is known as a priori knowledge, the problem in \((2.11)\) can be written as follows:

\[
\hat{x} = \arg \min_x \|y - \Phi x\|_2 \quad \text{subject to } \|x\|_0 \leq s.
\]

The minimization \((2.12)\) searches for the best approximation of \( x \) with the given maximum sparsity level \( s \). In [17], an iterative algorithm called **iterative hard thresholding** (IHT) is introduced which solves \((2.12)\) iteratively. IHT includes two steps. The first step consists of a gradient descent to reduce
Algorithm 1 IHT

1. **Inputs:** measurements vector $y$, measurement matrix $\Phi$, sparsity level of the signal $s$, descent step size $\tau$, number of iterations $t$

2. **Initialization:** Initial estimate $x^{[0]} = 0$

3. **Iteration:** For $n = 1, \ldots, t$
   
   (a) **Gradient descent:** $z^{[n]} \leftarrow x^{[n-1]} - \tau \Phi^T (\Phi x^{[n-1]} - y)$
   
   (b) **Projection onto “$\ell_0$-ball”:** $x^{[n]} \leftarrow \mathcal{H}_s(z^{[n]})$

4. **Output:** $\hat{x} = \frac{x^{[n]}}{\|x^{[n]}\|_2}$

$\|y - \Phi x\|_2$. The second step generates a sparse signal model by projecting the output of the step one onto the “$\ell_0$-ball” by selecting the $s$ largest elements in $x$ obtained from the previous step. Therefore, each iteration step is as follows:

$$x^{[n+1]} = \mathcal{H}_s(x^{[n]} + \Phi^T(y - \Phi x^{[n]}))$$ (2.13)

where $x^{[n]}$ denotes $x$ in $n$th iteration, $x^{[0]} = 0$ and $\mathcal{H}_s(\cdot)$ is a non-linear operator that keeps the $s$ largest elements of the argument and set the other elements to zero. It is shown that when $\Phi$ satisfies RIP the solution of (2.13) converges through iterations [18]. The steps of IHT are shown in detail in Algorithm 1.

The euclidean distance between the signal and its estimation, i.e., $\|x - \hat{x}\|_2$ is a measure of reconstruction quality. In Figure 2.2, the rate of successful signal reconstruction (i.e., $\|x - \hat{x}\|_2 < 10^{-5}$) via IHT among 4000 realizations is shown for different number of measurements $M$, where $N = 1000$, $s = 10$, $t = 200$ and $\Phi$ is a Gaussian random matrix. As it is expected, when $M$ tends to $N$ ($M/N \rightarrow 1$), the rate of successful signal reconstruction converges to 100%. That is, when $M = N$ the linear equation system is not undetermined and definitely has a unique solution.
Figure 2.2: The rate of successful signal reconstruction via IHT
3.1 1-bit compressive sensing problem

In Chapter 2 we introduced the classic compressive sensing problem. This problem is based on the fact that measurements vector $y$ has infinite bit precision. However, in practice measurements should be quantized for further storage and transmission. In the case the quantizer has only one bit, it is called 1-bit compressive sensing. That is, by sensing the signal through measuring matrix we obtain measurements vector and its 1-bit quantized version is called binary measurements vector. This binary vector is denoted by $b$ and we have

$$b = \text{sign} (\Phi x).$$

(3.1)

1-bit quantizer is a comparator to zero which is very fast and inexpensive hardware device. Therefore, it has many advantages in comparison to other types of quantizers in hardware implementation. As Boufounos et al. mentions [8], “1-bit quantizers do not suffer from dynamic range issues. If the analog side of the measurement system is properly implemented then the sign of the measurement remains valid even if the quantizer saturates”.

3.2 1-bit compressive sensing reconstruction

Inspired by (2.3), the problem of 1-bit compressive sensing reconstruction can be written as

$$\hat{x} = \arg \min_x \|x\|_1$$

subject to $\text{sign}(\Phi x) = b$. 

(3.2)
That is, by replacing the constraint $\Phi x = y$ in (2.3) with $\text{sign}(\Phi x) = b$, we obtain (3.2). However, the optimization problem in (3.2) is not convex and, therefore, is not easy to solve. In [19], it is suggested that by applying $\|\Phi x\|_1 = M$ as an extra constraint over (3.2), following convex optimization can be obtained

$$\hat{x} = \arg \min_x \|x\|_1$$
subject to $\text{sign}(\Phi x) = b$. 
and $\|\Phi x\|_1 = M$. \hspace{1cm} (3.3)

In fact, $\text{sign}(\Phi x) = b$ and $\|\Phi x\|_1 = M$ can be combined and represented as one linear equation

$$\|\Phi x\|_1 = \sum_{i=1}^{M} |\langle \phi_i, x \rangle| = \sum_{i=1}^{M} [b]_i \langle \phi_i, x \rangle = M$$ \hspace{1cm} (3.4)

where $[\cdot]_i$ denotes the $i$th element of the argument and $\phi_i$ denotes the $i$th row of $\Phi$. Therefore, (3.3) is indeed a convex minimization program [19]. In addition, for the sake of simplicity and without loss of generality, both the original and the reconstructed signals are imposed to be on “$\ell_2$-ball” with unite radius (energy normalization). Hence, constraint $\|x\|_2 = 1$ is added to (3.3) and we have

$$\hat{x} = \arg \min_x \|x\|_1$$
subject to $\text{sign}(\Phi x) = b$
and $\|\Phi x\|_1 = M, \|x\|_2 = 1$. \hspace{1cm} (3.5)

**Theorem 5** ([19, Corollary 1.2]). Let $\Phi$ be an $M \times N$ random Gaussian matrix. Set

$$\delta = C \left( \frac{s}{M} \log(2N/s) \log(2N/M + 2M/N) \right)^{1/5}. \hspace{1cm} (3.6)$$

Then for all $s$-sparse signals $x \in \mathbb{R}^N$, with probability at least $1 - C \exp(-c\delta M)$, the solution $\hat{x}$ of the convex minimization program (3.5) satisfies

$$\|\hat{x} - x\|_2 \leq \delta. \hspace{1cm} (3.7)$$

Here, $C$ and $c$ denote positive absolute constants and term $\|\hat{x} - x\|_2$ is a measure of the reconstruction quality. A useful conclusion of Theorem 5 can
be stated as follows:

An arbitrary accurate estimation of every $s$-sparse vector $x$ can be achieved from $M$ one-bit random measurements with order of $s \log^2(N/s)$. To be more precise $M \sim \delta^{-5}s\log^2(N/s)$ [19].

Due to time constraints, we do not prove Theorem 5 in this work. Theorem 5 has been proved in [19].

3.3 Iterative reconstruction algorithms for 1-bit compressive sensing

In this section, several iterative reconstruction algorithms for 1-bit compressive sensing are introduced. In Figure 3.1, the block diagrams of 1-bit compressive sensing and reconstruction part are shown. In this setup it is assumed that the binary measurements vector $b$ is noiseless. We will investigate the noisy scenario in the next chapter.

3.3.1 BIHT

Binary iterative hard thresholding (BIHT) is a reconstruction algorithm first introduced in [9] and further investigated in [7]. BIHT is derived from IHT and tries to minimize $\| b - \text{sign}(\Phi x) \|_2$ over $x$ where $b$ and $\Phi$ are given. Therefore, BIHT solves the following problem

$$\hat{x} = \arg \min_x \| b - \text{sign}(\Phi x) \|_2$$

subject to $\|x\|_0 \leq s$, $\|x\|_2 = 1$. \hfill (3.8)

It has been proved that minimization (3.8) is similar to [9, Lemma 5]

$$\hat{x} = \arg \min_x \| (b \odot \Phi x)^- \|_1$$

subject to $\|x\|_0 \leq s$, $\|x\|_2 = 1$, \hfill (3.9)

where $\odot$ denotes the element-wise vector multiplication, i.e., $[u \odot v]_i = u_i v_i$ and $(\cdot)^-$ denotes the negative function, i.e. $[(v)^-]_i = \begin{cases} -v_i & v_i \leq 0 \\ 0 & v_i > 0 \end{cases}$.
**Algorithm 2 BIHT**

1. **Inputs:** binary measurements vector \( b \in \{\pm 1\}^M \), measurement matrix \( \Phi \), sparsity level of the signal \( s \), descent step size \( \tau \), number of iterations \( t \)

2. **Initialization:** Initial estimate \( x^{[0]} = 0 \)

3. **Iteration:** For \( n = 1, \ldots, t \)
   
   (a) **Gradient descent:** \( z^{[n]} \leftarrow x^{[n-1]} - \tau \Phi^T (\text{sign}(\Phi x^{[n-1]}) - b) \)
   
   (b) **Projection onto \( \ell_0 \)-ball:** \( x^{[n]} \leftarrow H_s(z^{[n]}) \)

4. **Output:** \( \hat{x} = \frac{x^{[t]}}{\|x^{[t]}\|_2} \)

In other words, when the signal is approximated perfectly we have \( b = \text{sign}(\Phi x) \) and \( [b \odot \Phi x]_i > 0 \) for all \( i \). BIHT tries to minimize the summation of the absolute values of the negative elements in \( b \odot \Phi x \) iteratively. The iterative step of BIHT is

\[
x^{[n+1]} = H_s(x^{[n]} - \Phi^T(\text{sign}(\Phi x^{[n]}) - b)). \tag{3.10}
\]

In Algorithm 2 the steps of BIHT are shown in detail.
3.3.2 BIHT-$\ell_2$

If $\ell_2$-norm is applied in BIHT instead of $\ell_1$-norm, we obtain BIHT-$\ell_2$ which has been discussed in [7, 9, 20]. BIHT-$\ell_2$ algorithm can be thought of as trying to solve:

$$\hat{x} = \arg \min_x \| (b \odot \Phi x)^- \|_2$$

subject to $\|x\|_0 \leq s$, $\|x\|_2 = 1$. \hfill (3.11)

In fact, BIHT-$\ell_2$ tries to minimize

$$\sum_{i \in V} ((b \odot \Phi x)_i)^2,$$ \hfill (3.12)

where $x$ is $s$-sparse and $V$ is a set of indices in which $b \odot \Phi x$ is negative. In [9], it is shown that the solution of (3.11) can be obtained by simply iterating the following step:

$$x^{[n+1]} = H_s(x^{[n]} - (\text{diag}(b)\Phi)^T (b \odot \Phi x^{[n]}^-), \hfill (3.13)$$

where $\text{diag}(b)$ denotes a square diagonal matrix whose diagonal is vector $b$. The steps of (3.11) are depicted in Algorithm 3.

Algorithm 3 BIHT-$\ell_2$

1. **Inputs:** binary measurements vector $b \in \{\pm 1\}^M$, measurement matrix $\Phi$, sparsity level of the signal $s$, descent step size $\tau$, number of iterations $t$

2. **Initialization:** Initial estimate $x^{[0]} = 0$

3. **Iteration:** For $n = 1, \ldots, t$

   (a) **Gradient descent:**
   $$z^{[n]} \leftarrow x^{[n-1]} - \tau (\text{diag}(b)\Phi)^T (\text{diag}(b)\Phi x^{[n-1]}^-)$$

   (b) **Projection onto “$\ell_0$-ball”:** $x^{[n]} \leftarrow H_s(z^{[n]})$

4. **Output:** $\hat{x} = \frac{x^{[n]}}{\|x^{[n]}\|_2}$
3.3.3 RFPI

Another approach to signal reconstruction for 1-bit compressive sensing is introduced in [8]. In this method, the recovered signal is obtained by solving minimization program:

$$\hat{x} = \arg \min_x \|x\|_1$$

subject to $b \odot \Phi x \succeq 0$

and

$$\|x\|_2 = 1,$$

(3.14)

where $\succeq$ denotes element-wise inequality. To solve (3.14) efficiently, a barrier cost function is introduced in [8], which, together with Lagrange multiplier method [21], yields the following approximation of (3.14)

$$\hat{x} = \arg \min_x \|x\|_1 + \lambda \sum_i f ([b \odot \Phi x]_i)$$

subject to $\|x\|_2 = 1,$

(3.15)

where

$$f(x) = \begin{cases} 
x^2/2, & \text{if } x < 0 \\
0, & \text{otherwise}
\end{cases}$$

(3.16)

is the barrier cost function. Note that when $\lambda$ is sufficiently large, the solution of (3.14) coincides with the solution of (3.15). Since (3.16) is convex and smooth, the gradient descent can be applied to solve the minimization problem (3.15). **Renormalized fixed point iteration (RFPI)** is the algorithm introduced in [8] and solves (3.15). In Algorithm 4 the steps of RFPI are shown. In this algorithm, $x^{[0]}$ is initialized by $\Phi^{-1}b / \|\Phi^{-1}b\|_2$. The calculation of pseudo-inverse considerably increases the complexity of the algorithm. However, it is shown that random initialization of RFPI converges with high probability [8].

