A Friendly Introduction to Compressed Sensing

Lawrence J. Zhang
Advisor: Ery Arias-Castro
University of California San Diego
An Undergraduate Honors Thesis

June 7, 2021

Abstract

Compared to other signal processing techniques, compressed sensing (or sparse sampling) has caught the interest of many mathematicians, electrical engineers, and computer scientists. The field of compressed sensing is still rapidly evolving. As most papers and textbooks about compressed sensing are at graduate level, the purpose of this paper is to offer a gentler exposure to compressed sensing from a mathematical perspective. By synthesizing my study on compressed sensing as an undergraduate, this thesis covers important concepts in CS such as coherence and restricted isometry property. Several key algorithms in compressed sensing will also be introduced with discussions of their stability, robustness, and performance. In the end, we investigate single-pixel camera as an example of real-world application of compressed sensing.

Contents

1 Introduction

The first part of this paper introduces general ideas about compressed sensing and the motivation that drives the development of such signal recovery techniques. This thesis is motivated by Simon Foucart's book *A Mathematical Introduction to Compressed Sensing*, which is also frequently referenced throughout this thesis.

1.1 Background

Before introducing compressed sensing, it is important to recognize that compressed sensing is a sub-field of signal processing. In the field signal processing, one of the major research interests is the reconstruction of signal from different measurements. To mathematically formulate the signal recovery problem, we define $\mathbf{y} \in \mathbb{C}^N$ as the observed data, and it is associated with a corresponding signal $\mathbf{x} \in \mathbb{C}^N$. The observed data and its corresponding signal are connected via the measurement matrix $\mathbf{A} \in \mathbb{C}^{m\times N}$ as follows

$$\mathbf{A}\mathbf{x} = \mathbf{y}, \tag{1}$$

where N is the signal length and m is the number of measurements. Now it is obvious that the signal recovery problem concerns about solving the linear system above with respect to the signal $\mathbf{x} \in \mathbb{C}^N$. However, in the case of compressed sensing, when $m < N$, the linear system is underdetermined and solving the linear system becomes impossible if there is no additional information available regarding the measurement matrix $\mathbf{A} \in \mathbb{C}^{m\times N}$. For the sake of the viability of compressed sensing, the measurement matrix itself requires careful designs under specific criteria. As a result of the Shannon sampling theorem, traditional signal reconstruction requires the sampling rate of a continuous-time signal to be twice its highest frequency. This fact will be elaborated in next section, where we take a careful look into sampling Theory. By exploiting the sparsity of a signal and the coherence of a measurement matrix, it takes much fewer measurements to achieve signal recovery via compressed sensing.

One of the key concepts in compressed sensing is sparsity, $\|\mathbf{x}\|_0$ which denotes the number of nonzero entries in the vector $\mathbf{x}$. The recovery algorithms are also essential to compressed sensing. With the concept of sparsity, one of the first algorithmic attempts ever made is ℓ_0-minimization. Conceptually, the goal of ℓ_0-minimization is to reconstruct the signal $\mathbf{x}$ by solving the following optimization problem:

$$\text{minimize } \|\mathbf{z}\|_0 \qquad \text{subject to } \mathbf{Az} = \mathbf{y}. \tag{2}$$

However, ℓ_0-minimization is NP-hard, meaning this problem is difficult to solve. In later chapter, the NP-hardness of ℓ_0-minimization will be derived when algorithms are specifically introduced. Today, the most popular compressed sensing algorithm is ℓ_1-minimization, or basis pursuit, is as following:

$$\text{minimize } \|\mathbf{z}\|_1 \qquad \text{subject to } \mathbf{Az} = \mathbf{y}. \tag{3}$$

Because of the convexity of ℓ_1-norm $\|\cdot\|_1$, the basis pursuit method can be conveniently solved with methods from convex optimization. Besides optimization methods, there also exist alternative reconstruction methods such as greedy methods and threshholding-based methods, which will also be lightly introduced later.

1.2 Sampling Theory

The applications of signal reconstruction are omnipresent in both scientific and technological fields. In the typical situation, we seek to reconstruct a continuous-time signal from a discrete set of sample measurements. Radio frequency(RF) and analog-to-digital(ADC) technologies are an important example. Shannon-Nyquist sampling theorem lays the mathematical foundation and dictates the rates of high-bandwidth signals for most traditional signal reconstruction technologies. Nonetheless, as a nontraditional signal reconstruction technique, compressed sensing breaks free from the sample number restriction by exploiting factors like sparsity and compressibilty. In this section, a comparison will be made between Shannon-Nyquist sampling theorem and the general idea of compressed sensing.

Shannon-Nyquist sampling theorem states that to ensure the reconstruction of a function of bandwith B, a sampling at the rate $2B$ is required. The Fourier transform of a continuous-time signal $f \in L^1(\mathbb{R})$ is defined as:

$$\hat{f}(\xi) = \int_{\mathbb{R}} f(t) e^{-2\pi i t \xi} dt, \quad \xi \in \mathbb{R}. \tag{4}$$

If $\hat{f}$ is supported in $[-B, B]$, we say f has a bandwidth of B. Shannon-Nyquist sampling theorem states that for a function f with bandwidth B, it can be reconstructed from its discrete set of samples $f(k/(2B)), k \in \mathbb{Z}$ by

$$f(t) = \sum_{k \in \mathbb{Z}} f(\frac{k}{2B}) sinc(2\pi Bt - \pi k), \tag{5}$$

where the sinc function is

$$sinc(t) = \begin{cases} \frac{\sin t}{t} & \text{if } t \neq 0, \\ 1 & \text{if } t = 0. \end{cases} \tag{6}$$

For the sake of comparison to compressed sensing, we consider Shannon-Nyquist sampling theorem in some finite dimensional space. In this case, we consider the trigonometric polynomial with a maximal degree M such that

$$f(t) = \sum_{k=-M}^{M} x_k e^{2\pi ikt}, \quad t \in [0, 1], \tag{7}$$

where M is a substitute to the bandwidth B. Note the space of trigonometric polynomials of maximal degree M has dimension of $N = 2M + 1$, f can be reconstructed if there are $N = 2M + 1$ samples. The finite-dimensional Shannon-Nyquist sampling theorem states that

$$f(t) = \frac{1}{2M+1} \sum_{k=0}^{2M} f\left(\frac{k}{2M+1}\right) D_M\left(t - \frac{k}{2M+1}\right), \quad t \in [0, 1], \tag{8}$$

where the Dirchlet kernel D_M is

$$D_M(t) = \sum_{k=-M}^{M} e^{2\pi ikt} = \begin{cases} \frac{\sin(\pi(2M+1)t)}{\sin(\pi t)} & \text{if } t \neq 0, \\ 2M+1 & \text{if } t = 0. \end{cases} \tag{9}$$

By the finite-dimensional Shannon-Nyquist sampling theorem, it is impossible to reconstruct such trigonometric polynomials with maximal degree M if the number of samples is less than $N = 2M + 1$. In realistic situations, such required number of samples are sometimes too large and computationally infeasible. However, if the vector $\mathbf{x} \in \mathbb{C}^N$ of Fourier coefficients of f is sparse or compressible, fewer samples will be required to produce exact signal recovery if these properties are properly exploited. In mathematical language, given a set $\{t_1, ..., t_m\} \in [0, 1]$ of m sampling points denoting the sparsity, the observed data vector $\mathbf{y} = (f(t_\ell))_{\ell=1}^m$ can be written as

$$\mathbf{y} = \mathbf{A}\mathbf{x} \tag{10}$$

where $\mathbf{A} \in \mathbb{C}^{m \times N}$ is is a Fourier-type matrix with entries

$$A_{\ell,k} = e^{2\pi i k t_\ell}, \quad l = 1, ..., m, \quad k = -M, ...M. \tag{11}$$

In this setting, to recover f from the vector $\mathbf{y} = (f(t_\ell))_{\ell=1}^m$, we only need to find the coefficient vector $\mathbf{x}$. And note when $m < N$, the linear system becomes underdetermined. Following the assumption that the coefficient vector $\mathbf{x}$ is sparse, we get our standard compressed sensing problem.

2 Sparse solutions of Underdetermined Systems

The second part of this paper defines some important notation like vector sparsity and compressibility that will be ubiquitously used in later chapters. In the end, a proof for the NP-hardness of ℓ_0-minimization will be provided, and this will help us navigate to the following chapter, where we discuss common compressed sensing algorithms.

