Compressive Sampling

Zhenjie Qu*

August 2021

Contents

*Supervised by Professor Nicholas Kevlahan

Abstract

In science and engineering, we need to convert the physical world into data which can be analyzed and processed to obtain information. In real life, all images, audio or video are determined as a signal which can be analyzed and modified by signal processing. The traditional technique of signal processing is well known as the Nyquist-Shannon sampling theorem. However, as the result of the high Nyquist rate and a large number of samples collected, a new technique called compressive sampling has been introduced which can acquire and reconstruct the signal using fewer samples than Shannon's theorem. The aim of this paper is to introduce the basic concepts, algorithm, and application of compressive sampling.

1 Introduction

In this information age, with the rapid development of digital revolution, numerous images and audio comprised with a large amount of data needs to be stored and transmitted in our computers or mobile phones. The general method of acquiring and reconstructing a signal such as images and audio is applying Shannon-Nyquist sampling theory. Shannon sampling theorem is the conventional way of sampling signals which asserts that the sampling rate must be at least twice the maximum frequency present in the signal [5][9].

Theorem 1 *If a continuous time signal contains no frequency components higher than W Hz, then it can be completely determined by uniform samples taken at a rate f_s samples per second where*

$$f_s \geq 2W$$

or, in term of the sampling period

$$T \leq \frac{1}{2W}$$

The idea of the Nyquist rate is at least sampling twice every second on time-frequency graph. For example, if a 10000-seconds signal is in the form of a single sine wave, we need to take at least 20000 samples for the successful recovery of the signal. In other words, in the case of discrete finite-dimensional signal, the number of measurements we sampled (m) should be at least as large as the size of the signal (n) to ensure the signal be reconstructed perfectly. Figure 1 is an example of sampling a signal formed by a single sine wave using Nyquist rate. Figure 1(a) is a sine wave with amplitude =10 and frequency=10 Hz; Figure 1(b) shows when sampling frequency is 25 Hz, we can perfectly reconstruct the signal and dash line is to fit sine wave; in figure 1(c) the sampling frequency is 15 Hz, which disobey the theorem, so we fail to do the recovery. When applying this theorem, it's always complicated and costly to build a device to store a thousand of data. A technique called compressive sampling(also called compressed sensing) was initiated around 2006 by original papers from Candes,