To summarize, we introduced three 1-bit reconstruction algorithms in this chapter. In Table 3.1 an overview of the reconstruction algorithms are shown. As it is obvious in this table, in order to estimate the signal accurately, BIHT and BIHT-$\ell_2$ need one more input than RFPI does, which is the sparsity level of the signal, $s$. We use this characteristic of RFPI as an advantage in the next chapter to design a new reconstruction algorithm that does not require $s$ as an input. In addition, the task of the next chapter is to analyze some other reconstruction algorithms which are robust against the noise in the binary measurements vector.
Algorithm 4 RFPI

1. **Inputs:** vector of 1-bit measurements $b \in \{\pm 1\}^M$, measuring matrix $\Phi$, number of outer iterations $t_1$, number of inner iterations $t_2$

2. **Initialization:** descent step-size $\delta$, $x^{[0]} = \Phi^{-1}b / \|\Phi^{-1}b\|_2$, initial coefficient $\lambda^{[1]} = M$

3. **Outer iteration:** For $k = 1, \ldots, t_1$
   
   (a) **Inner iteration:** For $n = 1, \ldots, t_2$
      
      i. One-sided quadratic gradient:
      
         $s \leftarrow (\text{diag}(b)\Phi)^T (b \odot \Phi x^{[n-1]})$

      ii. Gradient projection on sphere surface:
         
         $g \leftarrow \langle s, x^{[n-1]} \rangle x^{[n-1]} - s$

      iii. One-sided quadratic gradient descent:
         
         $h \leftarrow x^{[n-1]} - \delta g$

      iv. Shrinkage ($\ell_1$ gradient descent):
         
         $[u]_i \leftarrow \text{sign} ([h]_i) \max \left\{ |[h]_i| - \frac{\delta}{\lambda^{[n]}}, 0 \right\}$, for all $i$

      v. Normalization: $x^{[n]} \leftarrow \frac{u}{\|u\|_2}$

   (b) **Initialize next inner iteration:**
      
      $x^{[0]} \leftarrow x^{[n]}$, $\lambda^{[k+1]} \leftarrow c\lambda^{[k]}$, where $c$ is a fixed constant.

4. **Output:** $\hat{x} = x^{[n]}$
<table>
<thead>
<tr>
<th>Method</th>
<th>Formula</th>
<th>Inputs</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIHT</td>
<td>$\text{arg min}_x \left| (b \odot \Phi x)^- \right|_1$ subject to $|x| \leq s, |x|_2 = 1$</td>
<td>$\Phi, b, s$</td>
</tr>
<tr>
<td>BIHT-$\ell_2$</td>
<td>$\text{arg min}_x \left| (b \odot \Phi x)^- \right|_2$ subject to $|x| \leq s, |x|_2 = 1$</td>
<td>$\Phi, b, s$</td>
</tr>
<tr>
<td>RFPI</td>
<td>$\text{arg min}_x |x|_1$ subject to $b \odot \Phi x \succeq 0, |x|_2 = 1$</td>
<td>$\Phi, b$</td>
</tr>
</tbody>
</table>

**Table 3.1:** Three reconstruction algorithms for 1-bit compressive sensing in the noiseless scenario.
Chapter 4

1-bit compressive sensing in the presence of noise

In the previous chapter, we discussed about 1-bit compressive sensing and the reconstruction algorithms. However, in the case that the binary measurements are contaminated with the noise, the reconstruction algorithms fail to estimate the signal perfectly. We start this chapter by modelling the noise in 1-bit compressive sensing and we introduce two different sources of the noise. Then, we explain some reconstruction algorithms designed to perform robustly against the noise in the binary measurements.

4.1 Noise modelling in 1-bit compressive sensing

4.1.1 Measurement noise

In the case that there is noise in the compressive sensing measurements, we have
\[ \tilde{y} = \Phi x + n \]
where \( \tilde{y} \) is the noisy measurements vector and \( n \) denotes white Gaussian additive noise vector, i.e., \( [n]_i \sim \mathcal{N}(0, \sigma_n^2) \). After 1-bit quantization of (4.1) we obtain the binary measurements vector contaminated with the measurement noise and the noisy binary measurements vector is denoted by \( \tilde{b}_m \). Hence,
\[ \tilde{b}_m = \text{sign}(\tilde{y}) = \text{sign}(\Phi x + n) \]
In fact the noise in the measurements vector causes sign flips in the binary measurements vector. As the variance of the noise increases it is more likely
to occur bit flips in $\tilde{b}_m$. The probability of bit flip in each bit of $\tilde{b}_m$ is denoted by $P_{fm}$ and is calculated in Appendix B.

4.1.2 Channel noise

In practice, especially in communications applications, the binary measurements vector $\tilde{b}_m$ is transmitted through a channel. The noise of this transmission channel causes extra sign flips on the binary measurements. We model the channel by a binary symmetric channel (BSC). The probability of sign flips in the BSC is

$$P_{fc} = P(\tilde{b}_m_i = +1 | \tilde{b}_c_i = -1) = P(\tilde{b}_m_i = -1 | \tilde{b}_c_i = +1)$$  \hspace{1cm} (4.3)

where $\tilde{b}_c$ denotes binary measurements vector contaminated with the channel noise.

4.1.3 A combined model for binary noise

As mentioned in the previous section, there are two different sources of the binary noise: the measurement noise and the channel noise. Having the sign flip probability, $P_{fm}$, caused by measurement noise, we can show the effect of measurement noise by a BSC. Then, we put the physical transmission channel after this model. Therefore, we have two BSC channels connected in serial. As Figure 4.1 depicts, the noiseless binary measurements $b$ passes through first BSC with sign flip probability $P_{fm}$ then its output is put to the second BSC which is the physical channel with sign flip probability $P_{fc}$. 

![Figure 4.1: Binary symmetric channel model](image)
\[ \tilde{b} \text{ denotes the final noisy binary data after the measurement noise and the channel noise.} \]

We replace these two BSCs by a simple equivalent BSC with sign flip probability \( P_f \). The value of \( P_f \) given \( P_{fm} \) and \( P_{fc} \) can be obtained from

\[
\begin{pmatrix}
1 - P_f & P_f \\
1 - P_f & P_f
\end{pmatrix}
= \begin{pmatrix}
1 - P_{fm} & P_{fm} \\
P_{fm} & 1 - P_{fm}
\end{pmatrix}\begin{pmatrix}
1 - P_{fc} & P_{fc} \\
P_{fc} & 1 - P_{fc}
\end{pmatrix}
\]

(4.4)

and

\[
P_f = P_{fc}(1 - P_{fm}) + P_{fm}(1 - P_{fc}).
\]

(4.5)

The approximate number of the total sign flips between \( b \) and \( \tilde{b} \) which is denoted by \( L \) can be obtain from

\[
L \approx MP_f.
\]

(4.6)

In fact, \( L \) is a measure of the noise level in 1-bit compressive sensing. Note that as \( M \) tends to infinity, \( L \) converges to \( MP_f \).

### 4.2 Iterative reconstruction algorithms for 1-bit compressive sensing in the presence of noise

The block diagram of 1-bit compressive sensing in the presence of noise is shown in Figure 4.2. The task of reconstruction algorithms in the presence of noise is to reconstruct the signal form the noisy binary measurements \( \tilde{b} \).

In the first part of this section, we explain an algorithm which reconstructs the signal in the presence of noise. Then, we introduce our reconstruction algorithm that works robustly against the binary noise and does not require sparsity level of the signal as an input.

#### 4.2.1 Adaptive outlier pursuit

**AOP and AOP-\( \ell_2 \)**

In [10], a reconstruction algorithm called *adaptive outlier pursuit* (AOP) has been proposed. In this algorithm, a binary vector \( \Lambda \) is defined in which the
position of 0’s shows the position of the sign flips and the position of 1’s shows the position of the unchanged bits in $\tilde{b}$, i.e., $[\Lambda]_i \in \{0, 1\}^M$. Therefore, the number of sign flips in $\tilde{b}$ can be obtained from

$$L = \sum_i (1 - [\Lambda]_i).$$

(4.7)

Given the number of sign flips, $L$, and the sparsity level of the signal, $s$, AOP tries to solve the following optimization problem

$$(\hat{x}, \hat{\Lambda}) = \arg\min_{x, \Lambda} \| (\Lambda \odot \tilde{b} \odot \Phi x)^- \|_1$$

subject to

$$\sum_i (1 - [\Lambda]_i) \leq L$$

$$\|x\|_0 \leq s$$

$$\|x\|_2 = 1$$

(4.8)

where $\hat{x}$ denotes the estimated signal and $\hat{\Lambda}$ denotes the estimated binary vector. Note that in the noiseless case $[\Lambda]_i = 1$ for all $i$ and the problem in (4.8) converts to BIHT. AOP solves (4.8) by iterating between two following steps:

Step 1: given $\hat{\Lambda}$ find

$$\hat{x} = \arg\min_x \| (\hat{\Lambda} \odot \tilde{b} \odot \Phi x)^- \|_1$$

subject to

$$\|x\|_0 \leq s$$

$$\|x\|_2 = 1$$

(4.9)

Step 2: given $\hat{x}$ find

$$\hat{\Lambda} = \arg\min_{\Lambda} \| (\Lambda \odot \tilde{b} \odot \Phi \hat{x})^- \|_1$$

subject to

$$\sum_i (1 - [\Lambda]_i) \leq L$$

(4.10)

First $\hat{x}$ is determined by solving (4.9), based on the current estimation of $\hat{\Lambda}$. In other words, AOP tries to minimize

$$\sum_{i \in V} [(\tilde{b} \odot \Phi x)^-]_i,$$

(4.11)

where $x$ is $s$-sparse, $V = \text{Supp}(\hat{\Lambda})$ and $\text{Supp}(\cdot)$ gives the indices of the non-zero elements in the argument. In the second step, $\hat{\Lambda}$ is updated by solving
Algorithm 5 AOP

1. **Inputs:** Measurement signs $\tilde{b} \in \{\pm 1\}^M$, Measurement matrix $\Phi$, Signal sparsity $s$, Constant $\tau$

2. **Initialization:** Initial estimate $x[0] = \Phi^T \tilde{b} / \|\Phi^T \tilde{b}\|_2$

3. **Iteration:** For $n = 1, \ldots, t$
   
   (a) **Update support:** $V \leftarrow \text{Supp}(\Lambda)$

   (b) **Gradient descent:**
   
   $z[n] \leftarrow x[n-1] - \tau (\Phi_{[V,:]}^T \left( \text{sign}(\Phi_{[V,:]}x[n-1]) - \tilde{b}_V \right)$

   (c) **Projection onto \( \ell_0 \)-ball:** $x[n] \leftarrow H_s(z[n])$

   (d) **Update \( \Lambda \):** Update $\Lambda$ from (4.12)

4. **Output:** $\hat{x} = \frac{x[n]}{\|x[n]\|_2}$

(4.10), based on the new estimation of $\hat{x}$. In [10], it is shown that (4.10) can be solved analytically and its solution is given by

$$\left[\hat{\Lambda}\right]_i = \begin{cases} 
0, & \text{if } \left[\left(\tilde{b} \odot \Phi \hat{x}\right)^-\right]_i \geq \beta \\
1, & \text{otherwise.}
\end{cases}$$

(4.12)

Here, $\beta$ is the $L$th largest entry of the vector $\left(\tilde{b} \odot \Phi \hat{x}\right)^-$. In Algorithm 5 the steps of AOP is illustrated. This algorithm is initialized with $x[0] = \Phi^T \tilde{b} / \|\Phi^T \tilde{b}\|_2$ and $\Phi_{[V,:]}$ denotes a sub-matrix of $\Phi$ which is restricted to the rows of $\Phi$ in set $V$.