2.1 Sparsity and Compressibility

Sparsity is one of the key assumptions in compressed sensing. In this section, we will rigorously define sparsity and compressibility and introduce several important theorems that are necessary to establish algorithms of compressed sensing. Let $[N]$ denote the set $\{1, 2, ..., N\}$. Let S denote a set in $[N]$, then $\bar{S}$ is the complementary set $[N] \setminus S$. Then the support of a vector $\mathbf{x} \in \mathbb{C}^N$ is the index set of its nonzero entries such that

$$\text{supp}(\mathbf{x}) := \{j \in [N] : x_j \neq 0\}. \tag{12}$$

The sparsity of $\mathbf{x} \in \mathbb{C}^N$ is equivalent to the cardinality of the support of $\mathbf{x}$. The vector $\mathbf{x}$ is s-sparse if at most s of its entries are nonzero as following:

$$\|\mathbf{x}\|_0 := \text{card}(\text{supp}(\mathbf{x})) \leq s. \tag{13}$$

The sparsity of $\mathbf{x}$, $\|\mathbf{x}\|_0$ may look nonsensical because "zero norm" does not make any mathematical sense, it is in fact the customary notation for sparsity. Unlike the sparsity which offers a strong constraint to the recovery problem, the concept of compressibility is more useful because it is relatively weaker, and this makes it more versatile in practice. In this case, we consider vectors that are near s-sparse, which are measured by the *error of best s-term approximation.*

For $p > 0$, *ℓ_p-error of best s-term approximation* to a vector $\mathbf{x} \in \mathbb{C}^N$ is defined by

$$\sigma_s(\mathbf{x})_p := inf\{\|\mathbf{x} - \mathbf{z}\|_p, \quad \mathbf{z} \in \mathbb{C}^N \text{is } s\text{-sparse}\}. \tag{14}$$

As we have defined $\sigma_s(\mathbf{x})_p$, the infimum is achieved by an s-sparse vector $\mathbf{z} \in \mathbb{C}^N$ whose nonzero entries equal the s largest absolute entries of x.

The vector $\mathbf{x} \in \mathbb{C}^N$ may be informally called a compressible vector if the error of its best s-term approximation quickly converges to 0 as s approaches N. If $\mathbf{x}$ belongs to the unit ℓ_p-ball for some small $p > 0$, where the unit ℓ_p-ball is defined by

$$B_p^N := \{\mathbf{z} \in \mathbb{C} : \|\mathbf{z}\|_p \leq 1\}. \tag{15}$$

For any $q > p > 0$ and any $\mathbf{x} \in \mathbb{C}^N$,

$$\sigma_s(\mathbf{x})_p \leq \frac{1}{s^{1/p-1/q}} \|x\|_p. \tag{16}$$

To prove this important inequality, we first define the nonincreasing rearrangement of $\mathbf{x} \in \mathbb{C}^N$ is the $\mathbf{x}^* \in \mathbb{R}^N$ for which

$$x_1^* \geq x_2^* \geq ... \geq x_N^* \geq 0 \tag{17}$$

and there is a permutation $\pi : [N] \to [N]$ with $x_j^* = |x_{\pi(j)}|$ for all $j \in [N]$. The nonincreasing rearrangement of a vector satisfies, for $\mathbf{x}, \mathbf{z} \in \mathbb{C}^N$,

$$\|\mathbf{x}^* - \mathbf{z}^*\|_\infty \leq \|\mathbf{x} - \mathbf{z}\|_\infty. \tag{18}$$

Now with the definition of nonincreasing rearrangement of a vector. We can give a proof to inequality (16). We first assume $\mathbf{x}^* \in \mathbb{R}_+^N$ is the nonincreasing arrangement of $\mathbf{x} \in \mathbb{C}^N$, then we get

$$\sigma_s(\mathbf{x})_q^q = \sum_{j=s+1}^{N} (x_j^*)^q \leq (x_s^*)^{q-p} \sum_{j=s+1}^{N} (x_j^*)^p \leq \left(\frac{1}{s} \sum_{j=1}^{s} (x_j^*)^p \right)^{(q-p)/p} \left(\sum_{j=s+1}^{N} (x_j^*)^p \right) \tag{19}$$

$$\leq \left(\frac{1}{s} \|\mathbf{x}\|_p^p \right)^{(q-p)/p} = \frac{1}{s^{1/p-1/q}} \|x\|_p^q.$$

Now we take the q-th root on each side of the inequality, we get

$$\sigma_s(\mathbf{x})_p \leq \frac{1}{s^{1/p-1/q}} \|x\|_p. \tag{20}$$

2.2 Minimal Number of Measurements

From the first chapter, we have formalized the compressed sensing problem, which is essentially to reconstruct an s-sparse vector $\mathbf{x} \in \mathbb{C}^N$ from

$$\mathbf{y} = \mathbf{A}\mathbf{x} \tag{21}$$

where $\mathbf{A} \in \mathbb{C}^{m \times N}$ is our measurement matrix for $m < N$. Although the linear system is underdetermined, the sparsity assumption imposes a strong condition in recovering the vector $\mathbf{x}$.

This begs the question: what is the minimal number of measurement needed to reconstructed s-sparse vectors? When we reconstruct the sparse vectors, we want to minimize their sparsity s, there are two distinct cases we are concerning about:

1. Uniform Recovery: recovery of All Sparse Vectors Simultaneously.
2. Nonuniform Recovery: recovery of Individual Sparse Vectors.

Before we discuss each case separately, notice that the following problems are equivalent given sparsity s, matrix $\mathbf{A} \in \mathbb{C}^{m \times N}$, and s-sparse $\mathbf{x} \in \mathbb{C}^N$:

(a) The vector $\mathbf{x}$ is the unique s-sparse solution of $\mathbf{A}\mathbf{z} = \mathbf{y}$ with $\mathbf{y} = \mathbf{A}\mathbf{x}$, that is, $\{\mathbf{z} \in \mathbb{C}^N : \mathbf{A}\mathbf{z} = \mathbf{A}\mathbf{x}, \|\mathbf{z}\|_0 \leq s\} = \{x\}$.
(b) The vector $\mathbf{x}$ can be reconstructed as the unique solution of

$$\underset{\mathbf{z} \in \mathbb{C}^N}{\text{minimize}} \ \|\mathbf{z}\|_0 \qquad \text{subject to } \mathbf{A}\mathbf{z} = \mathbf{y}. \tag{22}$$

Let $\mathbf{A}_S$ denote the submatrix of $\mathbf{A}$ with columns indexed by $S \subset [N]$. And similarly, $\mathbf{x}_S \in \mathbb{C}^S$ is the sub-vector where the non-empty entries are indexed by S. With these notations, a theorem can be introduced:
Given $\mathbf{A} \in \mathbb{C}^{m \times N}$, the following properties are equivalent:
(a) Every s-sparse vector $\mathbf{x} \in \mathbb{C}^N$ is the unique s-sparse solution of $\mathbf{A}\mathbf{z} = \mathbf{A}\mathbf{x}$, that is, if $\mathbf{A}\mathbf{z} = \mathbf{A}\mathbf{x}$ and both $\mathbf{x}$ and $\mathbf{z}$ are s-sparse, then $\mathbf{x} = \mathbf{z}$.
(b) The null space ker $\mathbf{A}$ does not contain any $2s$-sparse vector other than the zero vector, that is, $\ker A \cap \{\mathbf{z} \in \mathbb{C}^N : \|\mathbf{z}\|_0 \leq 2s\} = \{\mathbf{0}\}$.
(c) For every $S \subset [N]$ with $\text{card}(S) < 2s$, the submatrix $\mathbf{A}_S$ is injective as a map from $\mathbb{C}^S$ to $\mathbb{C}^m$.
(d) Every set of $2s$ columns of $\mathbf{A}$ is linearly independent.

Note if the reconstruction of all s-sparse vectors are possible based on the measurement matrix $\mathbf{A} \in \mathbb{C}^{m \times N}$ and measurement vector $\mathbf{y} \in \mathbb{C}^N$, the (a) and (d) automatically hold true. In turn, this implies

$$m \geq rank(\mathbf{A}) \geq 2s. \tag{23}$$

And the following theorem is introduced:
For any integer $N \geq 2s$, there exists a measurement matrix $\mathbf{A} \in \mathbb{C}^{m \times N}$ with $m = 2s$ rows such that every s-sparse vector $\mathbf{x} \in \mathbb{C}^N$ can be recovered from its measurement vector $\mathbf{y} \in \mathbb{C}^m$ as a unique solution of

$$\underset{\mathbf{z} \in \mathbb{C}^N}{\text{minimize}} \, \|\mathbf{z}\|_0 \qquad \text{subject to } \mathbf{Az} = \mathbf{y}. \tag{24}$$

And for any $N \geq 2s$, there exists a practical procedure for the reconstruction of every $2s$-sparse vectors from its first $m = 2s$ discrete Fourier measurements.