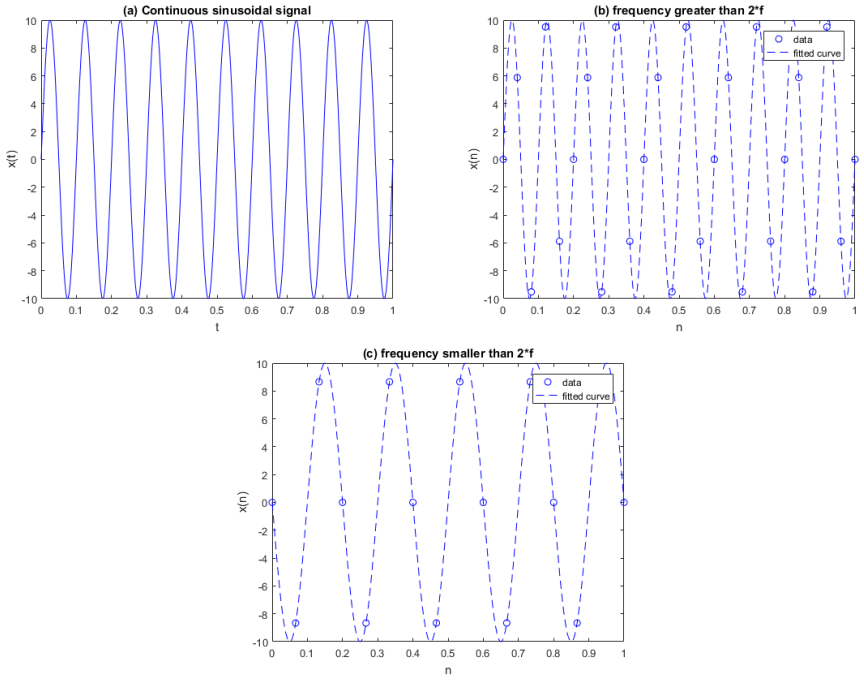

Figure 1: Nyquist sampling

Romberg, Tao and Donoho [5][2].The main idea of compressed sampling is to reconstruct the original data from undersampled measurements using different algorithms, such as optimization methods and greedy methods. The number of samples we obtain in compressive sampling is $S \log n$ where S is a constant and much smaller than n. Obviously, instead of using the traditional way, we have a huge decrease in the number of samples. Nowadays, this method is widely used in applied mathematics, engineering, biology, photography and other different areas. The structure of this paper will be separated as follows: In "Introduction" section I compare compressive sampling with Shannon-Nyquist Theorem, "Basic theory and developments" section will discuss about what needs for compressive sampling, then "Algorithm" section will introduce one of the algorithm–convex optimization problem, "Application" section will talk about "Magnetic resonance imaging" and finally I will make a conclusion about this paper.

2 Basic theory and developments

2.1 Sensing problem

In mathematical expression, any signal can be expressed as vectors $f \in \mathbb{R}^n$. We are interested in taking a small number samples m from the data f to obtain the observed data(also called coefficient; sampled values)$y \in \mathbb{R}^m$ via an $m \times n$ "sensing matrix" A:[2]

$$y = Af$$

where A is comprised of the vectors $\{\varphi_k\}_{k=1}^m \in \mathbb{R}^n$ as rows. $\{\varphi_k\}_{k=1}^n \in \mathbb{R}^n$ is sensing basis vectors, and can be also denoted as an $n \times n$ full matrix Φ which has n basis vectors. Clearly, sensing matrix A takes m out of n basis vectors from the big Φ matrix. Matrix A has m rows and matrix Φ has n rows. If the reconstructed signal is f^*, our goal is to minimize $\|f - f^*\|_2$, which means we want to recover the signal with much less perceptual loss. Since $m << n$, it is an underdetermined linear system, which means there are fewer equations than unknowns. To solve this ill-posed problem, the algorithm is [5]

$$\min \|x^*\|_1 \quad \text{subject to} \quad y = Af, \text{ where } f^* = \Psi x^*$$

$x^*$is the sparse solution of linear system $f^* = \Psi x^*$ which will be explained in detail in sparsity section. And the approximation f* is new signal after reconstruction. This algorithm will be discussed in algorithm section.

2.2 Sparsity and Compressibility

Compressive sampling is also called sparse recovery and exploits signal's sparsity to achieve signal processing. A signal $f \in \mathbb{R}^n$ is sparse means its coefficients x_i is zero or close to zero in Ψ domain. The expression of vector f can be written as follows:[5]

$$f = \sum_{i=1}^{n} x_i \psi_i$$

where $\Psi = \{\psi_1, \psi_1 .... \psi_n\}$ is an orthonormal basis such as Fourier basis and wavelet basis. For example, the signal in Figure 2(a) and its FFT transform in (b). Figure 2(a) is a signal consisted by two sine waves and some random noise, which only displays $0 - 50 millionseconds$ of original data for showing the trend of sine wave. Figure 2(b) is FFT transform of this signal. The matrix notation is $f = x\Psi$, Ψ is an nxn matrix. The signal is also called S-sparse (S means at most S non-zero entries in coefficients x_i). In mathematical expression, sparsity can be represented using l_0 norm[11]:

Definition 1 *The support of a vector $x \in \mathbb{C}^N$ is the index set of its nonzero entries, i.e.,*

$$supp(x) := \{j \in [N] : x_j \neq 0\}$$

The vector $x \in \mathbb{C}^N$ is called s-sparse if at most s of its entries are nonzero, i.e., if

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

This notation is also justified by the following[1]:

$$\|x\|_0 = \lim_{p \to 0} \|x\|_p^p \quad \text{for all} \quad x \in \mathbb{R}^n$$

According to the above definitions, we could obtain the number of non-zero elements in coefficients $\{x_i\}$. After keeping only the S largest coefficients and setting the small coefficients to zero, we obtain a new vector x_S, and the approximation of f is defined as $f_S = x_S \Psi$. For example in figure 2(b), we can capture two important largest coefficients for recovery and ignore the rest of small values. Sparse signals are often called compressible. The good compressibility of a signal depends on whether its coefficients decays more quickly. If we arrange the coefficients x_i in descending order as $|x_1| \geq |x_2| \geq \ldots \geq |x_n|$ [8], they decay with a power law if there exists constants C_1 and $q > 0$ such that $|x_i| \leq C_1 i^{-q}$. When q is larger, decay is faster and signal is more compressible. That is to say, the coefficient x_i is well-approximated by x_S and also the original signal f is well-approximated by f_S. The best S-term approximation is keeping the largest S coefficients and setting the rest as zero, then minimizing error between the original coefficient x and approximation x_S, defined as:[8]