Another version of AOP is AOP-$\ell_2$ which minimizes $\left\|\left(\Lambda \odot \tilde{b} \odot \Phi x\right)^-\right\|_2$ instead of $\left\|\left(\Lambda \odot \tilde{b} \odot \Phi x\right)^-\right\|_1$. Therefore, AOP-$\ell_2$ is basically derived from BIHT-$\ell_2$ and the steps are the same as in Algorithm 5 while the step (b) is replaced by

$$z[n] = x[n-1] - \left(\text{diag}(\tilde{b})_{[V,:]} \Phi_{[V,:]}\right)^T \left(\text{diag}(\tilde{b})_{[V,:]} \Phi_{[V,:]}x[n-1]\right)^-.$$  (4.13)
AOP and AOP-ℓ₂ with sign flips

Another version of AOP is called *adaptive outlier pursuit with sign flips* (AOP-f) \cite{10} which applies binary vector $\Omega \in \{\pm 1\}^M$ where $\hat{b} \odot b = \Omega$. That is, the position of $-1$’s shows the position of flipped bits and the position of $1$’s specifies the position of unchanged bits in $\hat{b}$. Therefore,

$$[\Omega]_i = 2[A]_i - 1 \quad (4.14)$$

In fact, if $\Omega$ is estimated perfectly, we have

$$b = \Omega \odot \hat{b}. \quad (4.15)$$

and

$$L = \frac{1}{2} \sum_i (1 - [\Omega]_i). \quad (4.16)$$

AOP-f is obtained by replacing $b$ with $\Omega \odot \hat{b}$ (which is the flipped version of $\hat{b}$) in BIHT. Both $x$ and $\Omega$ are estimated in each iteration. The minimization problem that AOP-f solves is as follows:

$$(\hat{x}, \hat{\Omega}) = \arg \min_{x, \Omega} \left\| \left( \Omega \odot \hat{b} \odot \Phi x \right)^- \right\|_1$$

subject to

$$\frac{1}{2} \sum_i (1 - [\Omega]_i) \leq L$$

$$\|x\|_0 \leq s$$

$$\|x\|_2 = 1 \quad (4.17)$$

The steps of solving (4.17) are the same as the steps in (4.8). That is, first by fixing $\hat{\Omega}$, AOP-f estimates $\hat{x}$. Hence,

$$\hat{x} = \arg \min_x \left\| \left( \Omega \odot \hat{b} \odot \Phi \hat{x} \right)^- \right\|_1$$

subject to

$$\|x\|_0 \leq s$$

$$\|x\|_2 = 1 \quad (4.18)$$

Then, AOP-f updates $\hat{\Omega}$ given $\hat{x}$ obtained from previous step. Therefore,

$$\hat{\Omega} = \arg \min_{\Omega} \left\| \left( \Omega \odot \hat{b} \odot \Phi \hat{x} \right)^- \right\|_1$$

subject to

$$\frac{1}{2} \sum_i (1 - [\Omega]_i) \leq L. \quad (4.19)$$
Algorithm 6 AOP-f

1. **Inputs:** Measurement signs $\tilde{b} \in \{\pm 1\}^M$, Measurement matrix $\Phi$, Signal sparsity $s$, Constant $\tau$

2. **Initialization:** Initial estimate $x^{[0]} = \Phi^T \tilde{b} / \left\| \Phi^T \tilde{b} \right\|_2$

3. **Iteration:** For $n = 1, \ldots, t$
   
   (a) **Gradient Descent:** $z^{[n]} \leftarrow x^{[n-1]} - \tau \Phi^T \left( \text{sign} \left( \Phi x^{[n-1]} \right) - \tilde{b} \right)$
   
   (b) **Projection onto “$\ell_0$-ball”:** $x^{[n]} \leftarrow H_s(z^{[n]})$
   
   (c) **Update $\Omega$:** Update $\Omega$ from (4.20)
   
   (d) **Update $\tilde{b}$:** $\tilde{b} \leftarrow \Omega \odot \tilde{b}$

4. **Output:** $\hat{x} = \frac{x^{[n]}}{\left\| x^{[n]} \right\|_2}$

The estimation of $\hat{\Omega}$ in the second step is given by

$$\hat{\Omega}_i = \begin{cases} -1, & \text{if } \left[ (\tilde{b} \odot \Phi \hat{x})^- \right]_i \geq \beta \\ 1, & \text{otherwise} \end{cases} \quad (4.20)$$

where $\beta$ has the same value as in (4.12). Like in AOP, we can replace $\ell_2$-norm by $\ell_1$-norm which yields AOP-$\ell_2$-f. Note that since elements in $(\Omega \odot \tilde{b} \odot \Phi \hat{x})^-$ are non-negative values, the estimation of $\hat{\Omega}$ in AOP-f-$\ell_2$ is also given by (4.20). If the binary measurements vector is noiseless then $[\Omega]_i = 1$ for all $i$ and AOP-f (AOP-$\ell_2$-f) converts to BIHT (BIHT-$\ell_2$). Algorithm 6 shows the steps in AOP-f. To convert AOP-f to AOP-$\ell_2$-f, we just need to replace step (a) by

$$z^{[n]} = x^{[n-1]} - \left( \text{diag}(\tilde{b}) \Phi \right)^T \left( \text{diag}(\tilde{b}) \Phi x^{[n-1]} \right)^- \quad (4.21)$$

4.2.2 Noise-adaptive renormalized fixed point iterative

In this section, we introduce noise-adaptive renormalized fixed point iterative (NARFPI) which is a reconstruction algorithm mostly derived from RFPI.
By applying (4.15) we modify (3.14) to account for bit flips as follows:

\[
\hat{x}, \hat{\Omega} = \arg \min_{x, \Omega} \|x\|_1 \\
\text{subject to} \quad \Omega \odot \tilde{b} \odot \Phi x \succeq 0 \\
\frac{1}{2} \sum_i (1 - [\Omega]_i) \leq L \\
\|x\|_2 = 1.
\] (4.22)

To solve (4.22) efficiently, we can apply the same relaxation step as in (3.15) and approximate (4.22) by

\[
\hat{x}, \hat{\Omega} = \arg \min_{x, \Omega} \|x\|_1 + \lambda \sum_i f([\hat{\Omega} \odot \tilde{b} \odot \Phi x]_i) \\
\text{subject to} \quad \frac{1}{2} \sum_i (1 - [\Omega]_i) \leq L \\
\|x\|_2 = 1.
\] (4.23)

The optimization problem in (4.23) is still non-convex and consists of a combination of discrete and continuous variables. Similarly to the approach in [8] to solve (3.15), we use two steps algorithm to find \(\hat{\Omega}\) and \(\hat{x}\) in (4.23). In the first step, \(\hat{\Omega}\) is fixed and the algorithm finds the optimum \(\hat{x}\) as follows:

\[
\hat{x} = \arg \min_x \|x\|_1 + \lambda \sum_i f([\hat{\Omega} \odot \tilde{b} \odot \Phi x]_i) \\
\text{subject to} \quad \|x\|_2 = 1.
\] (4.24)

Note that the only difference between (3.15) and (4.24) is that \(b\) is replaced by \(\tilde{b} \odot \hat{\Omega}\). Hence, we can use RFPI to solve (4.24). In the second step, we use \(\hat{x}\) obtained from (4.24), to find \(\hat{\Omega}\) as follows:

\[
\hat{\Omega} = \arg \min_{\Omega} \sum_i f([\hat{\Omega} \odot \tilde{b} \odot \Phi \hat{x}]_i) \\
\text{subject to} \quad \frac{1}{2} \sum_i (1 - [\Omega]_i) \leq L.
\] (4.25)

We can rewrite (4.25) as

\[
\hat{\Omega} = \arg \min_{\Omega} \left\| \left(\Omega \odot \tilde{b} \odot \Phi \hat{x}\right)^{-}\right\|_2 \\
\text{subject to} \quad \frac{1}{2} \sum_i (1 - [\Omega]_i) \leq L.
\] (4.26)
The minimization in (4.26) is identical to the second step of AOP-ℓ₂-f (when ℓ₁-norm is replaced by ℓ₂-norm in (4.19)). Therefore, the solution of (4.26) is (4.20). The details of NARFPI is shown in Algorithm 7. In Table 4.1, the algorithms designed for 1-bit compressive sensing in the presence of the binary noise are described briefly. The main advantage of NARFPI in comparison to the other algorithms is that it does not require a priori knowledge of s as an input.

Algorithm 7 NARFPI

1. **Inputs**: vector of 1-bit measurements \( \tilde{b} \in \{\pm 1\}^M \), measuring matrix \( \Phi \), number of bit flips \( L \), number of outer iterations \( t_1 \), number of inner iterations \( t_2 \)

2. **Initialization**: descent step-size \( \delta \), initial estimate \( [\Omega]_i = 1 \) for all \( i \), \( x^{[0]} = \Phi^{-1}b / \| \Phi^{-1}b \|_2 \), initial coefficient \( \lambda^{[1]} = M \)

3. **Outer iteration**: For \( k = 1, \ldots, t_1 \)
   
   (a) **Inner iteration**: For \( n = 1, \ldots, t_2 \)
      
      i. One-sided quadratic gradient:
      \[ s \leftarrow \left(\text{diag}(\tilde{b})\Phi\right)^T \left( \tilde{b} \odot \Phi x^{[n-1]} \right) \]

      ii. Gradient projection on sphere surface:
      \[ g \leftarrow \langle s, x^{[n-1]} \rangle x^{[n-1]} - s \]

      iii. One-sided quadratic gradient descent:
      \[ h \leftarrow x^{[n-1]} - \delta g \]

      iv. Shrinkage (ℓ₁ gradient descent):
      \[ [u]_i \leftarrow \text{sign}(|h_i|) \max \left\{ |h_i| - \frac{\delta}{\lambda^{[i]}}, 0 \right\} , \text{for all } i \]

      v. Normalization: \( x^{[n]} \leftarrow \frac{u}{\|u\|_2} \)

   (b) Find the location of noisy bits and flip them:
   Update \( \Omega \) from (4.20). \( \tilde{b} \leftarrow \Omega \odot \tilde{b} \).

   (c) Initialize next inner iteration:
   \( x^{[0]} \leftarrow x^{[n]} \), \( \lambda^{[k+1]} \leftarrow c\lambda^{[k]} \), where \( c \) is a fixed constant.

4. **Output**: \( \hat{x} = x^{[n]} \)
<table>
<thead>
<tr>
<th>Method</th>
<th>Formula</th>
<th>Inputs</th>
</tr>
</thead>
<tbody>
<tr>
<td>AOP</td>
<td>$\arg\min_{x,\Lambda} \left| (\Lambda \odot \tilde{b} \odot \Phi x)^{-} \right|_1$</td>
<td>$\Phi, \tilde{b}, s, L$</td>
</tr>
<tr>
<td>subject to $\sum_i (1 - [\Lambda]_i) \leq L$</td>
<td>$|x|_0 \leq s$</td>
<td>$|x|_2 = 1$</td>
</tr>
<tr>
<td>AOP-$\ell_2$</td>
<td>$\arg\min_{x,\Lambda} \left| (\Lambda \odot \tilde{b} \odot \Phi x)^{-} \right|_2$</td>
<td>$\Phi, \tilde{b}, s, L$</td>
</tr>
<tr>
<td>subject to $\sum_i (1 - [\Lambda]_i) \leq L$</td>
<td>$|x|_0 \leq s$</td>
<td>$|x|_2 = 1$</td>
</tr>
<tr>
<td>AOP-f</td>
<td>$\arg\min_{x,\Omega} \left| (\Omega \odot \tilde{b} \odot \Phi x)^{-} \right|_1$</td>
<td>$\Phi, \tilde{b}, s, L$</td>
</tr>
<tr>
<td>subject to $\frac{1}{2} \sum_i (1 - [\Omega]_i) \leq L$</td>
<td>$|x|_0 \leq s$</td>
<td>$|x|_2 = 1$</td>
</tr>
<tr>
<td>AOP-f-$\ell_2$</td>
<td>$\arg\min_{x,\Omega} \left| (\Omega \odot \tilde{b} \odot \Phi x)^{-} \right|_2$</td>
<td>$\Phi, \tilde{b}, s, L$</td>
</tr>
<tr>
<td>subject to $\frac{1}{2} \sum_i (1 - [\Omega]_i) \leq L$</td>
<td>$|x|_0 \leq s$</td>
<td>$|x|_2 = 1$</td>
</tr>
<tr>
<td>NARFPI</td>
<td>$\arg\min_{x,\Omega} |x|_1$</td>
<td>$\Phi, \tilde{b}, L$</td>
</tr>
<tr>
<td>subject to $\Omega \odot \tilde{b} \odot \Phi x \succeq 0$</td>
<td>$\frac{1}{2} \sum_i (1 - [\Omega]_i) \leq L$</td>
<td>$|x|_2 = 1$</td>
</tr>
</tbody>
</table>