Although the reconstruction procedure described is seemingly well-designed, its insufficiency in stability and robustness causes major drawbacks. For the recovery of individual sparse vectors (Non-uniform recovery), we have our s-sparse vector $\mathbf{x} \in \mathbb{C}^N$ fixed prior to choosing the measurement matrix $\mathbf{A}$. Then we take measurement accordingly. And we want $\mathbf{x}$ to be a unique solution, where such results require conditions depend on both $\mathbf{x}$ itself and the choice of $\mathbf{A}$. The core idea behind nonuniform recovery is that the required condition will be satisfied by most $(s+1) \times N$ matrices, where $\mathbf{A}$ is randomly chosen. In conclusion, for any $N \geq (s+1)$, given an s-sparse vector $\mathbf{x} \in \mathbb{C}^N$ there exists a measurement matrix $\mathbf{A} \in \mathbb{C}^{m \times N}$ with $m = s+1$ rows such that the vector $\mathbf{x}$ be reconstructed from its measurement vector as a solution of $\mathbf{y} = \mathbf{Ax} \in \mathbb{C}^N$:

$$\underset{\mathbf{z} \in \mathbb{C}^N}{\text{minimize}} \, \|\mathbf{z}\|_0 \qquad \text{subject to } \mathbf{Az} = \mathbf{y}. \tag{25}$$

2.3 NP-Hardness of ℓ_0-Minimization

In previous sections, without taking the viability into account, we have conveniently formulated the reconstruction of an s-sparse vector $\mathbf{x} \in \mathbb{C}^N$ as the ℓ_0-Minimization problem

$$\underset{\mathbf{z}\in\mathbb{C}^N}{\text{minimize}} \ \|\mathbf{z}\|_0 \qquad \text{subject to } \mathbf{Az} = \mathbf{y}, \tag{26}$$

where we achieve the signal recovery goal by minimizing the sparsity of $\mathbf{z}$. ℓ_0-norm based signal recovery is attractive in compressed sensing as it can facilitate exact recovery of sparse signal. Unfortunately, direct ℓ_0-norm minimization problem is NP-hard [**?**]. In this section, we will demonstrate a more generalized (noise-aware) version of ℓ_0-minimization, that is, for any $\eta \geq 0$, the ℓ_0-minimization

$$\underset{\mathbf{z}\in\mathbb{C}^N}{\text{minimize}} \ \|\mathbf{z}\|_0 \qquad \text{subject to } \|\mathbf{Az} - \mathbf{y}\|_2 \leq \eta. \tag{27}$$

for general $\mathbf{A} \in \mathbb{C}^{m\times N}$ and $\mathbf{y} \in \mathbb{C}^m$ is NP-hard.

By definition, NP-hard problems consist of all problems for a solving algorithm for any NP-problem. In simple language, the NP-hard problem is at least as hard as any NP-problem. And the complexity class NP-Complete contains all problems from both NP and NP-hard class. To prove the the statement above, we first consider a NP-complete problem, the *exact cover by 3-sets problem* as follows:
Given a collection $\{C_i, i \in [N]\}$ of 3-element subsets of $[m]$, does there exist an exact cover of [m] a set $J \subset [N]$ such that $\cup_{j\in J} C_j = [m]$ and $C_j \cup C_{j'} = \emptyset$ for all $j, j' \in J$ with $j \neq j'$?

To prove the NP-hardness of ℓ_0-minimization, we are going to show that solving ℓ_0-minimization implies a solution to the exact cover by 3-sets problem. For (29), we assume $\eta < 1$. Let $\{C_i, i \in [N]\}$ be a collection of 3-element subsets of $[m]$. We first define a vector $\mathbf{y} \in \mathbb{C}^m$ as $\mathbf{y} = [1, 1, ..., 1]^{\mathrm{T}}$. For our matrix $\mathbf{A} \in \mathbb{C}^{m\times N}$, we define its columns $\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_N \in \mathbb{C}^m$ by

$$(\mathbf{a}_i)_j = \begin{cases} 1 \text{ if } j \in C_i, \\ 0 \text{ if } j \notin C_i. \end{cases} \tag{28}$$

And $\mathbf{A} \in \mathbb{C}^{m\times N}$ is

$$\begin{bmatrix} | & | & & | \\ \mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_N \\ | & | & & | \end{bmatrix}.$$

Note $N \leq \binom{m}{3}$, so the construction of matrix $\mathbf{A} \in \mathbb{C}^{m\times N}$ is viable in polynomial time. If there exists vector $\mathbf{z} \in \mathbb{C}^N$ such that $\|\mathbf{Az} - \mathbf{y}\|_2 \leq \eta$, then each of the m components has at most a distance η to 1, and the sparsity is m as $\|\mathbf{Az}\|_0 = m$. But as we previously defined, each vector $\mathbf{a}_i$ has exactly 3 nonzero components, and the vector $\mathbf{Az} = \sum_{j=1}^{N} z_j \mathbf{a_j} = 1$ has at most $3\|\mathbf{z}\|_0$ nonzero components, meaning there is an inequality of sparsity $\|\mathbf{Az}\|_0 \leq 3\|\mathbf{z}\|_0$. Thus, if such $\mathbf{z} \in \mathbb{C}^N$ exists, it must satisfy $\|\mathbf{z}\|_0 \geq m/3$. Now let $\mathbf{x} \in \mathbb{C}^N$ denote the output of our ℓ_0-minimization problem, then there are two cases to consider for $\|\mathbf{z}\|_0 \geq m/3$:

1. If $\|\mathbf{z}\|_0 = m/3$, then the collection $C_j, j \in \text{supp}(\mathbf{x})$ forms an exact cover of $[m]$. Otherwise the sparsity of $\mathbf{Ax} = \mathbf{Ax} = \sum_{j=1}^{N} x_j \mathbf{a_j}$ would not be m.

2. If $\|\mathbf{z}\|_0 > m/3$, then there does not exist exact cover $C_j, j \in J$. Otherwise the vector $\mathbf{z} \in \mathbb{C}^N$ defined by $z_j = 1$ if $j \in J$ and $z_j = 0$ if $j \notin J$ would satisfy $\mathbf{y} = \mathbf{Az}$ and $\|\mathbf{z}\|_0 = m/3$, and this contradicts the ℓ_0-minimization of $\mathbf{x}$.

And this verifies the ℓ_0-minimization is NP-hard because solving ℓ_0-minimization also implies a solution to the exact cover by 3-sets problem, which is NP-hard.

Although the NP-hardness of ℓ_0-minimization looks discouraging, it only restricts the tractability of recovery algorithms using any choice of matrices $\mathbf{A}$ and vectors $\mathbf{y}$. In compressed sensing, we only concern about specially designed measurement matrices $\mathbf{A}$ and $\mathbf{y}$ for some sparse signal $\mathbf{x}$. Many recovery algorithms for such matrices will be introduced in the following chapter.

3 Algorithms of Compressed Sensing

In this chapter we introduce three types of algorithms in compressed sensing: optimization methods, greedy methods, and thresholding-based methods. We will mainly focus on basis pursuit (an optimization method) while discussing its null space property, stability and robustness. In later chapters, more thorough algorithm analyses will be provided as we introduce mathematical tools like coherence and restricted isometry property.

3.1 Optimization Methods: Basis Pursuit

In last chapter, we have proved the NP-hardness of ℓ_0-minimization. As mentioned in the introductory chapter, ℓ_1-minimization, also known as basis pursuit, is one of the most popular optimization method for compressed sensing. In contrast to ℓ_0-minimization, which is a nonconvex optimization problem, basis pursuit on the other hand, is a convex optimization problem, meaning there are many algorithms from convex optimization can be applied to the basis pursuit sparse recovery scheme. Basis pursuit has the following procedure:

Basis Pursuit

Input: measurement matrix $\mathbf{A} \in \mathbb{C}^{m \times N}$, measurement vector $\mathbf{y} \in \mathbb{C}^m$.
Instruction:

$$\mathbf{x}^\sharp = \operatorname{argmin} \|\mathbf{z}\|_1 \quad \text{subject to } \mathbf{A}\mathbf{z} = \mathbf{y}. \tag{BP}$$

Output: the vector $\mathbf{x}^\sharp$.

Let $\mathbf{A} \in \mathbb{R}^{m \times N}$ be a measurement matrix with columns $\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_N \in \mathbb{C}^m$. Assume the ℓ_1-minimizer $\mathbf{x}^\sharp$ is unique, then the system $\{\mathbf{a_j}, \operatorname{supp}(\mathbf{x}^\sharp)\}$ is linearly independent. More importantly,

$$\left\|\mathbf{x}^\sharp\right\|_0 = \operatorname{card}(\operatorname{supp}(\mathbf{x}^\sharp)) \leq m. \tag{29}$$

Note there is a more generalized version of basis pursuit where the procedure is noise aware:

Quadratically Constrained Basis Pursuit

Input: measurement matrix $\mathbf{A} \in \mathbb{C}^{m \times N}$, measurement vector $\mathbf{y} \in \mathbb{C}^{m}$.
Instruction:

$$\mathbf{x}^{\sharp} = \text{argmin}\|\mathbf{z}\|_1 \quad \text{subject to } \|\mathbf{A}\mathbf{z} - \mathbf{y}\|_2 \leq \eta.$$

Output: the vector $\mathbf{x}^{\sharp}$.