Definition 2 *For $p > 0$, the l_p-error of best S-term approximation to a vector $x \in \mathbb{C}^n$ is defined by:*

$$\sigma_s(x)_p := inf \|x - x_S\|_p$$

After we rearrange the coefficients from largest one to smallest one, we can imagine a decay curve with S as x-axis and coefficients x_i as y-axis. If S is larger, the error $\|x - x_S\|$ is smaller. And if the largest coefficients x_S is fewer, the signal compresses even better. When the error of this best s-term approximation decays quickly, the vector is compressible. Thus, we can define $\sigma_s(x)_p$ with power law as[8]

$$\sigma_s(x)_p \leq C_2 K^{1/2-s}$$

Note:$\sigma_s(x)_2$ will decay as K^{-r} if and only if the sorted coefficients x_i decay as $i^{-r+1/2}$

2.3 Incoherence and Random incoherent sampling theorem

Initially, the definition of coherence in mathematics is

Definition 3 *In linear algebra, the coherence or mutual coherence of a matrix A is the maximum value of the cross correlations between the columns of*

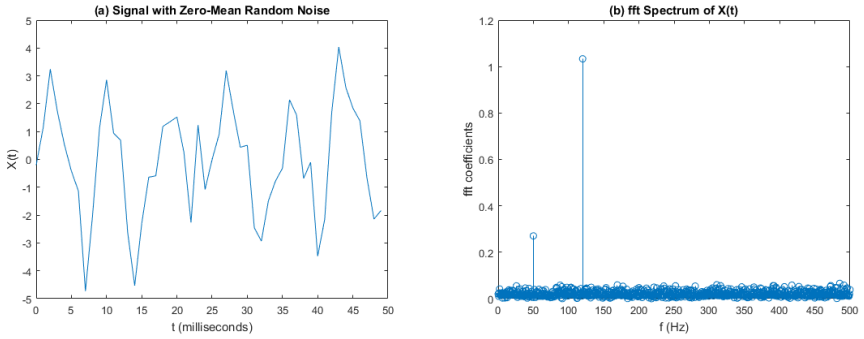

Figure 2: fft transform

A, which is denoted by $a_1, a_2, \ldots, a_n$*..Assume the vectors in the column are normalized*($\|a_i\|_2 = 0$)*, the mutual coherence of A is then defined as: [12]*

$$M = \max_{1 \le i \ne j \le m} | < a_i, a_j > |$$

, where cross relation means a measure of similarity of two functions of the displacement of one relative to the other in signal processing.

Similarly, in compressive sampling, the key idea of compressive sampling is to measure how the sensing basis Φ and representation basis Ψ different from each other. The expression stated as following:[5]

Definition 4 *The coherence between then sensing basis* Φ *and representation basis* Ψ *is:*

$$\mu(\Phi, \Psi) = \sqrt{n} \max_{1 \le i \ne j \le n} | < \phi_i, \psi_j > |$$

where $\mu(\Phi, \Psi) \in [1, \sqrt{n}]$. In compressed sampling, our expectation is to maximize the incoherence, which means sensing matrix should be largely distinguish from the sparsity basis. In this case, if the inner product of ϕ and ψ is $1/\sqrt{n}$, the incoherence is maximized, while as the inner product closes to one, coherence increases. The key role of coherence is shown in random incoherent sampling theorem[5]:

Theorem 2 *Fix* $f \in \mathbb{R}^n$ *and suppose that the coefficient sequence* x *of* f *in the basis* Ψ *is S-sparse. Select* m *measurements in the* Φ *domain uniformly at random. Then if*

$$m \ge C \cdot \mu^2(\Phi, \Psi) \cdot S \cdot logn$$

for some positive constant C, the solution to

$$\min \|x^*\|_1 \quad subject\ to \quad y = A\Psi x^*$$

is exact with overwhelming probability.(It is shown that the probability of success exceeds $1 - \delta$ *if* $m \ge C \cdot \mu^2(\Phi, \Psi) \cdot S \cdot logn$*)*