Table 4.1: Five reconstruction algorithms for 1-bit compressive sensing in the presence of the binary noise.
Chapter 5

Simulations and numerical results

In this chapter, we simulate the 1-bit compressive sensing reconstruction algorithms and analyze their performances in the noiseless and the noisy scenarios. The quality of reconstruction is measured in terms of the received signal to noise ratio (RSNR) defined as:

\[
\text{RSNR} = \frac{\mathbb{E}(\|x\|_2^2)}{\mathbb{E}(\|\hat{x} - x\|_2^2)}.
\]

Throughout this chapter the dimension of \( x \) is \( N = 1000 \). The position of non-zero elements in \( x \) is chosen uniformly at random and the amplitude of non-zero elements is generated according to zero-mean Gaussian variable with unit variance. The \( M \times N \) measuring matrix \( \Phi \) has independent entries following a zero-mean Gaussian distribution with variance \( 1/M \), i.e., \( \phi_{i,j} \sim \mathcal{N}(0, 1/M) \). We choose the number of binary measurements from multiples of 100 between 100 to 2000. This setting is beyond the classical compressive sensing goal of few measurements, i.e., \( M \ll N \). Note that, we can afford more measurements for a same bit budget compared to more sophisticated quantized compressive sensing approaches. In other words, in 1-bit compressive sensing each measurement is shown by one bit while in other quantized compressive sensing each measurement is shown by two or more bit. Therefore, 1-bit compressive sensing gives the most number of measurements in comparison to others given a fixed number of bits. In the case that the algorithms need the value of \( L \) as an input, we feed them by \( MP_f \) which is the estimated value of \( L \) (see (4.6)). Moreover, we set the number of iterations \( t = 1000 \), \( t_1 = 20 \) and \( t_2 = 200 \).
5.1 Algorithms designed for noiseless 1-bit compressive sensing

In this section, we simulate BIHT, BIHT-\(\ell_2\) and RFPI in three different binary noise levels. The sparsity level of \(x\) is set to \(s = 10\). We set the probability of bit flips to 0, 1% and 3%. The performance of the three algorithms is averaged over 100 realizations for each \(M/N\). In Figures 5.1, 5.2 and 5.3, the RSNR(dB) of the three algorithms is shown respectively for \(P_f = 0\), \(P_f = 1\%\) and \(P_f = 3\%\). Generally, as \(P_f\) increases, the overall reconstruction performance of these three algorithms decreases significantly. When \(P_f = 0\), the performance of BIHT is considerably higher than BIHT-\(\ell_2\) and RFPI. However, as the probability of sign flips increases BIHT-\(\ell_2\) outperforms BIHT and RFPI. This superior performance of BIHT-\(\ell_2\) in the presence of the noise claims that for scenarios with high level of binary noise, BIHT-\(\ell_2\) has the best reconstruction quality among the 1-bit reconstruction algorithms designed for the noiseless case.

5.2 Algorithms designed for noisy 1-bit compressive sensing

In [10], it is shown that the performance of AOP (AOP-\(\ell_2\)) is almost identical to the performance of AOP-f (AOP-f-\(\ell_2\)). Therefore, for the sake of simplicity, we just evaluate the performance of AOP-f (AOP-f-\(\ell_2\)) through out this
Figure 5.2: The performance of BIHT, BIHT-$\ell_2$ and RFPI when $P_f = 1\%$

Figure 5.3: The performance of BIHT, BIHT-$\ell_2$ and RFPI when $P_f = 3\%$
Figure 5.4: The performance of AOP-f, AOP-f-ℓ₂ and NARFPI when $P_f = 3\%$.

5.2.1 Signals with fixed sparsity level

Similarly to the previous section, the sparsity level of $x$ is set to $s = 10$. We simulate AOP-f, AOP-f-ℓ₂ and NARFPI in the scenario that $P_f = 3\%$. In Figure 5.4, the performance of the three algorithms are shown by RSNR(dB). AOP-f outperforms NARFPI and AOP-f-ℓ₂. In addition, the performance of AOP-f-ℓ₂ is almost identical to the performance of NARFPI since both of these algorithms try to minimize an ℓ₂-norm.

5.2.2 Signals with random sparsity level

In practice, the exact sparsity level of the signal is not known. As we mentioned before AOP-f and AOP-f-ℓ₂ need a priori knowledge of sparsity level of the signal to be reconstructed as an input. In this section, we simulate the case in which sparsity level of the signal varies randomly based on discrete truncated triangular distribution with mean 10 and $s \in [1, 19]$. The probability mass function (PMF) of the random $s$ for three different $\sigma^2$ is shown in Figure 5.5.

We simulate RFPI, AOP-f, AOP-f-ℓ₂ and NARFPI when $M/N = 2$ and $s$ is random based on the triangular distribution. The probability of sign flips is set to $P_f = 3\%$. Since the mean of $s$ is 10, we put 10 as an estimated $s$
into AOP-f and AOP-f-ℓ₂. In Figure 5.6 the average performance of the four algorithms over 100 realizations for each σ²_s is shown. As it is illustrated, the performance of NARFPI is constant and independent of σ²_s. RFPI exhibits also a constant RSNR as σ²_s varies, but its performance is poor because of the presence of bit flips. In contrast, the reconstruction quality of AOP-f and AOP-f-ℓ₂ decreases as deviation of s from its mean increases. When σ²_s > 10, NARFPI outperforms all the other algorithms.

To investigate whether NARFPI apparent superior performance occurs also for other M values, we consider another scenario in which s is fixed to a value between 1 to 19 (but AOP-f and AOP-f-ℓ₂ is still given 10 as estimate) and consider different values of M. The other parameters in this numerical experiment are the same as in the previous simulations. We plot 1/(RSNR) (i.e., the reconstruction error) in linear scale as a function of s for NARFPI (Figure 5.7), AOP-f (Figure 5.8) and AOP-f-ℓ₂ (Figure 5.9).

As expected, the reconstruction error of AOP-f and AOP-f-ℓ₂ is almost constant for s ≤ 10 but appears to grow faster than linearly in s when s exceeds 10. The error growth rate for 10 ≤ s in AOP-f is considerably faster than the error growth rate in AOP-f-ℓ₂. This behaviour seems natural given that AOP-f (AOP-f-ℓ₂) minimizes the ℓ₁-norm (ℓ₂-norm) of x under the constraint that ℓ₀ ≤ s and that we give s = 10 to AOP-f (AOP-f-ℓ₂) as estimate of the signal sparsity level. By comparing Figures 5.7, 5.8 and 5.9 we see that in the regime where the number of measurements is large compared to the signal dimension (e.g., M/N = 2), NARFPI outperforms
Figure 5.6: The performance of RFPI, AOP-f, AOP-f-$\ell_2$ and NARFPI when $P = 3\%$ and $s$ is random based on triangular distribution.

Figure 5.7: The estimation error of NARFPI as a function of the number of binary measurements $M$ and of the sparsity level $s$ of $x$. 
Figure 5.8: The estimation error of AOP-f as a function of the number of binary measurements $M$ and of the sparsity level $s$ of $x$.

Figure 5.9: The estimation error of AOP-f-$\ell_2$ as a function of the number of binary measurements $M$ and of the sparsity level $s$ of $x$. 
AOP-f and AOP-f-$\ell_2$. However, in the regime where the number of measurements is small compared to the signal dimension (e.g., $M/N = 0.5$), AOP-f outperforms NARFPI for $s > 10$. 
In this thesis, we discussed the general compressive sensing problem and the conditions on measuring matrix $\Phi$ which ensures the uniqueness of the reconstructed signal through compressive sensing reconstruction algorithms. We showed that RIP holds for a random matrix $\Phi$ when: 1) the entries of $\Phi$ are strictly sub-Gaussian and 2) the number of rows in $\Phi$ is greater than a particular value (related to the sparsity level and the dimension of the signal). Therefore, $\Phi$ can be applied in compressive sensing and signals measured through this class of matrices can be reconstructed perfectly.

Furthermore, we focused on a quantized version of compressive sensing problems, which is 1-bit compressive sensing. We investigated several 1-bit compressive sensing reconstruction algorithms (BIHT, BIHT-$\ell_2$ and RFPI) for noiseless scenario. We modelled the measurement and the channel noise in 1-bit compressive sensing and discussed four reconstruction algorithms (AOP, AOP-$\ell_2$, AOP-f, AOP-f-$\ell_2$) which are designed to work robustly in the presence of the noise. After that, we introduced our contribution, which is NARFPI, an algorithm obtained by merging RFPI and AOP-f. The main advantage of NARFPI over the previously introduced 1-bit compressive algorithms is that NARFPI does not require a priori knowledge of the sparsity level of the signal as an input to reconstruct the signals robustly in the presence of the noise.

The performance of all the 1-bit compressive algorithms in both the noiseless and the noisy scenarios were evaluated numerically in Chapter 5. From the simulations we found that:

- When the binary measurements vector $b$ is noiseless, BIHT has less reconstruction error than BIHT-$\ell_2$. However, as the level of the noise
increases BIHT-ℓ₂ outperforms BIHT. The reason of this behaviour lies in the difference between ℓ₁-norm and ℓ₂-norm minimization applied in these two algorithms. The ℓ₁-norm minimization gives more accurate signal estimation than the ℓ₂-norm minimization when there is no noise in the binary measurements. In contrast, ℓ₂-norm minimization outperforms ℓ₁-norm minimization when there is noise in the binary measurements.

- In the scenario that the sparsity level of the signal is perfectly known, the performance of NARFPI is similar to the performance of AOP-f-ℓ₂ in different noisy scenarios. The ℓ₂-norm minimization in both NARFPI and AOP-f-ℓ₂ causes this similarity. In contrast, AOP-f outperforms NARFPI and AOP-f-ℓ₂ because of the ℓ₁-norm applied in AOP-f. Generally, we can conclude that in the noisy case and when the sparsity level of the signal is perfectly known, the reconstruction algorithms based on the ℓ₁-norm minimization have better performance than the reconstruction algorithms based on the ℓ₂-norm minimization.

- When the sparsity level of the signal deviates from its estimated value, NARFPI outperforms both AOP-f and AOP-f-ℓ₂ because NARFPI does not require a priori knowledge of the signal sparsity level. In other words, NARFPI is appealing in practical scenarios in which there is no perfect knowledge about the sparsity level. For instance, the sparsity level of the transformed images through wavelet transformation is not a deterministic value but has a calculable distribution with particular mean and variance. Therefore, in this scenario NARFPI outperforms AOP-f and AOP-f-ℓ₂.

6.1 Suggestion for future work

In this work, we proposed NARFPI which is an iterative algorithm mainly derived from RFPI. The cost function inside RFPI works based on the ℓ₂-norm. Therefore, NARFPI is generally an ℓ₂-norm based algorithm. This ℓ₂-norm in NARFPI is the reason of the similarity between performance of NARFPI and AOP-f-ℓ₂. As mentioned in the previous section, in the low-noise regime ℓ₁-norm based algorithms, e.g., AOP-f reconstruct the signal more accurately than the ones based on ℓ₂-norm. As a future work, one may apply a new cost function in NARFPI that is based on ℓ₁-norm. It is expected that the performance of the new resulting algorithm is identical to the performance of AOP-f when the sparsity level of the signal is known.
Appendices
Proof of Theorems in Chapter 2

A.1 Proof of Theorem 1

Proof. First, by assuming that $\Phi$ satisfies the null space property, we prove the uniqueness of the solution. Let $z \in \mathbb{C}^N$ be such that $\Phi z = \Phi x$, $z \neq x$. Then we have to show that $\|x\|_1 < \|z\|_1$, in other words, $x$ is the sparsest possible solution. By assumption $v = x - z \in \ker \{\Phi\} \setminus \{0\}$. Let $T = \text{supp}(x)$, $v_T$ is the vector $v$ spanned to entries with indices in $T$ and $T^c = [N] \setminus T$ where $[N] = \{1, \ldots N\}$. Then we have

$$
\|x\|_1 = \|x - z_T + z_T\|_1 \\
\leq \|x - z_T\|_1 + \|z_T\|_1 = \|v_T\|_1 + \|z_T\|_1 \\
< \|v_{T^c}\|_1 + \|z_T\|_1 = \|-z_{T^c}\|_1 + \|z_T\|_1 = \|z\|_1 .
$$

(A.1)

Therefore, $x$ is the unique solution of (2.3).