The solution $\mathbf{x}^{\sharp}$ is closed related to the output of basis pursuit denoising, which is the following problem with some parameter $\lambda \geq 0$,

$$\underset{\mathbf{z} \in \mathbb{C}^N}{\text{minimize}} \ \lambda\|\mathbf{z}\|_1 + \|\mathbf{A}\mathbf{z} - \mathbf{y}\|_2^2. \tag{30}$$

And the solution $\mathbf{x}^{\sharp}$ also closely relates to the output of LASSO, which has the following scheme for some parameter $\tau \geq 0$,

$$\underset{\mathbf{z} \in \mathbb{C}^N}{\text{minimize}} \|\mathbf{A}\mathbf{z} - \mathbf{y}\|_2^2 \quad \text{subject to } \|\mathbf{z}\|_1 \leq \tau. \tag{31}$$

We can summarize some relationships between the outputs of these three schemes as follows:
(a) If $\mathbf{x}$ is a minimizer of the basis pursuit denoising with $\lambda \geq 0$, then there exist some noise $\eta = \eta_{\mathbf{x}} \geq 0$ such that $\mathbf{x}$ is a minimizer of the quadratically constrained basis pursuit.
(b) If $\mathbf{x}$ is a minimizer of the quadratically constrained basis pursuit with a noise $\eta > 0$, then there exists $\tau = \tau_x \geq 0$ such that $\mathbf{x}$ is a unique minimizer to the LASSO.
(c) If $\mathbf{x}$ is a minimizer of the LASSO with some $\tau > 0$, then there exists $\lambda = \lambda_{\mathbf{x}} \geq 0$ such that $\mathbf{x}$ is a minimizer of the basis pursuit denoising.

3.1.1 Null Space Property

Null space property is a necessary condition for the success of exact recovery of sparse vectors via basis pursuit. And it is defined as follows:
A matrix $\mathbf{A} \in \mathbb{C}^{m \times N}$ is said to satisfy the null space property relative to a set $S \subset [N]$ if

$$\|\mathbf{v}_S\|_1 < \|\mathbf{v}_{\bar{S}}\|_1 \text{ for all } \mathbf{v} \in \ker\mathbf{A} \setminus \{\mathbf{0}\}. \tag{32}$$

It is said to satisfy the null space property of order s if it satisfies the null space property relative to any set $S \subset [N]$ with $\text{card}(S) \leq s$.
Given a matrix $\mathbf{A} \in \mathbb{C}^{m\times N}$, every vector $\mathbf{x} \in \mathbb{C}^N$ is the unique solution of with $\mathbf{y} = \mathbf{A}\mathbf{x}$ if and only if $\mathbf{A}$ satisfies the null space property of order s.

3.1.2 Stability

In real world, without the idealized situations, the vector we ought to recover via basis pursuit are approximately close to sparse vectors. Thus, when we recover a vector $\mathbf{x} \in \mathbb{C}^N$, we want to have an error controlled by its distance to a s-sparse vectors.

A matrix $\mathbf{A} \in \mathbb{C}^{m\times N}$ is said to satisfy the stable null space property with constant $0 < \rho < 1$ relative to a set $S \subset [N]$ if

$$\|\mathbf{V}_S\|_1 \leq \rho \|\mathbf{V}_{\bar{S}}\|_1 \text{ for all } \mathbf{v} \in \ker\mathbf{A} \setminus \{\mathbf{0}\}. \tag{33}$$

With an enhanced (stable) null space property, we can then introduce the theorem which concludes the stability of basis pursuit.
Suppose that a matrix $\mathbf{A} \in \mathbb{C}^{m\times N}$ satisfies the stable null space property of order s with constant $0 < \rho < 1$. Then, for any $\mathbf{x} \in \mathbb{C}^N$, a solution $\mathbf{x}^\sharp$ of ℓ_1-minimization with $\mathbf{y} = \mathbf{A}\mathbf{x}$ approximates the vector $\mathbf{x}$ with ℓ_1-error

$$\left\|\mathbf{x} - \mathbf{x}^\sharp\right\|_1 \leq \frac{2(1+\rho)}{1-\rho}\sigma_s(\mathbf{x})_1. \tag{34}$$

3.1.3 Robustness

Realistically speaking, the measurement of a signal $\mathbf{x} \in \mathbb{C}^N$ is impossible to be exact. This fact implies $\mathbf{y} \in \mathbb{C}^m$ is merely an approximation of the vector $\mathbf{A}\mathbf{x} \in \mathbb{C}^m$ where

$$\|\mathbf{A}\mathbf{x} - \mathbf{y}\| \leq \eta, \tag{35}$$

for some norm (in most cases, it's the ℓ_2-norm). The null space property can be reinforced by the robustness and gives us a better condition for the exact recovery of basis pursuit. The matrix $\mathbf{A} \in \mathbb{C}^{m\times N}$ is said to satisfy the robust null space property with constants $0 < \rho < 1$ and $\tau > 0$ relative to a set $S \subset [N]$ if

$$\|\mathbf{v}_S\|_1 \leq \rho \|\mathbf{v}_{\bar{S}}\|_1 + \tau \|\mathbf{A}\mathbf{v}\| \quad \text{for all } \mathbf{v} \in \mathbb{C}^N \tag{36}$$

It is said to satisfy the robust null space property of order s with constants $0 < \rho < 1$ and $\tau > 0$ if it satisfies the robust null space property with constants ρ, τ relative to any set $S \in [N]$ with $\text{card}(S) \leq s$. And if a matrix $\mathbf{A} \in \mathbb{C}^{m \times N}$ satisfies the robust null spacer property of order s with constants $0 < \rho < 1$ and $\tau > 0$, then for any $\mathbf{x} \in \mathbb{C}^N$, a solution $\mathbf{x}^\sharp$ of the quadratically constrained basis pursuit, with $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$ and $\|\mathbf{e}\| \leq \eta$ approximates the vector $\mathbf{x}$ with ℓ_1 error

$$\left\|\mathbf{x} - \mathbf{x}^\sharp\right\| \leq \frac{2(1+p)}{1-p} \sigma_s(\mathbf{x})_1 + \frac{4\tau}{1-\rho} \eta. \tag{37}$$

3.2 Greedy Methods

Frankly speaking, greedy algorithm has its name because it makes optimal choices heuristically at each step to get an optimal solution to the overall problem. For sparse vector recovery, there are two commonly used iterative greedy algorithms:

1. Orthogonal Matching Pursuit (OMP)
2. Compressive Sampling Matching Pursuit (CoSaMP)

For Orthogonal Matching Pursuit, we have the following procedure:
Let S^n be the target support at each iteration. We update our target vector $\mathbf{x}^n$ as its supported on S^n that fits the measurements the best. Then we have the following algorithm:

Orthogonal Matching Pursuit

Input: measurement matrix $\mathbf{A} \in \mathbb{C}^{m \times N}$, measurement vector $\mathbf{y} \in \mathbb{C}^m$.
Initialization: $S^0 = \emptyset, \mathbf{x^0} = \mathbf{0}$.
Iteration: repeat until a stopping criterion is met at $n = \bar{n}$:

$$S^{n+1} = S^n \cup \{j_{n+1}\}, \quad j_{n+1} := \underset{j \in [N]}{\operatorname{argmax}}\{|(\mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}^n))_j|\}, \tag{OMP1}$$

$$\mathbf{x}^{n+1} = \underset{\mathbf{z} \in \mathbb{C}^N}{\operatorname{argmin}}\{\|\mathbf{y} - \mathbf{A}\mathbf{z}\|_2, \operatorname{supp}(\mathbf{z}) \subset S^{n+1}\}, \tag{OMP2}$$

Output: the $\bar{n}$-sparse vector $\mathbf{x}^\sharp = \mathbf{x}^{\bar{n}}$.

In OMP, the following step is the most computationally expensive,

$$\mathbf{x}^{n+1} = \underset{\mathbf{x} \in \mathbb{C}^N}{\operatorname{argmin}}\{\|\mathbf{y} - \mathbf{A}\mathbf{z}\|_2, \operatorname{supp}(\mathbf{z}) \in S^{n+1}\}, \tag{38}$$

but it can be accelerated by using the QR-decomposition of $\mathbf{A}_{S_n}$. And the choice of the index j_{n+1} is determined by the greedy strategy where the aim is to reduce the ℓ_2 norm of the residual $\mathbf{y} - \mathbf{A}\mathbf{x}^n$ as much as possible. However, orthogonal matching pursuit is not without any weakness. If an incorrect index has been selected in a target support S^n, it remains in all the subsequent target supports $S^{n'}$ for $n' \geq n$. In this situation if an incorrect index has been selected, s iterations of the orthogonal matching pursuit will not suffice in recovering a s-sparse vector. The easiest possible solution will be just to increase the number of iterations until it meets the criteria, but there is a different greedy algorithm using another strategy which is mentioned in this section later. Note there are certain conditions to hold for orthogonal matching pursuit to work as expected. In particular, given a matrix $\mathbf{A} \in \mathbb{C}^{m \times N}$, every nonzero vector $\mathbf{x} \in \mathbb{C}^N$ supported on a set S of size s is recovered from $\mathbf{y} = \mathbf{A}\mathbf{x}$ after at most s iterations of orthogonal matching pursuit if and only if the column submatrix $\mathbf{A}_S$ is injective and

$$\max_{j \in S}|(\mathbf{A}^*\mathbf{r})_j| > \max_{\ell \in S}|(\mathbf{A}^*\mathbf{r})_\ell| \tag{39}$$

for all nonzero $\mathbf{r} \in \{\mathbf{A}\mathbf{z}, \operatorname{supp}(\mathbf{z}) \subset S\}$. This can be formulated in a more concise way as the exact recovery condition as follows:

$$\left\|\mathbf{A}_S^\dagger \mathbf{A}_{\bar{S}}\right\|_{1 \to 1} < 1. \tag{40}$$

where $\mathbf{A}_S^{\dagger}$ is the Moore-Penrose pseudo-inverse of the matrix $\mathbf{A}$.