Based on this theorem, it's obvious that if $\mu(\Phi, \Psi) = 1$, we need the fewest number of samples m. The smaller coherence is, the fewer samples we needed. However, if $\mu(\Phi, \Psi) = \sqrt{n}$, we may not successfully recover our signal from a limited number of samples. Furthermore, we can notice that in this time we only need the order of $m = logn$ samples for recovery rather than at least the order of n in traditional idea. It means the use of compressive sampling is much more efficient than Nyquist sampling.
As stated in this theorem, we should select the measurements randomly .Random sampling is an appropriate way of sensing to recover the signal. Considering a data is a single sine wave. if we sample it using Fourier transform, we can get a graph in terms of frequency like a Dirac delta function, it means the rest of the coefficients are all zero. But this time we focus on discrete data, so when we do sampling, this single non-zero value might be missed. Since we don't know the priori knowledge for whether the basis we choose is the proper basis which we can exactly find all non-zero coefficients for signal recovery, the best solution for this problem is to sample the data randomly using random matrix such as Gaussian matrix and Bernoulli matrix[6]. So randomness is one of the major conditions of compressive sampling. The more detailed information about random matrices will be discussed later.

2.4 Setup of sensing matrix

In sensing problem section, the sampled value $y \in \mathbb{C}^n$ is denoted by $y = R\Phi f$ where Φ is the big $n \times n$ matrix which is comprised of a set of vectors $\{\varphi_k\}_{k=1}^n \in \mathbb{R}^n$, and R refers to an $m \times n$ matrix that randomly select m sensing vectors φ_k from matrix Φ, thus we get a random matrix $B = R\Phi$. So this time, according to the importance of randomness, we select m vectors from n basis vectors $\{\varphi_k\}$ at random instead of directly select the first m vectors to form a sensing matrix A previously. Then we know sparse signal $f = \Psi x$ where Ψ is a representation basis. After combining this two linear system, $y = B\Psi x = R\Phi\Psi x = Ax$, $A = R\Phi\Psi$[5] Here we should recover our signal f by finding x in $y = Ax$ via solving convex optimization problem.

2.5 Null Space Property

The condition of sensing matrix A plays an important role to successfully recover our sparse signal f from measurements y. We should consider the null space of matrix A [8] :

$$N(A) = \{z : Az = 0\}$$

For any different pair of S-sparse vectors $f, f' \in \Sigma_S = \{f : \|f\|_0 \leq S\}$, where Σ_S denotes the set of all S-sparse signals. We must have $Af \neq Af'$, then $A(f - f') \neq 0$ with $(f - f') \in \Sigma_{2S}$, where Σ_{2S} denotes the set of all $2S$-sparse vectors. Since if $Af = Af'$,we get $y = Af = Af'$, then we therefore have $A(f - f') = 0$ with $(f - f') \in \Sigma_{2S}$. This should be avoided because we want every different signals f, f' must corresponds to unique measurement value y.

Otherwise, it's hard to reconstruct the signal successfully. So $N(A)$ cannot contain vectors in Σ_{2S}. This property can be identified using *spark*[8]:

Definition 5 *The spark of a given matrix A is the smallest number of columns of A that are linearly dependent.*

Theorem 3 *For any vector $y \in \mathbb{R}^n$, there exists at most one signal $f \in \Sigma_S$ such that $y = Af$ if and only if $spark(A) > 2S$.*

However, just few real-world signals are truely sparse, most of the signals are approximately sparse. *Spark* can be only applied to truely signal, so we should consider a more restrictive condition called null space property[7]:

Definition 6 *An $m \times n$ complex matrix A has the nullspace property of order s, if for all index sets S with $s = |S| \leq n$, we have that: $\|\eta_S\|_1 < \|\eta_S c\|_1$ for all $\eta \in kerA \setminus \{0\}$.*

The null space property is both necessary and sufficient for establishing guarantees for recovery. But these guarantees do not account for noise.[8]

2.6 Restricted Isometry Property

In realistic situation, it is impossible to successfully recover a signal without the presence of noise. Thus compressive sampling should deal with both approximately signal and with noise. According to the uniform uncertainty principle(UUP) [6] [4], $M \times N$ sensing matrix A in $y = Ax$ obeys a "restricted isometry hypothesis:

Definition 7 *Uniform Uncertainty principle(UUP): we say that a measurement matrix F_Ω obeys the uniform uncertainty principle with oversampling factor λ if for every sufficiently small $\alpha > 0$, the following statement is true with probability at least $1-O(N^{-\frac{\rho}{\alpha}})$ for some fixed positive constant $\rho > 0$: for all subsets T such that*

$$|T| \leq \alpha \cdot M/\lambda$$

the marix $A_{\Omega T}$ obeys the bounds

$$\frac{1}{2} \cdot \frac{M}{N} \leq \lambda_{min}(A^*_{\Omega T} A_{\Omega T}) \leq \lambda_{max}(A^*_{\Omega T} A_{\Omega T}) \leq \frac{3}{2} \cdot \frac{M}{N}$$

It is equivalent to the inequality

$$\frac{1}{2}\frac{M}{N}\|f\|^2_{l_2} \leq \|A_\Omega f\|^2_{l_2} \leq \frac{3}{2}\frac{M}{N}\|f\|^2_{l_2}$$

holding for all signals f with support size less or equal to $\alpha A/\lambda$

where A_Ω is an abstract measurement matrix. It is a random $|\Omega| \times N$ matrix following some probability distribution (e.g. the Gaussian, Bernoulli, or Fourier ensembles).We also allow the number of measurements $|\Omega|$ to be a random variable taking values between 1 and N, and set $M := E(|\Omega|)$ be the expected

number of measurements. Matrices $F_{\Omega T} := F^*_{\Omega T} : l_2(T) \to l_2(\Omega)$ which are the $|\Omega|$ by $|T|$ matrices obtained by extracting $|T|$ columns from F_Ω[4].
Restricted Isometry Property is a condition of sensing matrix A, which stated as[5]:

Definition 8 *Restricted Isometry property(RIP): Let A_T, $T \subset \{1, \dots N\}$ be the $M \times |T|$ submatrix obtained by extracting the columns vector of A corresponding to the indices in T. For each integer $S = 1, 2, \dots$, we define the S-restricted isometry constant δ_S of the matrix A which is the smallest quantity such that*

$$(1 - \delta_S)\|c\|^2_{l_2} \leq \|A_T c\|^2_{l_2} \leq (1 + \delta_S)\|c\|^2_{l_2}$$

for all subsets T with $|T| \leq S$ and coefficient sequences $(c_j)_{j \in T}$, where constants $\delta_S \in (0, 1)$.

In our CS problem, matrix A has the restricted isometry property if the isometry constant δ_S is not too close to one. This means the best situation is $\delta_S = 0$, then we get $\|Ax\|_2 = \|x\|_2$, sensing matrix A perfectly preserves the l_2 norm of S-sparse signals this time. The worst situation is obviously $\delta_S = 1$, we get $0 \leq \|Ax\|_2 \leq \sqrt{2}\|x\|_2$ here, which means matrix A has no control of l_2 norm. If $\delta_S = 0.2$, we get $0.8\|x\|^2_2 \leq \|Ax\|^2_2 \leq 1.2\|x\|^2_2$, matrix A approximately preserves the l_2 norm of S-sparse signals. After applying sensing matrix, the l_2 norm has changed, but this change is acceptable, which is between 0.8 and 1.2. Thus, when the upper and lower bound of $\|Ax\|^2_2$ is much closer to $\|x\|^2_2$, matrix A will better control the l_2 norm.
We know each degree of sparsity corresponds to one δ_S, the special case is when $S = 1$, there is only one non-zero coefficient. Let $\{v_i\}$ be the column vectors of sensing matrix A, We get

$$1 - \delta_1 \leq \|v_i\|^2 \leq 1 + \delta_1$$

Since the value δ_1 only bring us the magnitude information about vectors $\{v_i\}$, δ_1 is the best isometry constant. Specially, $\delta_1 = 0$ if and only if all of the vector $\{v_i\}$ have unit length.[6]
Moreover, another important way to describe RIP is to say that the columns of the sensing matrix A are nearly orthogonal. For example we know the best case is when $\delta_S = 0$, then $\|Ax\|_2 = \|x\|_2$, A is obviously a orthonormal basis. Next, we will discuss about restricted orthogonality constants,[3]