In order to prove the converse of above, we take $v = x - z \in \ker \{\Phi\} \setminus \{0\}$ and $T \subset [N]$, $|T| = s$. By assumption, $v_T$ is the unique solution of $\arg \min_{z} \|z\|_1$ and $\Phi z = \Phi v_T$, i.e.,

$$
\Phi v_T = \Phi z \text{ and } \|v_T\|_1 < \|z\|_1 .
$$

(A.2)

Also, we have

$$
\Phi v_T = -\Phi v_{T^c}, (v_T \neq v_{T^c}) .
$$

(A.3)

Hence, (A.2) and (A.3) give

$$
\|v_T\|_1 < \|v_{T^c}\|_1 .
$$

(A.4)

Therefore, $\Phi$ satisfies the null space property. \qed
A.2 Proof of Theorem 2

In the first part of this section, we introduce some theorems and lemmas helping us to prove Theorem 2.

Proposition 1. For every matrix \( \Phi \in \mathbb{C}^{M \times N} \), restricted isometry constants are increasingly ordered, i.e.,
\[
\delta_1 \leq \delta_2 \leq \cdots \leq \delta_s \leq \cdots.
\] (A.5)

Proof. It is obvious that an \( s \)-sparse vector \( x \) can also be considered as an \( s + 1 \)-sparse vector. Therefore, we have
\[
(1 - \delta_{s+1}) \| x \|_2^2 \leq (1 - \delta_s) \| x \|_2^2 \leq \| \Phi x \|_2^2 \leq (1 + \delta_s) \| x \|_2^2 \leq (1 + \delta_{s+1}) \| x \|_2^2
\] (A.6)
and, consequently,
\[
\delta_s \leq \delta_{s+1}.
\] (A.7)

Theorem 6 (Rayleigh-Ritz, [22, Theorem 4.2.2]). Let \( \Phi \in \mathbb{C}^{N \times N} \) be Hermitian, and let the eigenvalues of \( \Phi \) be ordered as
\[
\lambda_{\min} = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_{N-1} \leq \lambda_N = \lambda_{\max}.
\] (A.8)

Then
\[
\lambda_1 x^\dagger x \leq x^\dagger \Phi x \leq \lambda_N x^\dagger x \quad \text{for all } x \in \mathbb{C}^N
\] (A.9)
\[
\lambda_{\max} = \lambda_N = \max_{x \neq 0} \frac{x^\dagger \Phi x}{x^\dagger x} = \max_{x^\dagger x = 1} x^\dagger \Phi x
\] (A.10)
\[
\lambda_{\min} = \lambda_1 = \min_{x \neq 0} \frac{x^\dagger \Phi x}{x^\dagger x} = \min_{x^\dagger x = 1} x^\dagger \Phi x.
\] (A.11)

Lemma 2. For every hermitian matrix \( \Phi \in \mathbb{C}^{N \times N} \) we have
\[
\max_{\| x \|_2 = 1} \| \Phi x \|_2 = \max_{\| x \|_2 = 1} |\langle \Phi x, x \rangle|.
\] (A.12)
Proof. By applying Theorem 1 we obtain

$$\max_{\|x\|_2=1} \|\Phi x\|_2 = \max \sqrt{\|\Phi x\|_2^2} = \max \sqrt{\frac{x^\dagger \Phi^\dagger \Phi x}{x^\dagger x}} = \sqrt{\lambda_{\max} (\Phi^\dagger \Phi)} = \sigma_{\max} (\Phi),$$

(A.13)

where $\sigma_{\max} (\Phi)$ is the maximum singular value of $\Phi$. In addition, $|\lambda (\Phi)| = \sigma (\Phi)$ since $\Phi$ is hermitian. Therefore,

$$\sigma_{\max} (\Phi) = |\lambda_{\max} (\Phi)| = \max \frac{x^\dagger \Phi x}{x^\dagger x} = \max \|x\|_2=1 \langle \Phi x, x \rangle. \quad (A.14)$$

$\Box$

**Proposition 2** ([12, Proposition 2.5.c]). Let $\Phi \in \mathbb{C}^{M \times N}$ with restricted isometry constant $\delta_s$ and $u, v \subseteq \mathbb{C}^N$ with disjoint support, i.e., $\text{supp} (u) \cap \text{supp} (v) = \emptyset$ and $s = |\text{supp} (u)| + |\text{supp} (v)|$. Then

$$|\langle \Phi u, \Phi v \rangle| \leq \delta_s \|u\|_2 \|v\|_2. \quad (A.15)$$

Proof. We can write (2.5) as

$$\|\Phi x\|_2^2 - \|x\|_2^2 \leq \delta_s \|x\|_2^2 \quad (A.16)$$

$$\frac{|\langle \Phi x, \Phi x \rangle - \langle x, x \rangle|}{\langle x, x \rangle} = \frac{|x^\dagger \Phi^\dagger \Phi x - x^\dagger x|}{x^\dagger x} = \frac{|\langle \Phi^\dagger \Phi x, x \rangle - \langle x, x \rangle|}{\langle x, x \rangle} = \frac{|\langle (\Phi^\dagger \Phi - I) x, x \rangle|}{\langle x, x \rangle} \leq \delta_s. \quad (A.17)$$

Therefore, from Lemma 2

$$\max_{\|x\|_2=1} |\langle (\Phi^\dagger \Phi - I) x, x \rangle| = \max_{\|x\|_2=1} \| (\Phi^\dagger \Phi - I) x \|_2 \leq \delta_s. \quad (A.18)$$

By definition $\text{supp} (v) = V$ and $\text{supp} (u) = U$ where $|V| + |U| = s$ and $V \cap U = \emptyset$. Let $0_U$ and $0_V$ be zero vectors with lengths $|U|$ and $|V|$, and $[0_V^\dagger, 0_U^\dagger]^\dagger$ is a vector obtained by merging vectors $0_V$ and $0_U$. In addition, $\Phi_{[\cdot, S]}$ denotes a sub-matrix of $\Phi$ spanned to columns in $S$ where $S$ is the
combination of indices in $V$ and $U$ in order. Then, we can write

$$
\langle \Phi_u, \Phi_v \rangle = \begin{bmatrix} v^\dagger & 0^\dagger \end{bmatrix} \begin{bmatrix} \Phi_v & \Phi_u \end{bmatrix} \begin{bmatrix} v^\dagger & 0^\dagger \end{bmatrix}_U = \begin{bmatrix} v^\dagger & 0^\dagger \end{bmatrix}_V \Phi_u \begin{bmatrix} v^\dagger & 0^\dagger \end{bmatrix}_U
$$

(A.19)

From Cauchy-Schwartz inequality we have

$$|p^\dagger \Delta q| = |\langle \Delta q, p \rangle| \leq \|\Delta q\|_2 \|p\|_2.$$  

(A.20)

Moreover, from (A.18) we know that

$$\frac{\|\Delta q\|_2}{\|q\|_2} \leq \max \frac{\|\Delta q\|_2}{\|q\|_2} \leq \delta_s.$$  

(A.21)

Hence,

$$|p^\dagger \Delta q| \leq \delta_s \|q\|_2 \|p\|_2 \Rightarrow |\langle \Phi_u, \Phi_v \rangle| \leq \delta_s \|u\|_2 \|v\|_2.$$  

(A.22)

Now we are ready to prove Theorem 2 by applying Propositions 1 and 2.

**Proof.** Take $v \in \ker \{\Phi\}$, let $T_0$ be the set of $s$ largest modulus entries of $v$ and $v_{T_0}$ is a vector spanned to entries of $v$ with indeces in $T_0$. Then $T_0^c = T_1 \cup T_2 \cup ...$ where $T_i$, for every $i > 0$, is the set containing $s$ largest modulus entries of $v_Q$, $Q = [N] \setminus \bigcup_{j=0}^{i-1} T_j$. Therefore, $\Phi(v_{T_0}) = -\Phi(v_{T_1} + v_{T_2} + ...)$ and by using (2.5) and Proposition 1 we obtain

$$1 - \delta_2 \leq 1 - \delta_s \quad \text{and} \quad (1 - \delta_2) \|v_{T_0}\|_2^2 \leq \|\Phi(v_{T_0})\|_2^2.$$  

(A.23)

Now we can write

$$\|\Phi v_{T_0}\|_2^2 = \langle \Phi v_{T_0}, \Phi v_{T_0} \rangle = \langle \Phi v_{T_0}, -\Phi(v_{T_1} + v_{T_2} + ...) \rangle = \sum_{k \geq 1} \langle \Phi v_{T_0}, -\Phi v_{T_k} \rangle \quad \text{for} \quad k \geq 1.$$  

(A.24)
Substituting (A.23) in (A.24) gives
\[ \|v_T\|_2^2 \leq \frac{1}{1 - \delta_{2s}} \sum_{k \geq 1} \langle \Phi v_T, -\Phi v_k \rangle \]  
(A.25)
and Proposition 2 yields
\[ \langle \Phi v_T, -\Phi v_k \rangle \leq \delta_{2s} \|v_T\|_2 \|v_k\|_2 . \]  
(A.26)
By substituting (A.26) in (A.25) we get
\[ \|v_T\|_2^2 \leq \frac{\delta_{2s}}{1 - \delta_{2s}} \sum_{k \geq 1} \|v_T\|_2 \|v_k\|_2 \]  
(A.27)
The \(s\) entries of \(v_k\) do not exceed the \(s\) entries of \(v_{k-1}\) so
\[ \forall j \in T_k \ |v_j| \leq \frac{1}{s} \sum_{l \in T_{k-1}} |v_l| = \frac{1}{s} \|v_{k-1}\|_1 \]
By using Cauchy-Schwartz inequality we have
\[ \|v_{T_{k}}\|_2 = \left( \sum_{j=1}^{\infty} |v_j|^2 \right)^{1/2} \leq \left( s \cdot \left( \frac{1}{s} \|v_{T_{k-1}}\|_1 \right)^2 \right)^{1/2} \]  
= \frac{1}{\sqrt{s}} \|v_{T_{k-1}}\|_1 . \]  
(A.28)
Assume that \(s_0\) is a \(N\)-dimensional vector and
\[ s_0 = \begin{cases} s_j = 1 & j \in T_0 \\ s_j = 0 & j \notin T_0 \end{cases} \]  
(A.29)
By using Cauchy-Schwartz inequality we have
\[ \|v_{T_0}\|_1 = |\langle v_{T_0}, s_0 \rangle| \leq \|v_{T_0}\|_2 \|s_0\|_2 = \sqrt{s} \|v_{T_0}\|_2 . \]  
(A.30)
Applying (A.27) and (A.28) in (A.30) gives
\[ \|v_{T_0}\|_1 \leq \sqrt{s} \|v_{T_0}\|_2 \leq \frac{\delta_{2s}}{1 - \delta_{2s}} \sum_{k \geq 1} \sqrt{s} \|v_k\|_2 \leq \frac{\delta_{2s}}{1 - \delta_{2s}} \sum_{k \geq 1} \|v_{k-1}\|_1 \]  
= \frac{\delta_{2s}}{1 - \delta_{2s}} \left( \|v_{T_0}\|_1 + \sum_{k \geq 1} \|v_k\|_1 \right) \leq \frac{\delta_{2s}}{1 - \delta_{2s}} (\|v_{T_0}\|_1 + \|v_{T_{0}}\|_1) \]  
(A.31)
Now if $\delta_{2s} < 1/3$ or $\frac{\delta_{2s}}{1-\delta_{2s}} < 1/2$, then
\[
\|v_{T_0}\|_1 \leq \frac{\delta_{2s}}{1-\delta_{2s}} (\|v_{T_0}\|_1 + \|v_{T_0}\|_1) < \frac{1}{2} (\|v_{T_0}\|_1 + \|v_{T_0}\|_1).
\] (A.32)

Therefore, we have that
\[
\|v_{T_0}\|_1 < \|v_{T_0}\|_1
\] (A.33)
so the null space property follows.