Compressed sampling matching pursuit (CoSaMP) is an alternative greedy algorithm which can be very useful providing an estimation of the sparsity s. We introduce the following notations:

$L_s(\mathbf{z}) :=$ index set of s largest absolute entries of $\mathbf{z} \in \mathbb{C}^N$,

$H_s(\mathbf{z}) := Z_{L_s(\mathbf{z})}$.

where H_s is a nonlinear operator called hard thresholding operator of order s. And $H_s(\mathbf{z})$ is the best s-term approximation to $\mathbf{z} \in \mathbb{C}^N$. Then we have the following procedure for CoSaMP:

Compressed sampling matching pursuit

Input: measurement matrix $\mathbf{A} \in \mathbb{C}^{m \times N}$, measurement vector $\mathbf{y}$, sparsity level s.

Initialization: s-sparse vector $\mathbf{x^0}$, typically $\mathbf{x^0} = \mathbf{0}$.

Iteration: repeat until a stopping criterion is met at $n = \bar{n}$:

$$U^{n+1} = \mathrm{supp}(\mathbf{x}^n) \cup L_{2s}(\mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}^n)), \tag{CoSaMP1}$$

$$\mathbf{u}^{n+1} = \underset{\mathbf{z} \in \mathbb{C}^N}{\mathrm{argmin}} \{\|\mathbf{y} - \mathbf{A}\mathbf{z}\|_2, \mathrm{supp}(\mathbf{z}) \subset U^{n+1}\}, \tag{CoSaMP2}$$

$$\mathbf{x}^{n+1} = H_s(\mathbf{u}^{n+1}). \tag{CoSaMP3}$$

Output: the $\bar{n}$-sparse vector $\mathbf{x}^{\sharp} = \mathbf{x}^{\bar{n}}$.

---

4 Coherence

For compressed sensing, coherence is a very simple measure of the suitability of the measurement matrix during the analysis of recovery algorithms.

4.1 Definitions

Coherence is defined as follows: Let $\mathbf{A} \in \mathbb{C}^{m \times N}$ be a matrix with ℓ_2-normalized columns $\mathbf{a}_1, ..., \mathbf{a}_N$, i.e., $\|\mathbf{a}_i\|_2 = 1$ for all $i \in [N]$. The *coherence* $\mu = \mu(A)$ of the matrix $\mathbf{A}$ is defined as

$$\mu := \max_{1 \leq i \neq j \leq n} |\langle \mathbf{a}_i, \mathbf{a}_j \rangle|. \tag{41}$$

And the ℓ_1 coherence function μ_1 of the matrix $\mathbf{A}$ is defined for $s \in [N-1]$ by

$$\mu_1(s) := \max_{i \in [N]} \Big\{ \sum_{j \in S} |\langle \mathbf{a}_i, \mathbf{a}_j \rangle| \quad S \subset [N], \quad card(S) = s, \quad i \notin S \Big\}. \tag{42}$$

A small coherence implies that column submatrices of moderate size are well-conditioned. By definition, we know that for $1 \leq s, t \leq N-1$,

$$\max\{\mu_1(s), \mu_1(t)\} \leq \mu_1(s+t) \leq \mu_1(s) + \mu_1(t) \tag{43}$$

Note the ℓ_1-$coherence function$ μ_1 is invariant under multiplication by a unitary matrix $\mathbf{U}$, and by Cauchy-Schwarz inequality $|\langle \mathbf{U}\mathbf{a}_i, \mathbf{U}\mathbf{a}_j \rangle| \leq \|\mathbf{a}_i\| \|\mathbf{a}_j\|$ we get that $\mu \leq 1$. And this result serves as an upper bound to the coherence.
However, there is a tighter upper bound. Given a matrix $\mathbf{A} \in \mathbb{C}^{m \times N}$ with ℓ_2-*normalized columns* and an integer $s \geq 1$, if $\mu_1(s) + \mu_1(s+1) < 1$, then for each set $s \subset [N]$ with $card(S) \leq 2S$, the matrix $\mathbf{A}_S^* \mathbf{A}_S$ is invertible and the matrix $\mathbf{A}_S$ injective. In particular, the conclusion holds if

$$\mu < \frac{1}{2s-1}. \tag{44}$$

There is a more accurate lower bound than $\mu \geq 0$ called *Welch bound*. Before introducing it, it is important to know the definitions of equiangularity and tight frame as a prerequisite.

Equiangularity: A system of ℓ_2-normalized vectors $(\mathbf{a}_1, ...\mathbf{a}_N)$ in $\mathbb{C}^m$ is called equiangular if there is a constant $c \geq 0$ such that

$$|\langle \mathbf{a}_i, \mathbf{a}_j \rangle| = c, \text{for all } i, j \in [N], i \neq j. \tag{45}$$

Tight Frame: A system of vectors $(\mathbf{a}_1, ...\mathbf{a}_N)$ is called a tight frame if there exist a constant $\lambda > 0$ such that one of the following conditions holds:

$$\text{(a) } \|\mathbf{x}\|_2^2 = \lambda \sum_{j=1}^{N} |\langle \mathbf{x}, \mathbf{a}_j \rangle|^2, \text{ for all } \mathbf{x} \in \mathbb{C}^m, \tag{46}$$

$$\text{(b) } \mathbf{x} = \lambda \sum_{j=1}^{N} \langle \mathbf{x}, \mathbf{a}_j \rangle \mathbf{a}_j, \text{ for all } \mathbf{x} \in \mathbb{C}^m \tag{47}$$

$$\text{(c) } \mathbf{A}\mathbf{A}^* = \frac{1}{\lambda} \mathbf{Id}_m \text{where } \mathbf{A} \text{ is the matrix with columns } \mathbf{a}_1, ..., \mathbf{a}_N. \tag{48}$$

The coherence of a matrix $\mathbf{A} \in \mathbb{C}^{m \times N}$ with ℓ_2-normalized column satisfies

$$\mu \geq \sqrt{\frac{N-m}{m(N-1)}}. \tag{49}$$

Equality holds if and only if the columns $\mathbf{a}_1, ..., \mathbf{a}_N$ of the matrix $\mathbf{A}$ form an equiangular tight frame.

4.2 Analysis of Orthogonal Matching Pursuit

The success of sparse recovery via analysis of orthogonal matching pursuit can be guaranteed by a coherence that is small enough. In this section, we introduce this key result as below.

Let $\mathbf{A} \in \mathbb{C}^{m \times N}$ be a matrix with ℓ_2-normalized columns. If

$$\mu_1(s) + \mu_1(s-1) < 1, \tag{50}$$

then every s-sparse vector $\mathbf{x} \subset \mathbb{C}^N$ can be recovered from the measurement vector $\mathbf{y} = \mathbf{A}\mathbf{x}$ after at most s iterations of orthogonal matching pursuit. To prove this theorem, we let $\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_N$ denote the ℓ_2-normalized columns of

A. From the exact recovery condition of Orthogonal Matching Pursuit, we want to show that for any $S \in [N]$ with $\text{card}(S) = s$, the column submatrix $\mathbf{A}_S$ is injective and that

$$\max_{j \in S} |\langle(\mathbf{r}, \mathbf{a}_j)\rangle| > \max_{\ell \in \bar{S}} |\langle(\mathbf{r}, \mathbf{a}_\ell)\rangle| \tag{51}$$

for all nonzero $\mathbf{r} \in \{\mathbf{Az}, \text{supp}(\mathbf{z}) \subset S\}$. Let $\mathbf{r} := \sum_{i \in S} r_i \mathbf{a}_i$ be such a vector and choose $k \in S$ so that $|r_k| = \max_{i \in S}(|r_i|)$. Note for $\ell \in \bar{S}$, we have

$$|\langle \mathbf{r}, \mathbf{a}_\ell \rangle| = |\sum_{i \in S} r_i \langle \mathbf{a}_i, \mathbf{a}_\ell \rangle| \leq \sum_{i \in S} |r_i||\langle \mathbf{a}_i, \mathbf{a}_\ell \rangle| \leq |r_k|\mu_1(s). \tag{52}$$

On the other hand, we get the following inequality

$$|\langle \mathbf{r}, \mathbf{a}_k \rangle| = |\sum_{i \in S} r_i \langle \mathbf{a}_i, \mathbf{a}_k \rangle| \geq |r_k||\langle \mathbf{a}_k, \mathbf{a}_k \rangle| - \sum_{i \in S, i \neq k} |r_i||\langle \mathbf{a}_i, \mathbf{a}_k \rangle|. \tag{53}$$

With the above inequalities, we have proved the inequality associated with the exact recovery condition holds because our assumption $1 - \mu_1(s-1) > \mu_1(s)$. And it is also injective because (42).