Definition 9 *we define the S, S'-restricted orthogonality constants $\theta_{S,S'}$ for $S + S' \leq N$ to be the smallest quantity such that*

$$| < A_T x, A_T x' > | \leq \theta_{S,S'} \cdot \|x\|_{l_2}\|x'\|_{l_2}$$

holds for all disjoint sets $T, T' \subseteq \{1, \dots N\}$ of cardinality $|T| \leq S$ and $|T'| \leq S'$ where $\theta_{S,S'} \geq 0$

Similar to the expression of inner product: $< u, v >= cos\theta \cdot \|u\|\|v\|$, the constant $\theta_{S,S'}$ is the cosine of the smallest angle between the two subspaces spanned by

the columns in T and T'[6]. According to the definition of orthogonality of two vectors in an inner product space, smaller value of $\theta_{S,S'}$ also indicates the column vectors of matrix A behaves more similarly to an orthonormal system. Now, let's consider the $2S$-restricted isometry constant δ_{2S} which is the smallest quantity such that[5]:

$$(1-\delta_{2S})\|x_1-x_2\|_{l_2}^2 \leq \|A_T x_1 - A_T x_2\|_{l_2}^2 \leq (1+\delta_{2S})\|x_1-x_2\|_{l_2}^2$$

for all subsets T with $|T| \leq 2S$ and $x_1 - x_2 \in \Sigma_{2S}$, where Σ_{2S} denotes the set of all $2S$-sparse vectors. If $\delta_{2S} < 1$ for example $\delta_{2S} = 0.2$, we get $0.8\|x_1-x_2\|_{l_2}^2 \leq \|A_T x_1 - A_T x_2\|_{l_2}^2 \leq 1.2\|x_1-x_2\|_{l_2}^2$, it means all pairwise distances between S-sparse signals are well-preserved. However, if $\delta_{2S} = 1, 0 \leq \|A_T x_1 - A_T x_2\|_{l_2} \leq \sqrt{2}\|x_1-x_2\|_{l_2}$. This also implies $2S$-sparse vectors cannot be in the null space of A, since as we discussed in null space condition section, the necessary condition is $Ax_1 \neq Ax_2$ and $A(x_1-x_2) \neq 0$, sparse vectors x_1, x_2 must be distinguished from each other for the unique y.

2.7 Noiseless data recovery

If RIP holds, then the l_1 norm minimization problem gives an accurate reconstruction:[5]

$$\min \|x^*\|_1 \quad \text{subject to} \quad y = Ax^*$$

where $A = R\Phi\Psi$.

Theorem 4 *[5]Assume that $\delta_{2S} = \sqrt{2}-1$. Then the solution x^* to l_1 minimization problem obeys*

$$\|x^* - x\|_{l_2} \leq C_0 \cdot \|x - x_S\|_{l_1}/\sqrt{S}$$

and

$$\|x^* - x\|_{l_1} \leq C_0 \cdot \|x - x_S\|_{l_1}$$

for some constant C_0, where X_S is the vector x with all but the largest S ones setting to zero.

Where $\|x^*-x\|$ means the error in solution of optimization problem, and $\|x-x_S\|$ refers to the error in sparse representation. Here we can minimize the error $\|x^* - x\|$ by minimizing the S-term approximation error $\|x - x_S\|$, it is also called best S-term approximation in "Sparsity" section before. In other words, we are interested in how well we can recover a nearly sparse signal, and the recovery error is not much worse than $C_0\|x-x_S\|$. Firstly, [6]we should arrange the elements of coefficient x_i in decreasing order of magnitude, the ith entry obeys

$$|x| \leq R \cdot i^{-1/p}, 1 \leq i \leq N$$

x_i decays like a power law, when p is smaller, the decay is faster. The best S-term approximation of x_i shows

$$\|x - x_S\|_{l_2} \leq C_2 \cdot R \cdot S^{1/2-1/p}$$

in the l_2 norm, and

$$\|x - x_S\|_{l_1} \leq C_1 \cdot R \cdot S^{1-1/p}$$

in the l_1 norm, where C_1, C_2, R are constants.