\section*{A.3 Proof of Theorem 3}

First we introduce some intermediate theorems and definitions which help us to prove Theorem 3.

\textbf{Proposition 3.} If $\Phi \in \mathbb{C}^{M \times N}$ satisfies RIP with Restricted Isometry Constant $\delta_s$ then for all $T \subset [N]$ and $|T| \leq s$ we have
\[
1 - \delta_s \leq \lambda_{\min} \left( \Phi_{[:,T]} \Phi_{[:,T]}^\dagger \right) \leq \lambda_{\max} \left( \Phi_{[:,T]} \Phi_{[:,T]}^\dagger \right) \leq 1 + \delta_s \quad \text{(A.34)}
\]
and
\[
\sqrt{1 - \delta_s} \leq \sigma_{\min} \left( \Phi_{[:,T]} \right) \leq \sigma_{\max} \left( \Phi_{[:,T]} \right) \leq \sqrt{1 + \delta_s}. \quad \text{(A.35)}
\]

\textbf{Proof.} We know that
\[
(1 - \delta_s) \|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1 + \delta_s) \|x\|_2^2 \quad \text{for every \ s-sparse} \ x.
\]
Then we can write
\[
\|\Phi x\|_2^2 - \|x\|_2^2 \leq \delta_s \|x\|_2^2. \quad \text{(A.36)}
\]
By applying $\|x\|_2^2 = \langle x, x \rangle = x^\dagger x$ we can write the left-hand side of (A.36) as
\[
\|\Phi x\|_2^2 - \|x\|_2^2 = |\langle \Phi x, \Phi x \rangle - \langle x, x \rangle| = |x^\dagger (\Phi \Phi - I) x| \quad \text{(A.37)}
\]
Substituting (A.37) in (A.36) yields
\[
x^\dagger (\Phi \Phi - I) x \leq \delta_s x^\dagger x.
\]
Since $x$ is $s$-sparse we obtain

$$x^\dagger_T x_T = x^\dagger x.$$ (A.38)

Therefore,

$$x^\dagger_T (\Phi^\dagger_{[\cdot,T]} \Phi_{[\cdot,T]} - I) x_T = x^\dagger (\Phi^\dagger \Phi - I) x,$$ (A.39)

where $T = \text{supp}(x)$. From (A.38), (A.39) and Theorem 6 we have (for all $s$-sparse $x$ with support $I$)

$$\lambda_{\min} (\Phi^\dagger_{[\cdot,T]} \Phi_{[\cdot,T]} - I) x^\dagger x \leq x^\dagger (\Phi^\dagger \Phi - I) x \leq \lambda_{\max} (\Phi^\dagger_{[\cdot,T]} \Phi_{[\cdot,T]} - I) x^\dagger x$$

$$|x^\dagger (\Phi^\dagger \Phi - I) x| \leq \max\{\lambda_{\max} (\Phi^\dagger_{[\cdot,T]} \Phi_{[\cdot,T]} - I) x^\dagger x, -\lambda_{\min} (\Phi^\dagger_{[\cdot,T]} \Phi_{[\cdot,T]} - I) x^\dagger x\}.$$ (A.40)

In other words, (A.40) implies that if $0 < \lambda_{\min} (\Phi^\dagger_{[\cdot,T]} \Phi_{[\cdot,T]} - I) x^\dagger x$ then

$$|x^\dagger (\Phi^\dagger \Phi - I) x| \leq \lambda_{\max} (\Phi^\dagger_{[\cdot,T]} \Phi_{[\cdot,T]} - I) x^\dagger x,$$ (A.41)

but in the case that $0 > \lambda_{\min} (\Phi^\dagger_{[\cdot,T]} \Phi_{[\cdot,T]} - I) x^\dagger x$ then simply $|x^\dagger (\Phi^\dagger \Phi - I) x| \leq \max\{\lambda_{\max} (\Phi^\dagger_{[\cdot,T]} \Phi_{[\cdot,T]} - I) x^\dagger x, -\lambda_{\min} (\Phi^\dagger_{[\cdot,T]} \Phi_{[\cdot,T]} - I) x^\dagger x\}$. Comparing (2.5) and (A.40) gives

$$\delta_s = \max\{\lambda_{\max} (\Phi^\dagger_{[\cdot,T]} \Phi_{[\cdot,T]} - I), -\lambda_{\min} (\Phi^\dagger_{[\cdot,T]} \Phi_{[\cdot,T]} - I)\}$$

or

$$\delta_s = \max\{\lambda_{\max} (\Phi^\dagger_{[\cdot,T]} \Phi_{[\cdot,T]}), 1, 1 - \lambda_{\min} (\Phi^\dagger_{[\cdot,T]} \Phi_{[\cdot,T]}), \}

1 - \delta_s \leq \lambda_{\min} (\Phi^\dagger_{[\cdot,T]} \Phi_{[\cdot,T]}) \leq \lambda_{\max} (\Phi^\dagger_{[\cdot,T]} \Phi_{[\cdot,T]}) \leq 1 + \delta_s.$$ (A.42)

□

**Definition 6** (Lipschitz condition, [23 page 46]). Let $f(x)$ be defined on an interval $I$ and suppose we can find a positive constant $\alpha$ such that

$$|f(x_1) - f(x_2)| \leq \alpha \|x_1 - x_2\|_2$$ (A.43)

for all $x_1, x_2 \in I$. Then $f$ is said to satisfy a Lipschitz condition with Lipschitz constant $\alpha$. 

48
**Example 1.** The maximum and minimum singular values of any $\Phi \in \mathbb{C}^{M \times N}$ are Lipschitz functions with Lipschitz constant 1.

**Proof.** From Theorem 6

\[
\max \frac{x^i \Phi^i \Phi x}{x^i x} = \max \frac{\|\Phi x\|_2^2}{\|x\|_2^2} = \lambda_{\max} (\Phi^i \Phi) = \sigma_{\max}^2 (\Phi) \quad (A.44)
\]

\[
\min \frac{x^i \Phi^i \Phi x}{x^i x} = \min \frac{\|\Phi x\|_2^2}{\|x\|_2^2} = \lambda_{\min} (\Phi^i \Phi) = \sigma_{\min}^2 (\Phi) . \quad (A.45)
\]

By definition we have

\[
\|\Phi\|_F = \sqrt{\text{tr} (\Phi^i \Phi)} = \sqrt{\sum \sigma_i^2 (\Phi)} \quad (A.46)
\]

where $\sigma_i (\Phi)$ is $i$-th singular value of $\Phi$ and $\|\cdot\|_F$ denotes Frobenius-norm. Therefore

\[
\sigma_{\min} (\Phi) \leq \sigma_{\max} (\Phi) \leq \|\Phi\|_F . \quad (A.47)
\]

We know from Triangular inequality that

\[
\text{for any } \Phi, \Xi \in \mathbb{C}^{M \times N} \quad \|\Phi x\|_2 \leq \|\Xi x\|_2 + \|(\Phi - \Xi) x\|_2 . \quad (A.48)
\]

From (A.48) we obtain

\[
\max \|\Phi x\|_2 \leq \max \|\Xi x\|_2 + \max \|(\Phi - \Xi) x\|_2 \quad (A.49)
\]

and

\[
\min \|\Phi x\|_2 \leq \min \|\Xi x\|_2 + \min \|(\Phi - \Xi) x\|_2 . \quad (A.50)
\]

Therefore, by substituting (A.44) and (A.45) in (A.49) and (A.50) we obtain

\[
\sigma_{\max} (\Phi) \leq \sigma_{\max} (\Xi) + \sigma_{\max} (\Phi - \Xi) \quad (A.51)
\]

\[
\sigma_{\min} (\Phi) \leq \sigma_{\min} (\Xi) + \sigma_{\min} (\Phi - \Xi) . \quad (A.52)
\]

In addition, from (A.47) we have

\[
\sigma_{\max} (\Phi) - \sigma_{\max} (\Xi) \leq \sigma_{\max} (\Phi - \Xi) \leq \|\Phi - \Xi\|_F \quad (A.53)
\]

\[
\sigma_{\min} (\Phi) - \sigma_{\min} (\Xi) \leq \sigma_{\min} (\Phi - \Xi) \leq \|\Phi - \Xi\|_F . \quad (A.54)
\]
Since we can interchange $\Phi$ and $\Xi$, we get
\[
|\sigma_{\text{max}}(\Phi) - \sigma_{\text{max}}(\Xi)| \leq \|\Phi - \Xi\|_F
\]
\[
|\sigma_{\text{min}}(\Phi) - \sigma_{\text{min}}(\Xi)| \leq \|\Phi - \Xi\|_F.
\] (A.55)

Therefore, $\sigma_{\text{max}}(\cdot)$ and $\sigma_{\text{min}}(\cdot)$ satisfy Lipschitz condition with constant $\alpha = 1$.

**Theorem 7** ([24, Lemma 2.2]). Let $\mathbf{x}$ be a $N$-dimensional vector of unit variance, independent Gaussian variables, if $f : \mathbb{R}^N \to \mathbb{R}$ is Lipschitz function with Lipschitz constant $\alpha$ then for all $t > 0$
\[
P(f(\mathbf{x}) - \mathbb{E}(f(\mathbf{x})) > t) \leq e^{-t^2/2\alpha^2}.
\] (A.56)

**Theorem 8** (Gordon’s theorem, [25, Theorem 5.32]). Let $\Phi$ be an $M$ by $N$ array of i.i.d $\mathcal{N}(0, 1)$ then
\[
\mathbb{E}(\sigma_{\text{min}}(\Phi)) \geq \sqrt{M} - \sqrt{N} \quad \text{and} \quad \mathbb{E}(\sigma_{\text{max}}(\Phi)) \leq \sqrt{M} + \sqrt{N}.
\] (A.57)

**Proposition 4.** For any matrix $\Phi \in \mathbb{C}^{M \times N}$ with i.i.d $\mathcal{N}(0, 1)$ modulus entries,
\[
P\left(\sigma_{\text{min}}(\Phi) < \sqrt{M} - \sqrt{N} - t\right) \leq e^{-t^2/2}
\]
\[
P\left(\sigma_{\text{max}}(\Phi) > \sqrt{M} + \sqrt{N} + t\right) \leq e^{-t^2/2}.
\] (A.58)

**Proof.** From Example 1 we know that $\sigma_{\text{min}}(\Phi)$ and $\sigma_{\text{max}}(\Phi)$ are Lipschitz with constant 1. Therefore, applying Theorem 7 gives
\[
P\left(\mathbb{E}(\sigma_{\text{min}}(\Phi)) - \sigma_{\text{min}}(\Phi) > t\right) \leq e^{-t^2/2}
\]
\[
P\left(\sigma_{\text{max}}(\Phi) - \mathbb{E}(\sigma_{\text{max}}(\Phi)) > t\right) \leq e^{-t^2/2}.
\] (A.59)

From Theorem 8
\[
\sqrt{M} - \sqrt{N} \leq \mathbb{E}(\sigma_{\text{min}}(\Phi)) \quad \text{and} \quad \sqrt{M} + \sqrt{N} \geq \mathbb{E}(\sigma_{\text{max}}(\Phi)).
\] (A.60)
Therefore,
\[
P \left( \sigma_{\min}(\Phi) < \sqrt{M} - \sqrt{N} - t \right) \leq P \left( \sigma_{\min}(\Phi) < \mathbb{E}(\sigma_{\min}(\Phi)) - t \right) \leq e^{-t^2/2}
\]
\[
P \left( \sigma_{\max}(\Phi) > \sqrt{M} + \sqrt{N} + t \right) \leq P \left( \sigma_{\max}(\Phi) > \mathbb{E}(\sigma_{\max}(\Phi)) + t \right) \leq e^{-t^2/2}.
\]
(A.61)

Now by applying Propositions 3 and 4 we prove Theorem 3.