4.3 Analysis of Basis Pursuit

A small coherence also guarantees of success of basis pursuit. The following result for sparse vector recovery via basis pursuit is consequent.

Let $\mathbf{A} \in \mathbb{C}^{m \times N}$ be a matrix with ℓ_2-normalized columns. If

$$\mu_1(s) + \mu_1(s-1) < 1, \tag{54}$$

then every s-sparse vector $\mathbf{x} \in \mathbb{C}^N$ can be recovered from the measurement vector $\mathbf{y} = \mathbf{Ax}$ via basis pursuit.

In fact, any condition guarantees the success of recovery of orthogonal matching pursuit automatically guarantees the success of basis pursuit recovery. This holds true because the exact recovery condition of orthogonal matching pursuit implies the null space property, which is the property that validates the exact recovery of sparse vectors via basis pursuit.

5 Restricted Isometry Property

Though coherence is very useful in determining if a measurement matrix is well conditioned, however, when the measurement matrix has a large sparsity levels, the performance analysis will be limited as a result of the Welch Bound. In this case, we need a better measure to apply to such matrices. Here we introduce the concept of *restricted isometry property* (also known as $uniform$ *uncertainty principle*). Note coherence only samples some pairs of columns of a measurement matrix, on the other hand, the restricted isometry property of order s takes all s-tuples of columns into account which guarantees a better assessment quality.

5.1 Definitions

The sth restriced isometry constant $\delta_s = \delta_s(\mathbf{A})$ of a matrix $\mathbf{A} \in \mathbb{C}^{m \times N}$ is the smallest $\delta \geq 0$ such that

$$(1-\delta)\|\mathbf{x}\|_2^2 \leq \|\mathbf{A}\mathbf{x}\|_2^2 \leq (1+\delta)\|\mathbf{x}\|_2^2. \tag{55}$$

for all s-sparse vectors $\mathbf{x} \in \mathbb{C}^N$. Equivalently, it is given by by

$$\delta_s = \max_{S \subset [N], card(S) \leq s} \|\mathbf{A}_S^* \mathbf{A}_S - \mathbf{Id}\|_{2 \to 2}. \tag{56}$$

For the above equivalency to hold, there are three pre-established remarks.

First, the sequence of restricted isometry constant is nondecreasing such that,

$$\delta_1 \leq \delta_2 \leq ... \leq \delta_s \leq \delta_{s+1} \leq ... \leq \delta_N. \tag{57}$$

Second, in the situation of compressed sensing, we only consider the relevant case for $\delta_s < 1$ (Though it is possible that $\delta_s \geq 1$ when it's out of scope of compressed sensing).

Third, if the entries of the measurement matrix $\mathbf{A}$ are real, then δ_s could also be defined as the smallest $\delta \geq 0$ such that (8) holds for all real s-sparse vector $\mathbf{x} \in \mathbb{R}^N$.

Now we observe that

$$|\|\mathbf{A}_S\mathbf{x}\|_2^2 - \|\mathbf{x}\|_2^2| \leq \delta\|\mathbf{x}\|_2^2 \text{ for all } S \subset [N],\ card(S) \leq s,\ \text{and all } \mathbf{x} \in \mathbb{C}^S \tag{58}$$

And for $\mathbf{x} \in \mathbb{C}^S$,

$$\|\mathbf{A}_S\mathbf{x}\|_2^2 - \|\mathbf{x}\|_2^2 = \langle \mathbf{A}_S\mathbf{x}, \mathbf{A}_S\mathbf{x}\rangle - \langle \mathbf{x}, \mathbf{x}\rangle = \langle (\mathbf{A}_S^* - \mathbf{Id})\mathbf{x}, \mathbf{x}\rangle. \tag{59}$$

Note $\mathbf{A}_S^* \mathbf{A}_S - \mathbf{Id}$ is Hermitian (or self-adjoint, meaning it is equal to its own conjugate transpose), so we have

$$\max_{\mathbf{x} \in \mathbb{C}^s \setminus \{\mathbf{0}\}} \frac{\langle (\mathbf{A}_S^* \mathbf{A}_S - \mathbf{Id})\mathbf{x}, \mathbf{x} \rangle}{\|\mathbf{x}\|_2^2} = \|\mathbf{A}_S^x - \mathbf{Id}\|_{2\to 2} \tag{60}$$

which is equal to

$$\delta_s = \max_{S \subset [N], card(S) \leq s} \|\mathbf{A}_S^* \mathbf{A}_S - \mathbf{Id}\|_{2\to 2}. \tag{61}$$

Now we can make the comparison between restricted isometry constant of a matrix and its coherence μ and coherence function μ_1 with the following proposition:
If the matrix $\mathbf{A}$ has ℓ_2-normalized columns $\mathbf{a}_1, ..., \mathbf{a}_N$ i.e., $\|\mathbf{a}\|_2 = 1$ for all $j \in [N]$, then

$$\delta_1 = 0, \quad \delta_2 = \mu, \quad \delta_s \leq \mu_1(s-1) \leq (s-1)\mu, \quad s \geq 2. \tag{62}$$

The (s,t)-*restricted orthogonality constant* $\theta_{s,t} = \theta_{s,t}(\mathbf{A})$ of a matrix $\mathbf{A} \in \mathbb{C}^{m \times N}$ is the smallest $\theta \geq 0$ such that

$$|\langle \mathbf{A}\mathbf{u}, \mathbf{A}\mathbf{v} \rangle| \leq \theta \|\mathbf{u}\|_2 \|\mathbf{v}\|_2 \tag{63}$$

for all disjointly supported s-sparse and t-sparse vector $\mathbf{u}, \mathbf{v} \in \mathbb{C}^N$. Equivalently,

$$\theta_{s,t} = max\left\{ \left\| \mathbf{A}_T^2 \mathbf{A}_s \right\|_{2\to 2}, S \cap T = \emptyset, card(S) \leq s, card(T) \leq t \right\}. \tag{64}$$

The relationship between restricted isometry constants and restricted orthogonality constant is as follows:

$$\theta_{s,t} \leq \mu_{s,t} \leq \frac{1}{s+t}(s\delta_s + t\delta_t + 2\sqrt{st}\theta_{s,t}) \tag{65}$$

And there is a special case $s = t$ which gives the follow inequalities

$$\theta_{s,s} \leq \delta_{2s} \quad and \quad \delta_{2s} \leq \delta_s + \theta_{s,s}. \tag{66}$$

In order to obtain the restricted isometry property $\delta_s \leq \delta$ in the optimal regime for compressed sensing, random matrices will be used. And it is an open problem to find deterministic (non-random) matrices which satisfies restricted isometry property.

5.2 Analysis of Basis Pursuit

In sparse recovery via basis pursuit, small restricted isometry constants for the measurement matrices guarantees the success of the recovery. In this section, we present a simple algorithmic analysis on basis pursuit without considering stability and robustness.

Suppose that the 2sth restricted isometry constant of the matrix $\mathbf{A} \in {}^{m\times N}$ satisfies

$$\delta_{2s} < \frac{1}{3} \tag{67}$$

Then every s-sparse vector $\mathbf{x} \in \mathbb{C}^N$ is the unique solution of

$$\underset{\mathbf{z}\in\mathbb{C}^N}{\text{minimize}}\ \|\mathbf{z}\|_1 \qquad \text{subject to } \mathbf{Az} = \mathbf{Ax}. \tag{68}$$

Suppose that the 2sth restricted isometry constant of the matrix $\mathbf{A} \in {}^{m\times N}$ satisfies

$$\delta_{2s} < \frac{4}{\sqrt{41}} \tag{69}$$

Then for any $\mathbf{x} \in \mathbb{C}^N$ and $\mathbf{y} \in \mathbb{C}^m$ with $\|\mathbf{Ax} - \mathbf{y}\|_2 \le \eta$, a solution $\mathbf{x}^\sharp$ of

$$\underset{\mathbf{z}\in\mathbb{C}^N}{\text{minimize}}\ \|\mathbf{z}\|_1 \qquad \text{subject to } \|\mathbf{Az} - \mathbf{y}\|_2 \le \eta \tag{70}$$

approximates the vector $\mathbf{x}$ with errors

$$\left\|\mathbf{x} - \mathbf{x}^\sharp\right\|_1 \le C\sigma_s(\mathbf{x})_1 + D\sqrt{s}\eta, \tag{71}$$

$$\left\|\mathbf{x} - \mathbf{x}^\sharp\right\|_2 \le \frac{C}{\sqrt{s}}\sigma_s(\mathbf{x})_1 + D\eta, \tag{72}$$

where the constants C, $D > 0$ depend only on δ_{2s}.
If the 2sth restricted isometry constant of $\mathbf{A} \in {}^{m\times N}$ obeys the theorem above, then the matrix $\mathbf{A}$ satisfies the ℓ_2-robust null space property of order s with constants $0 < \rho < 1$ and $\tau > 0$ depending only on δ_{2s}.