2.8 Random matrices

Suppose A is a random matrix, such as [11]Gaussian matrix whose entries are formed by normal distribution with mean 0 and variance $1/m$ and Bernoulli matrix where the entries are independent random variables taking the values ± 1 with equal probability. For example, one of the application in compressive sampling is single-pixel camera. The idea is to recover the real-world image f from the random measurements y using a Bernoulli random matrix. [11]In single-pixel camera, there is a micro mirror array consisting of a large number of small mirrors. The light of the image is reflected at this mirror array and a lense combines all reflected light on the sensor, single pixel of the camera. Now, if these random matrices A satisfies RIP, the number of samples we need to take for reconstructing the signal is[5]

$$m \geq C \cdot Slog(\frac{n}{S})$$

where C is some constants. This is a sharp lower bound since there can be no other reconstruction algorithm that can recover the data with a lot fewer samples.[5]. Alternatively, with $A = R\Phi\Psi$ the samples we take for recovery with large probability is

$$m \geq C \cdot S(logn)^4$$

Although this time we take about $S(logn)^4$ samples instead of $Slog(\frac{n}{S})$, the number of samples are much less than the data length n.

2.9 Robust compressive sampling and Noisy data recovery

As mentioned before, in realistic application, we cannot completely reconstruct an object from a few number of samples without the effect of noise. So now we consider to recover a vector $x \in \mathbb{R}^n$ from data[5]

$$y = Ax + z$$

where $A = R\Phi\Psi$ is a sensing matrix, z is a stochastic or deterministic unknown error term. And l_1 minimization problem for noisy data is[5]:

$$\min \|x^*\|_{l_1} \quad \text{subject to} \quad \|Ax^* - y\|_{l_2} \leq \epsilon$$

where ϵ bounds the amount of noise in the data.

Theorem 5 *[5]Assume that $\delta_{2S} = \sqrt{2} - 1$. Then the solution x^* to above l_1 minimization problem obeys*

$$\|x^* - x\|_{l_2} \leq C_0 \cdot \|x - x_S\|_{l_1}/\sqrt{S} + C_1 \cdot \epsilon$$

for some constants C_0 and C_1. This time the reconstruction error is bounded by the error which would occur if one had noiseless data, and another error which is just proportional to the noise level.

3 Algorithm

3.1 Convex Optimization Problem

There are many algorithms of compressive sensing, such as Optimization methods, Greedy methods and Thresholding-Based methods.[11]. The most common algorithm of compressive sampling is called basis pursuit, which is exactly l_1 minimization problem mentioned before. In the real life, we concentrate on recovering the signal from a limited number of measurements with additive noise, this model is given before:

$$y = Ax + z$$

where matrix A is a sensing matrix, x is set of coefficient vectors and z is an unknown error term. Generally, our goal is to find the smallest number of non-zero entries in sparse coefficients x from this system to recover the data. The first algorithm comes to our mind is l_0-minimization. According to the definition of norm, $\|x\|_0$ means the number of non-zero entries of a coefficient x. The our optimization problem should be given as following:

$$\min \|x\|_0 \quad \text{subject to} \quad \|Ax - y\|_2 \leq \epsilon$$

where ϵ bounds the amount of noise in the data: $\|z\|_2 \leq \epsilon$. Unfortunately, l_0-minimization turns out to be NP hard. It is a computational problem which is not feasible when data size n is very large. l_0 minimization is an exponential-time algorithm to solve a problem of size n which means $e^{\alpha n}$.[11]However, we need polynomial-time algorithm to compute the optimization problem in a time that grows like n^{α}. Thus we use l_1 norm minimization:[13]

$$\min \|x\|_1 \quad \text{subject to} \quad \|Ax - y\|_2 \leq \delta$$

where the tolerance δ is related to the error bound ϵ. The substitution from l_0 quasi-norm to l_1 norm is referred to as convex relaxation[13].Alternatively, it is found closely to the convex problem:[13]

$$min \frac{1}{2}\|Ax - y\|_2^2 + \gamma\|x\|_1$$

where the parameter γ manages a tradeoff between the approximation error and the sparsity of the coefficient vector.

4 Application

Magnetic Resonance Imaging (MRI) is an essential medical imaging technique used in radiology to form pictures of the human body. It has an inherently slow

data acquisition process. The application of compressive sampling to MRI can significantly reduce the scan time which is beneficial to patients and health care economics[10]. Here MRI obeys two important requirements to successfully apply for compressive sampling: one is medical imagery is naturally compressible by sparse coding in an appropriate transform domain (e.g., by wavelet transform), another one is MRI scanners naturally acquire encoded samples, rather than direct pixel samples (e.g., in spatial-frequency encoding).[10]. The most common method of image reconstruction for MRI is basis pursuit in which convex optimization problem is solved.