**Proof.** Let \( T \subset [N] \) and \( |T| \leq 2s \). We are interested to find the case that Proposition 3 holds for \( \Phi \) with high probability. In other words, \( \Phi \) will satisfy RIP with high probability when
\[
P \left( \sigma_{\min}(\Phi_{[\cdot,T]}) < \sqrt{1 - \delta_{2s}} \right) \leq e^{-Mk^2/2}
\]
(A.62)
and
\[
P \left( \sigma_{\max}(\Phi_{[\cdot,T]}) > \sqrt{1 + \delta_{2s}} \right) \leq e^{-Mk^2/2}
\]
(A.63)
are small enough.

By using Proposition 4 and dividing left side statement by \( \sqrt{M} \), for every \( T \) we get
\[
P \left( \sigma_{\min}(\Phi_{[\cdot,T]}) < 1 - \sqrt{|T|/M} - k \right) \leq e^{-Mk^2/2}
\]
(A.64)
\[
P \left( \sigma_{\max}(\Phi_{[\cdot,T]}) > 1 + \sqrt{|T|/M} + k \right) \leq e^{-Mk^2/2}
\]
(A.65)
where \( k = t/\sqrt{M} \). Since \( \sigma_{\max}(\Phi_{[\cdot,T]}) \) and \( \sigma_{\min}(\Phi_{[\cdot,T]}) \) increase by decreasing \( |T| \), we have
\[
P_{T : |T| \leq 2s} \left( \sigma_{\max}(\Phi_{[\cdot,T]}) > \sqrt{1 + \delta_{2s}} \right) = P_{T : |T| = 2s} \left( \sigma_{\max}(\Phi_{[\cdot,T]}) > \sqrt{1 + \delta_{2s}} \right)
\]
(A.66)
and applying (A.65) over all \( T \)s satisfying \( |T| = 2s \) gives
\[
P_{T : |T| = 2s} \left( \sigma_{\max}(\Phi_{[\cdot,T]}) > 1 + \sqrt{2s/M} + k \right) \leq \# \{ T : |T| = 2s \} \cdot e^{-Mk^2/2}
\]
\[
\leq \binom{N}{2s} e^{-Mk^2/2}.
\]
(A.67)
We use following approximation
\[
\binom{N}{2s} \approx \left( \frac{N}{2s} \right)^{2s} = e^{2s \log(N/2s)}.
\]
(A.68)
Substituting above approximation in (A.67) gives
\[ \mathbb{P}_{T:|T|\leq 2s} \left( \sigma_{\max}(\Phi[:,T]) > 1 + \sqrt{2s/M + k} \right) < e^{2s \log(N/2s)} e^{-Mk^2/2}. \] (A.69)
Therefore we have
\[ \mathbb{P}_{T:|T|\leq 2s} \left( \sigma_{\max}(\Phi[:,T]) < 1 + \sqrt{2s/M + k} \right) \geq 1 - e^{2s \log(N/2s)} e^{-Mk^2/2}. \] (A.70)
Null space property holds when the conditions below are fulfilled. Firstly,
\[ 1 + \sqrt{2s/M + k} < \sqrt{1 + \delta_{2s}}, \] (A.71)
where \( \delta_{2s} < 1/3 \). In addition,
\[ 2s(\log(N/2s)) - Mk^2/2 \leq 0, \] (A.72)
which implies that the probability is always between zero and one. From (A.71) we have
\[ 0 < k < 0.15 - \sqrt{2s/M} \] (A.73)
On the other hand, from (A.72) we have
\[ \frac{4s(\log(N/s) - \log(2))}{k^2} = \frac{4s(\log(N/s) - 2.77s)}{k^2} \leq M. \] (A.74)
Therefore, by choosing \( 4s \log(N/s)/k^2 \leq M \), (A.74) holds. For \( \sigma_{\min}(\Phi) \) the proof is the same as above. In the case that (A.73) and (A.74) hold we have
\[ \mathbb{P}_{T:|T|\leq 2s} \left( \sigma_{\max}(\Phi[:,T]) > \sqrt{1 + \delta_{2s}} \right) \geq 1 - e^{2s \log(N/2s)} e^{-Mk^2/2} \] (A.75)
Similarly, for \( \sigma_{\min}(\Phi) \) we have
\[ \mathbb{P}_{T:|T|\leq 2s} \left( \sigma_{\min}(\Phi[:,T]) < \sqrt{1 - \delta_{2s}} \right) \geq 1 - e^{2s \log(N/2s)} e^{-Mk^2/2} \] (A.76)
Now if we name (A.75) \( \mathbb{P}(A) \) and (A.76) \( \mathbb{P}(B) \) then we are interested to find \( \mathbb{P}(A \cap B) \) and from basic probability theory we know
\[ \mathbb{P}(A \cap B) \geq \mathbb{P}(A) + \mathbb{P}(B) - 1. \] (A.77)
Therefore, from (A.75), (A.76) and (A.77) we obtain
\[ \mathbb{P}_{T:|T|\leq 2s} \left( \sqrt{1 - \delta_{2s}} \leq \sigma_{\min}(\Phi[:,T]) < \sigma_{\max}(\Phi[:,T]) < \sqrt{1 + \delta_{2s}} \right) \geq 1 - 2e^{2s \log(N/2s)} e^{-Mk^2/2} \] (A.78)
\[ \square \]
A.4 Proof of Theorem 4

We need to prove some intermediate theorems and lemmas which help us to prove Theorem 4. The main part of this section is derived from [16].

Lemma 3 ([16, Theorem 7.2]). Assume that \( x = [X_1, X_2, \ldots, X_N] \) where each \( X_i \sim \text{Sub}(c^2) \) is independent. Then for any \( a \in \mathbb{R}^N, \langle x, a \rangle \sim \text{Sub}(c^2 \|a\|_2^2) \). Similarly, if each \( X_i \sim \text{SSub}(\sigma^2) \), then for any \( a \in \mathbb{R}^N, \langle x, a \rangle \sim \text{SSub}(\sigma^2 \|a\|_2^2) \).

Proof. Since \( X_i \) are i.i.d, factorization gives

\[
\mathbb{E} \left( \exp \left( t \sum_{i=1}^{N} a_i X_i \right) \right) = \mathbb{E} \left( \prod_{i=1}^{N} \exp (t a_i X_i) \right) \\
= \prod_{i=1}^{N} \mathbb{E} (\exp(t a_i X_i)) \\
\leq \prod_{i=1}^{N} \exp \left( c^2 (t a_i)^2 /2 \right) \\
= \exp \left( \left( \sum_{i=1}^{N} a_i^2 \right) c^2 t^2 /2 \right). \quad (A.79)
\]

If the \( X_i \) are strictly sub-Gaussian, then setting \( c^2 = \sigma^2 \) gives \( \mathbb{E} \left( \langle x, a \rangle^2 \right) = \sigma^2 \|a\|_2^2 \). \( \square \)

Lemma 4 (Markov’s Inequality). For any random variable \( X > 0 \) and \( t > 0 \),

\[
\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}(X)}{t}. \quad (A.80)
\]

Proof. By noting that \( f(x) \) is the p.d.f. for \( X \)

\[
\mathbb{E}(X) = \int_{0}^{\infty} x f(x) \, dx \geq \int_{t}^{\infty} x f(x) \, dx \geq \int_{t}^{\infty} tf(x) \, dx = t \mathbb{P}(X \geq t). \quad (A.81)
\]

Lemma 5 ([16, Lemma 7.4]). Let \( X \sim \text{Sub}(c^2) \). Then

\[
\mathbb{E} \left( \exp \left( \lambda X^2 / 2c^2 \right) \right) \leq \frac{1}{\sqrt{1 - \lambda}} \quad \text{for any } \lambda \in [0, 1). \quad (A.82)
\]

53
Proof. If \( \lambda = 0 \), then the lemma simply holds. Now suppose that \( \lambda \in (0, 1) \). Since \( X \) is sub-Gaussian, we have
\[
\int_{-\infty}^{\infty} \exp (tx) f(x) \, dx \leq \exp \left( c^2 t^2 / 2 \right) \quad \text{for any} \; t \in \mathbb{R}. \tag{A.83}
\]
By multiplying both sides by \( \exp (-c^2 t^2 / 2\lambda) \) we get
\[
\int_{-\infty}^{\infty} \exp (tx - c^2 t^2 / 2\lambda) f(x) \, dx \leq \exp \left( c^2 t^2 (\lambda - 1) / 2 \right). \tag{A.84}
\]
Now integrating both sides with respect to \( t \) gives
\[
\int_{-\infty}^{\infty} \left( \int_{-\infty}^{\infty} \exp (tx - c^2 t^2 / 2\lambda) \, dt \right) f(x) \, dx \leq \int_{-\infty}^{\infty} \exp \left( c^2 t^2 (\lambda - 1) / 2 \right) \, dt \]
\[
= \frac{1}{c} \sqrt{2\pi\lambda} \int_{-\infty}^{\infty} \exp \left( \lambda x^2 / 2c^2 \right) f(x) \, dx \leq \frac{1}{c} \sqrt{\frac{2\pi\lambda}{1 - \lambda}}. \tag{A.85}
\]

Theorem 9 ([16, Theorem 7.2]). Suppose that \( x = [X_1, X_2, \ldots, X_N] \), where each \( X_i \) is i.i.d. with \( X_i \sim \text{Sub} (c^2) \) and \( \mathbb{E} (X_i^2) = \sigma^2 \). Then
\[
\mathbb{E} (\|x\|_2^2) = M\sigma^2. \tag{A.86}
\]
In addition, for any \( \alpha \in (0, 1) \) and for any \( \beta \in [c^2 / \sigma^2, \beta_{\max}] \), there exists a constant \( \mu \geq 4 \) depending only on \( \beta_{\max} \) and the ratio \( \sigma^2 / c^2 \) such that
\[
\mathbb{P} (\|x\|_2^2 \leq \alpha M\sigma^2) \leq \exp (-M (1 - \alpha)^2 / \mu) \tag{A.87}
\]
and
\[
\mathbb{P} (\|x\|_2^2 \geq \beta M\sigma^2) \leq \exp (-M (\beta - 1)^2 / \mu). \tag{A.88}
\]

Proof. Since \( X_i \) are independent,
\[
\mathbb{E} (\|x\|_2^2) = \sum_{i=1}^{M} \mathbb{E} (X_i^2) = \sum_{i=1}^{M} \sigma^2 = M\sigma^2, \tag{A.89}
\]
therefore, (A.86) holds. From Lemma 4 we have
\[
\mathbb{P} (\|x\|_2^2 \geq \beta M\sigma^2) = \mathbb{P} \left( \exp \left( \lambda \|x\|_2^2 \right) \geq \exp \left( \lambda \beta M\sigma^2 \right) \right)
\]
\[
\leq \frac{\mathbb{E} (\exp \left( \lambda \|x\|_2^2 \right))}{\exp \left( \lambda \beta M\sigma^2 \right)} = \prod_{i=1}^{M} \frac{\mathbb{E} (\exp \left( \lambda X_i^2 \right))}{\exp \left( \lambda \beta M\sigma^2 \right)}. \tag{A.90}
\]
From Lemma 5 we have
\[ E \left( \exp \left( \lambda X_i^2 \right) \right) = E \left( \exp \left( 2c^2 \lambda X_i^2 / 2c^2 \right) \right) \leq \frac{1}{\sqrt{1 - 2c^2} \lambda}. \] (A.91)

Therefore,
\[ \prod_{i=1}^{M} E \left( \exp \left( \lambda X_i^2 \right) \right) \leq \left( \frac{1}{1 - 2c^2 \lambda} \right)^{M/2} \] (A.92)
and hence
\[ P \left( \|x\|_2^2 \geq \beta M\sigma^2 \right) \leq \left( \frac{\exp \left( -2\lambda \beta \sigma^2 \right)}{1 - 2c^2 \lambda} \right)^{M/2}. \] (A.93)

The optimal \( \lambda \), which can be found by setting the derivative to zero and solving for \( \lambda \), is
\[ \lambda = \frac{\beta \sigma^2 - c^2}{2c^2 \sigma^2 (1 + \beta)}. \] (A.94)

Substituting the optimal \( \lambda \) in (A.93) gives
\[ P \left( \|x\|_2^2 \geq \beta M\sigma^2 \right) \leq \left( \frac{\exp \left( -2\lambda \beta \sigma^2 \right)}{1 - 2c^2 \lambda} \right)^{M/2}. \] (A.95)

Similarly for \( \alpha \) we have
\[ P \left( \|x\|_2^2 \geq \alpha M\sigma^2 \right) \leq \left( \frac{\exp \left( -2\lambda \beta \sigma^2 \right)}{1 - 2c^2 \lambda} \right)^{M/2}. \] (A.96)

If we define
\[ \mu = \max \left( 4, 2 \frac{(\beta_{\max}\sigma^2/c - 1)^2}{(\beta_{\max}\sigma^2/c - 1) - \log (\beta_{\max}\sigma^2/c)} \right) \] (A.97)
then we have that for any \( \gamma \in [0, \beta_{\max}\sigma^2/c] \) we have the bound
\[ \log (\gamma) \leq (\gamma - 1) - \frac{2(\gamma - 1)^2}{\mu}, \] (A.98)
and therefore
\[ \gamma \leq \exp \left( (\gamma - 1) - \frac{2(\gamma - 1)^2}{\mu} \right). \] (A.99)

By setting \( \gamma = \alpha \sigma^2/c^2 \), we can obtain (A.87). In the same way, setting \( \gamma = \beta \sigma^2/c^2 \) gives (A.88).
**Corollary 1** ([16 Corollary 7.1]). Suppose that \( x = [X_1, X_2, \ldots, X_N] \), where each \( X_i \) is i.i.d. with \( X_i \sim \text{SSub}(\sigma^2) \). Then

\[
\mathbb{E} \left( \|x\|_2^2 \right) = M\sigma^2 \tag{A.100}
\]

and for any \( \epsilon > 0 \),

\[
\mathbb{P} \left( \|x\|_2^2 - M\sigma^2 \geq \epsilon M\sigma^2 \right) \leq 2 \exp \left( -\frac{M\epsilon^2}{\mu} \right) \tag{A.101}
\]

with \( \mu = 2/(1 - \log(2)) \approx 6.52 \).