5.3 Analysis of Orthogonal Matching Pursuit

Similar to last section, we provide a simple analysis on Orthogonal Matching Pursuit. Recall for Orthogonal Matching Pursuit, we let S^n be the target support at each iteration. We update our target vector $\mathbf{x}^n$ as its supported on S^n that fits measurements the best.

Given $\mathbf{A} \in \mathbb{C}^{m\times N}$, let $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$ for some s-sparse vector $\mathbf{x} \in \mathbb{C}^N$ with $S = supp(x)$ and some $\mathbf{e} \in \mathbb{C}^m$. Let $(\mathbf{x}^n)$ denote the sequence defined by (OMP1), (OMP2) started at an index set S^0. With $s^0 = card(S^0)$ and $s' = card(S\backslash S^0)$, if $\delta_{s+s^0+12s'} < 1/6$, then there is a constant $C > 0$ depending only on $\delta_{s+s^0+12s'}$ such that

$$\|\mathbf{y} - \mathbf{A}\mathbf{x}^{\bar{n}}\|_2 \leq C\|\mathbf{e}\|_2, \quad \bar{n} = 12s'. \tag{73}$$

Suppose that $\mathbf{A} \in {}^{m\times N}$ has restricted isometry property

$$\delta_{13s} < \frac{1}{6}. \tag{74}$$

Then there is a constant $C > 0$ depending only on δ_{13s} such that, for all $\mathbf{x} \in \mathbb{C}^N$ and $\mathbf{e} \in \mathbb{C}^m$, the sequence $\mathbf{x}^n$ defined by (OMP1), (OMP2) with $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$ satisfies

$$\left\|\mathbf{y} - \mathbf{A}\mathbf{x}^{12s}\right\|_2 \leq C\|\mathbf{A}\mathbf{x}_{\bar{S}} + \mathbf{2}\|_2 \tag{75}$$

for any $S \subset [N]$ with $card(S) = s$.

6 Compressed Sensing with Random Matrices

We have demonstrated the guarantees of recovery of each algorithm type if the restricted isometry constant of the measurement matrix satisfy $\delta_{\kappa s} \leq \delta_*$ for some small integer κ and some $\delta_* \in (0, 1)$. In this chapter, we specifically discuss nonuniform recovery with random matrices, with a focus on subgaussian matrices.

6.1 Subgaussian Matrices

Before proceeding to establish important theorems about random matrix recovery, it is important to garner some understandings about random matrices. Gaussian and subgaussian random matrices are very important for the theory of compressive sensing because they provide a model of measurement matrices which can be analyzed very accurately [**?**]. Although there are different many other types of random matrices such Bernoulli matrices and rademacher matrices, for the simplicity of the demonstration, we only discuss Gaussian and subgaussian matrices in our context.

A random variable X is called subgaussian if there exist constant $\beta, \kappa > 0$ such that

$$\mathbb{P}(|X| \geq t) \leq \beta e^{-\kappa t^2} \quad \text{for all } t > 0. \tag{76}$$

Note if X is subgaussian with $\mathbb{E}X = 0$, then there exists a constant c (depending only on β and κ) such that

$$\mathbb{E}[\exp(\theta X)] \leq \exp(c\theta^2) \quad \text{for all } \theta \in \mathbb{R}. \tag{77}$$

And if the above inequality holds, then $\mathbb{E}X = 0$ and X is subgaussian with parameters $\beta = 2$ and $\kappa = 1/(4c)$.

A matrix is a random matrix $\mathbf{A} \in \mathbb{C}^{m \times N}$ if it has random variables as entries. If the entries of $\mathbf{A}$ are independent mean-zero subgaussian random variables with variance 1 and same subgaussian parameters β, κ as above, then $\mathbf{A}$ is a subgaussian matrix. And if the entries of $\mathbf{A}$ are independent Gaussian random variables, then $\mathbf{A}$ is called a Gaussian random matrix. Note both Bernoulli and Gaussian are subgaussian, and the entries of subgaussian matrices are not always identically distributed. Consider a random vector $\mathbf{Y} \in \mathbb{R}^N$. If for all $\mathbf{x} \in \mathbb{R}^N$ with $\|\mathbf{x}\|_2 = 1$, then the random variable $\langle \mathbf{Y}, \mathbf{x} \rangle$

is subgaussian with subgaussian parameter $\bar{c}$ being independent of $\mathbf{x}$, that is,

$$\mathbb{E}[exp(\theta\langle \mathbf{Y}, \mathbf{x}\rangle)] \leq exp(\bar{c}\theta^2), \quad \text{for all } \theta \in \mathbb{R}, \quad \|\mathbf{x}\|_2 = 1, \tag{78}$$

then $\mathbf{Y}$ is called a subgaussian random vector.

6.2 Restricted Isometry Property for Subgaussian Matrices

Now we introduce the key result on the restricted isometry property for subgaussian random matrices, which is widely cited in the literature.

Consider a subgaussian random matrix $\mathbf{A} \in \mathbb{C}^{m\times N}$ with independent entries, then there exists a constant $C > 0$ (dependent only on the subgaussian parameters β, κ) such that the restricted isometry constant of $\frac{1}{\sqrt{m}}\mathbf{A}$ satisfies $\delta_s \leq \delta$ with probability at least $1 - \epsilon$ provided

$$m \geq C\delta^{-2}(s \ln(eN/s) + \ln(2\epsilon^{-1})). \tag{79}$$

Then if we let $\epsilon = 2exp(-\delta^2 m/(2C))$, we get the following condition

$$m \geq 2C\delta^{-2}s \ln(eN/s), \tag{80}$$

which guarantees that $\delta_s \leq \delta$ with probability at least $1-2\exp(-\delta^2 m/(2C))$.

Note we normalize the matrix as $\frac{1}{\sqrt{m}}\mathbf{A}$ because $\mathbb{E}\|\frac{1}{\sqrt{m}}\mathbf{A}\mathbf{x}\|_2^2 = \|\mathbf{x}\|_2^2$ for a fixed vector $\mathbf{x}$ and subgaussian random $\mathbf{A}$ with variance 1 for all entries. To generalize the theorem to a larger class of random matrices, we introduce the definition of random isotropic vectors. Consider a random vector $\mathbf{Y} \in \mathbb{R}^N$,
(a) If $\mathbf{E}|\langle \mathbf{Y}, \mathbf{x}\rangle|^2 = \|\mathbf{x}\|_2^2$ for all $\mathbf{x} \in \mathbb{R}^N$, then $\mathbf{Y}$ is called isotropic.
(b) If for all $\mathbf{x} \in \mathbb{R}^N$ with $\mathbf{x} = 1$, the random variable $\langle \mathbf{Y}, \mathbf{x}\rangle$ is subgaussian with subgaussian parameter c being independent of x, that is,

$$\mathbb{E}[\exp(\theta\langle \mathbf{Y}, \mathbf{x}\rangle)] \leq \exp(c\theta^2) \quad \text{for all } \theta \in \mathbb{R}, \quad \mathbf{x} = 1, \tag{81}$$

then $\mathbf{Y}$ is called a subgaussian random vector.

Now the key results for subgaussian random matrices can be generalized to random matrices with independent, isotropic, and subgaussian rows as following:

Consider a random matrix with independent, isotropic, and subgaussian rows $\mathbf{A} \in \mathbb{R}^{m \times N}$ with the same subgaussian parameter as in (81), if

$$m \geq C\delta^{-2}(s \ln(eN/s) + \ln(2\epsilon^{-1})), \tag{82}$$

then the restricted isometry constant of $\dfrac{1}{\sqrt{m}}\mathbf{A}$ satisfies $\delta_s \leq \delta$ with probability at least $\mathbf{1} - \epsilon$.

6.3 Nonuniform Recovery with Subgaussian Matrices

When we discuss the minimal number of measurements in section 2.2, we mentioned that there are two types of recovery schemes in compressed sensing: uniform recovery and nonuniform recovery. As we discussed uniform recovery in previous chapters, we are going to introduce nonuniform recovery with subgaussian matrices in comparison to the uniform recovery.

The uniform recovery states with high probability on the draw of the random matrix, every sparse vector can be reconstructed under appropriate conditions. On the other hand the nonuniform recovery states that a given sparse vector x can be reconstructed with high probability on the the draw of the matrix under appropriate conditions. The difference to uniform recovery is that nonuniform recovery does not imply that there is a matrix that recovers all x simultaneously. Or in other words, the small exceptional set of matrices for which recovery fails may depend on $\mathbf{x}$ [**?**]. Let $\mathbf{x} \in \mathbb{C}^N$ be an s-sparse vector. Let $\mathbf{A} \in \mathbb{C}^{m \times N}$ be a randomly drawn subgaussian matrix with subgaussian parameter c in (77). If for some $\epsilon \in (0, 1)$,

$$m \geq \frac{4c}{1-\delta} s \ln(2N/\epsilon), \quad \text{with } \delta = \sqrt{\frac{C}{4c}\left(\frac{7}{\ln(2N/\epsilon) + \frac{2}{s}}\right)} \tag{83}$$

with the assumption that N and s are large enough so that $\delta < 1$, then with probability at least $1-\epsilon$, the vector $\mathbf{x}$ is the unique minimizer of $\|\mathbf{z}\|_1$ subject to $\|\mathbf{Az} = \mathbf{Ax}\|$. And the constant $C = 2/(3\bar{c})$ depends only the subgaussian parameter $\bar{c}$ in (78).

Note the parameter δ becomes near zero as N and s get large. In this sense, if $m > 4cs \ln(2N/\epsilon)$, the sparse recovery is said to be sufficient. For

Gaussian and Bernoulli random matrices where $c = 1/2$, the approximate sufficient condition is as follows:

$$m > 2s \ln(2N/\epsilon). \tag{84}$$

6.4 Null Space Property for Gaussian Matrices

Here we introduce a theorem that guarantees good constants and stability of uniform recovery with Gaussian random matrices by only directly using the stable null space property. Let $\mathbf{A} \in \mathbb{C}^{m\times N}$ be a random random draw of a Gaussian matrix. Assume that

$$\frac{m^2}{m+1} \geq 2s\ln(eN/s)\left(1+\rho^{-1}+D(s/N)+\sqrt{\frac{\ln(\epsilon^{-1})}{s\ln(eN/s)}}\right)^2, \tag{85}$$

where D is a function that satisfies $D(\alpha) \leq 0.92$ for all $\alpha \in (0,1]$ and $\lim_{\alpha\to 0} D(\alpha) = 0$. Then, for probability at least $1-\epsilon$, every vector $\mathbf{x} \in \mathbb{R}^N$ is approximated by a minimizer $\mathbf{x}^\sharp$ of $\|\mathbf{z}\|_1$ subject to $\mathbf{Az} = \mathbf{Ax}$ as

$$\left\|\mathbf{x}-\mathbf{x}^\sharp\right\|_1 \leq \frac{2(1+\rho)}{1-\rho}\sigma_s(\mathbf{x})_1. \tag{86}$$

The function D in the above theorem is

$$D(\alpha) = \frac{1}{\sqrt{2\ln(e\alpha^{-1})}} + \frac{1}{(8\pi e^3)^(1/4)\ln(e\alpha^{-1})}. \tag{87}$$

For the null space property to hold, this theorem only involves the case $N \geq 2s$. This gives us a better upper bound as

$$D(\alpha) \leq D(1/2) = \frac{1}{\sqrt{2\ln(2e)}} + \frac{1}{(8\pi e^3)^{1/4}\ln(2e)} \approx 0.668. \tag{88}$$

In compressed sensing, we are most interested in the situation where we have large N ,mildly large s , and a small ratio s/N. In this case the assumption in the theorem (88) will become

$$m > 2(1+\rho^{-1})^2 s \ln(eN/s). \tag{89}$$

7 An Application of Compressed Sensing

Most parts of this thesis are dedicated to the mathematical mechanism of compressed sensing. Due to the advantages of compressed sensing over traditional signal processing technologies, it goes without saying that compressed sensing also has profound real world applications. In the last chapter of this paper, an interesting example about single-pixel imaging is presented as an example to demonstrate the real world application of compressed sensing.

7.1 Single-Pixel Camera

In traditional digital cameras, taking pictures requires focusing light through the lenses onto a light sensor, which consists of a grid of light-sensitive tiny photosites. The photosite is often called a pixel, and there are usually millions of individual pixels in today's light sensor. However, via compressed sensing technique, a device called single-pixel camera is made possible. The name of this device is self-explanatory, only a single pixel is required to reconstruct the image. The single-pixel camera takes independent realizations of Bernoulli random vectors and measures these inner products on a single pixel. with sparse recovery methods, only a small number of these random inner products are required for image reconstruction.

In the single-pixel camera, a microarray of a large number of small mirrors can be turned on and off individually. And the light from the image is reflected onto this microarray, where the light become focused via a lens and further reflect to a single pixel sensor. The measurement of light intensity is manipulated by turning each mirror on and off. [**?**]

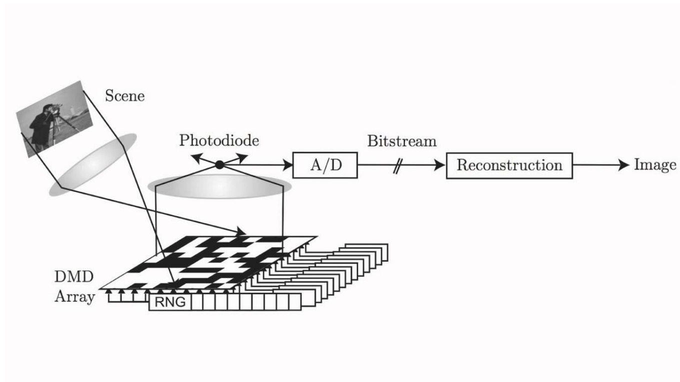

Figure 1: A picture of single-pixel camera setup [?]

Let $\mathbf{z} \in \mathbb{R}^N$ be the image vector which consists of the gray values of the collected pixels, $\mathbf{z} = \mathbf{Wx}$ where $\mathbf{x}$ is a sparse vector that we aim to recover, and $\mathbf{W} \in \mathbb{C}^{m\times N}$ is a matrix represents the sparse transform. For vectors $\mathbf{b}$, the inner products $\langle \mathbf{z}, \mathbf{b}\rangle$ stores the location of activated mirrors and zeros for the deactivated mirrors. With vectors $\mathbf{a}$ where each entry contains either 1 or -1 with equal probabilites, and two auxiliary vectors $\mathbf{b^1}, \mathbf{b^2} \in \{0,1\}^N$

$$b_j^1 = \begin{cases} 1 & \text{if } a_j = 1, \\ 0 & \text{if } a_j = -1, \end{cases} \quad b_j^2 = \begin{cases} 1 & \text{if } a_j = -1, \\ 0 & \text{if } a_j = 1, \end{cases} \tag{90}$$

we have inner products $\langle \mathbf{z}, \mathbf{a}\rangle = \langle \mathbf{z}, \mathbf{b^1}\rangle - \langle \mathbf{z}, \mathbf{b^2}\rangle$. The vectors $\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_N$ are then chosen randomly. And the measured intensities are inner products $y_\ell = \langle \mathbf{z}, \mathbf{a}_\ell\rangle$ with independent Bernoulli vectors. Thus we get $\mathbf{y} = \mathbf{Az}$ for a Bernoulli matrix $\mathbf{A} \in \mathbb{R}^{m\times N}$. Note $\mathbf{z} = \mathbf{Wx}$, so we have the system $\mathbf{Y} = \mathbf{AWx}$ where $\mathbf{x}$ is compressible or sparse. As the single-pixel camera is designed, the measurements are sampled each by each. However, with compressed sensing, only a few measurement needs to be taken to reconstruct the image.

It is natural to ask, since the current camera technology is well-developed and affordable, where does the single-pixel camera find its place on the market? It is important to note that the single-pixel camera can sample measurements and reconstruct images for certain wavelengths that are not within the visible spectrum, and it is a much more affordable approach than producing sensor chips for these wavelengths. In this scope, the single-pixel camera and compressed sensing have great potentials as a real world engineering approach.

References

[1] S. Foucart, H. Rauhut *A mathematical introduction to compressive sensing.* Birkhäuser, New York, 2013.

[2] Y. Eldar, G. Kutyniok *Compressed sensing: theory and applications.* Cambridge University Press, 2013. (Cited on p. 88.)

[3] M. Hyder, K. Mahata *An Approximate L0 Norm Minimization Algorithm for Compressed Sensing* 2009 IEEE International Conference on Acoustics, Speech and Signal Processing. 2009. (Cited on p. 1.)

[4] U. Ayaz, H. Rauhut *Nonuniform Sparse Recovery with Subgaussian Matrices* arXiv:1007.2354, 2001. (Cited on p. 2.)

[5] A.J. Jerri *The Shannon sampling theorem–its various extensions and applications: A tutorial review.* Proc. IEEE. 1565–1596 1977.

[6] E.J. Candes, J. Romberg, T. Tao, *Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information* IEEE Transactions on Information Theory, 2006.

[7] M. Duarte, et al. *Single-Pixel Imaging via Compressive Sampling* IEEE Signal Processing Magazine, 2008.

[8] D. Mackenzie *Compressed sensing makes every pixel count* Mackenzie2009CompressedSM, 2009.