**Proof.** Since \( X_i \sim \text{SSub}(\sigma^2) \), we have that \( X_i \sim \text{Sub}(\sigma^2) \) and by setting \( \alpha = 1 - \epsilon \) and \( \beta = 1 + \epsilon \) in Theorem 9 we obtain (A.101) with \( \beta_{\text{max}} = 2 \). By substituting \( c^2 = \sigma^2 \) and \( \beta_{\text{max}} = 2 \) in (A.97) we obtain \( \mu = 2/(1 - \log(2)) \).

**Corollary 2** ([16 Corollary 7.2]). Suppose that \( \Phi \in \mathbb{R}^{M \times N} \) whose modulus entries \( \phi_{ij} \) are i.i.d. with \( \phi_{ij} \sim \text{SSub}(1/M) \). Let \( y = \Phi x \) for any \( x \in \mathbb{R}^N \). Then for any \( \epsilon > 0 \),

\[
\mathbb{E} \left( \|y\|_2^2 \right) = \|x\|_2^2 \tag{A.102}
\]

and

\[
\mathbb{P} \left( \|y\|_2^2 - \|x\|_2^2 \geq \epsilon \|x\|_2^2 \right) \leq 2 \exp \left( -\frac{M\epsilon^2}{\mu} \right) \tag{A.103}
\]

with \( \mu = 2/(1 - \log(2)) \approx 6.52 \).

**Proof.** Let \( \phi_{ij} \) be the \( i \)-th row of \( \Phi \). Then \( i \)-th entry of \( y \) can be written as \( [y]_i = \langle \phi_{ij}, x \rangle \), and by applying Lemma 3 we have \( Y_i \sim \text{SSub}(\|x\|_2^2/M) \). The result follows by using Corollary 1 for \( y = [Y_1, Y_2, \ldots, Y_N] \).

**Lemma 6** ([16 Lemma 7.5]). Let \( \epsilon \in (0, 1) \). There exists a set of points \( Q \) such that \( |Q| \leq (3/\epsilon)^K \) and for any \( x \in \mathbb{R}^K \) with \( \|x\|_2 \leq 1 \) there is a point \( q \in Q \) which satisfies \( \|x - q\|_2 \leq \epsilon \).

**Proof.** We start adding arbitrary \( q_i \in \mathbb{R}^K \) to \( Q \) such that \( i = 1, 2, \ldots, l \), \( \|q_i\|_2 \leq 1 \) and \( \|q_i - q_j\|_2 > \epsilon \) for all \( i > j \) until we can add no more points (\( q_{i+1} \)) to \( Q \), where \( l = |Q| \). Therefore, for any \( x \in \mathbb{R}^K \) with \( \|x\|_2 \leq 1 \) there is a \( q \in Q \) that satisfies \( \|x - q\|_2 \leq \epsilon \). By centering balls of radius \( \epsilon/2 \) at each \( q_i \) the balls are disjoint and within a ball of radius \( 1 + \epsilon/2 \). If \( B^K(r) \) denotes a ball of radius \( r \) in \( \mathbb{R}^K \), then

\[
|Q| \text{Vol} \left( B^K(\epsilon/2) \right) \leq \text{Vol} \left( B^K(1 + \epsilon/2) \right) \tag{A.104}
\]
and therefore
\[ |Q| \leq \frac{\text{Vol}(B^K(1 + \epsilon/2))}{\text{Vol}(B^K(\epsilon/2))} = \frac{(1 + \epsilon/2)^K}{(\epsilon/2)^K} \leq (3/\epsilon)^K. \] (A.105)

Now, we are ready to prove Theorem 4.

Proof. Without losing generality it is enough to prove (2.5) in the case \( \|x\|_2 = 1 \). We define \( X_T \) as \( s \)-dimensional subspace of \( \Phi_{[\cdot,T]} \), \( T \subset [N] \) and \( |T| = s \). We choose finite set of points \( Q_T \) such that \( Q_T \subseteq X_T \), \( \|q\|_2 \leq 1 \) for all \( q \in Q_T \), and for all \( x \in Q_T \) with \( \|x\|_2 \leq 1 \) we have
\[ \min_{q \in Q_T} \|x - q\|_2 \leq \delta_s/14. \] (A.106)

Here, we have chosen particular value \( \delta_s/14 \) which makes the proof easier, however, it can be replaced by any smaller arbitrary value. From Lemma 6, we know that \( |Q_T| \leq (42/\delta_s)^s \). By collecting all points \( Q_T \) with possible sets of \( T \) together we obtain:
\[ Q = \bigcup_{T:|T|=s} Q_T. \] (A.107)

There are \( \binom{N}{s} \) possible index sets \( T \). From stirling’s approximation we have \( s! \approx \sqrt{2\pi s} \left( \frac{s}{e} \right)^s \), and we have
\[ \binom{N}{s} \leq \frac{N^s}{s!} \approx \frac{1}{\sqrt{2\pi s}} \left( \frac{eN}{s} \right)^s < \left( \frac{eN}{s} \right)^s. \] (A.108)

Therefore,
\[ |Q| \leq (42eN/\delta_s s)^s. \] (A.109)

From Corollary 1 we have (A.101). Hence
\[ \mathbb{P} \left( \| \Phi q \|_2^2 - \| q \|_2^2 \geq \epsilon \| q \|_2^2 \right) \leq 2 \exp \left( -\frac{Me^2}{\mu} \right), \text{ for all } q \in Q_T. \] (A.110)

By applying (A.109) we obtain
\[ \mathbb{P} \left( \| \Phi q \|_2^2 - \| q \|_2^2 \geq \epsilon \| q \|_2^2 \right) \leq 2 \left( \frac{42eN}{\delta_s s} \right)^s \exp \left( -\frac{Me^2}{\mu} \right), \text{ for all } q \in Q \] (A.111)

\[ \mathbb{P} \left( \| \Phi q \|_2^2 - \| q \|_2^2 \leq \epsilon \| q \|_2^2 \right) \geq 1 - 2 \left( \frac{42eN}{\delta_s s} \right)^s \exp \left( -\frac{Me^2}{\mu} \right), \text{ for all } q \in Q. \] (A.112)
Setting \( \epsilon = \delta_s/\sqrt{2} \) in (A.112) gives
\[
(1 - \delta_s/\sqrt{2}) \|q\|_2^2 \leq \|\Phi q\|_2^2 \leq (1 + \delta_s/\sqrt{2}) \|q\|_2^2, \text{ for all } q \in Q \tag{A.113}
\]
with probability greater than
\[
1 - 2 (42eN/\delta_s)^s e^{-M\delta_s^2/2\mu}. \tag{A.114}
\]
We observe that if \( M \) satisfies (2.10) then
\[
\log \left( \frac{42eN}{\delta_s^s} \right)^s \leq s \log \left( \frac{N}{s} \right) \log \left( \frac{42e}{\delta_s^s} \right) \leq M \log \left( \frac{42e/\delta_s}{\alpha} \right) \tag{A.115}
\]
and (A.114) will change to \( 1 - 2e^{-\beta M} \) as desired.
Now we define \( \theta \) as the smallest number such that
\[
\|\Phi x\|_2 \leq \sqrt{1 + \theta} \|x\|_2 \text{ for all } x \in X_T, \|x\|_2 \leq 1. \tag{A.116}
\]
If we show that \( \theta \leq \delta_s \) then the proof is completed. We pick \( q \in Q_T \subset X_T \) and for any \( x \in X_T \) we have \( \|x - q\|_2 \leq \delta_s/14. \) From (A.113) and (A.116) we have that
\[
\|\Phi x\|_2 \leq \|\Phi q\|_2 + \|\Phi (x - q)\|_2 \leq \sqrt{1 + \delta_s/\sqrt{2} + \delta_s/14\sqrt{1 + \theta}}. \tag{A.117}
\]
By definition \( \theta \) is the smallest number for which (A.116) holds, therefore,
\[
\sqrt{1 + \theta} \leq \sqrt{1 + \delta_s/\sqrt{2} + \delta_s/14\sqrt{1 + \theta}} \tag{A.118}
\]
and
\[
\sqrt{1 + \theta} \leq \frac{\sqrt{1 + \delta_s^2/2}}{1 - \delta_s/14} \leq \sqrt{1 + \delta_s}. \tag{A.119}
\]
We have proved the upper inequality in (2.5). Similarly, for lower bound we have
\[
\|\Phi x\|_2 \geq \|\Phi q\|_2 - \|\Phi (q - x)\|_2 \geq \sqrt{1 - \delta_s/\sqrt{2} - \delta_s/14\sqrt{1 - \theta}} \geq \sqrt{1 - \delta_s}, \tag{A.120}
\]
which completes the proof.

\( \square \)
Appendix B

A model for the measurement noise in 1-bit compressive sensing

In the case that there is measurement noise we have

$$\text{sign} (\Phi x + n) = \tilde{b}_m, \tag{B.1}$$

where $n \in \mathbb{R}^M$ is the measurement noise and $[n]_i \sim N(0, \sigma^2_n)$ for all $i$. First, we find the distribution of $[\Phi x]_i$. Each element of $\Phi x$ is the summation of $s$ multiplication pairs of $\phi_{i,j} \sim N(0, \sigma^2 \phi)$ and $[x]_j \sim N(0, \sigma^2 x)$. In [26, section 6.A], the distribution of multiplication of two independent $s$ dimension Gaussian vectors has been calculated. Recall that $s$ is the sparsity of $x$, $\sigma^2_x = 1$ and $\sigma^2 \phi = 1/M$ where $M$ is the number of rows in $\Phi$. We are looking for the probability of the bit flip in each bit of $\tilde{b}_m$ which is denoted by $P_{fm}$. Therefore,

$$P_{fm} = \mathbb{P} \left( \text{sign} ([\Phi x + n]_i) \neq \text{sign} ([\Phi x]_i) \right). \tag{B.2}$$

We assume that $[n]_i = n$, $[\Phi x]_i = \theta$ and $\theta > 0$ then we have

$$P_{fm} = 2 \int_0^\infty p_N (n < -\theta) p_\theta (\theta) d\theta$$

$$= 2 \int_0^\infty \left( 1 - Q \left( \frac{\theta}{\sigma_n^2} \right) \right) p_\theta (\theta) d\theta, \tag{B.3}$$

where $p_\theta (\theta)$ is shown in [26, section 6.A]. Intuitively, when $\sigma^2_n = 0$, $P_{fm} = 0$ and as $\sigma^2_n$ tends to infinity $P_{fm}$ converges to $\frac{1}{2}$. 

59
Bibliography