5 Conclusion

The development of compressive sampling greatly benefits the signal processing and this technique is widely in different areas. Compressed sensing is a much better way to acquire and reconstruct signals from far fewer samples than traditional method. The necessary conditions for signal recovery are sparseness, incoherence, randomness and optimization. Then compressive sampling should be always applied to a nearly signal with additive noise in the real world. Sensing matrix should obey RIP for accurate reconstruction of signal.

6 Acknowledgement

This report was done under the supervision of Dr. Nicholas Kevlahan on the topic of compressive sampling. Very thankful to Dr. Nicholas Kevlahan for teaching me and supporting me during this 6 weeks.

References

[1] Holger Boche et al. "A survey of compressed sensing". In: *Compressed sensing and its applications*. Springer, 2015, pp. 1–39.

[2] Emmanuel Candes and Justin Romberg. "Sparsity and incoherence in compressive sampling". In: *Inverse problems* 23.3 (2007), p. 969.

[3] Emmanuel J Candes and Terence Tao. "Decoding by linear programming". In: *IEEE transactions on information theory* 51.12 (2005), pp. 4203–4215.

[4] Emmanuel J Candes and Terence Tao. "Near-optimal signal recovery from random projections: Universal encoding strategies?" In: *IEEE transactions on information theory* 52.12 (2006), pp. 5406–5425.

[5] Emmanuel J. Candes and Michael B. Wakin. "An Introduction To Compressive Sampling". In: *IEEE Signal Processing Magazine* 25.2 (2008), pp. 21–30. DOI: 10.1109/MSP.2007.914731.

[6] Emmanuel J Candès et al. "Compressive sampling". In: *Proceedings of the international congress of mathematicians*. Vol. 3. Madrid, Spain. 2006, pp. 1433–1452.

[7] Albert Cohen, Wolfgang Dahmen, and Ronald DeVore. "Compressed sensing and best -term approximation". In: *Journal of the American mathematical society* 22.1 (2009), pp. 211–231.

[8] Mark A Davenport et al. *Introduction to compressed sensing*. 2012.

[9] Edmund Lai. "2 - Converting analog to digital signals and vice versa". In: *Practical Digital Signal Processing*. Ed. by Edmund Lai. Oxford: Newnes, 2003, pp. 14–49. ISBN: 978-0-7506-5798-3. DOI: https://doi.org/10.1016/B978-075065798-3/50002-3. URL: https://www.sciencedirect.com/science/article/pii/B9780750657983500023.

[10] Michael Lustig et al. "Compressed sensing MRI". In: *IEEE signal processing magazine* 25.2 (2008), pp. 72–82.

[11] Joel A Tropp. "A mathematical introduction to compressive sensing [Book Review]". In: *Bulletin of the American Mathematical Society* 54.1 (2017), pp. 151–165.

[12] Joel A Tropp. "Just relax: Convex programming methods for identifying sparse signals in noise". In: *IEEE transactions on information theory* 52.3 (2006), pp. 1030–1051.

[13] Joel A Tropp. "Just relax: Convex programming methods for identifying sparse signals in noise". In: *IEEE transactions on information theory* 52.3 (2006), pp. 1030–1051.

7 Appendix A

This section includes the images of my code.

```
%%orignal signal
amplitude=10; % Amplitude
fm=10;        % Frequency
t=0:0.001:1;
xa1=amplitude*sin(2*pi*fm*t);
plot(t,xa1,'b');
title('(a) Continuous sinusoidal signal');
xlabel('t');
ylabel('x(t)');

%%frequency greater than 2*fm
fs1=25;
n=0:1/fs1:1;
xa2=amplitude*sin(2*pi*n*fm);
figure
[fitresult, gof] = createFit(n, xa2);
title('(b) frequency greater than 2*f');
xlabel('n');
ylabel('x(n)');
axis([0 1 -10 10]);

%%frequency smaller than 2*fm
fs2=15;
n1=0:1/fs2:1;
xa3=amplitude*sin(2*pi*n1*fm);
figure
[fitresult, gof] = createFit(n1, xa3);
title('(c) frequency smaller than 2*f');
xlabel('n');
ylabel('x(n)');
axis([0 1 -10 10]);
```

```
function [fitresult, gof] = createFit(n1, xa3)
[xData, yData] = prepareCurveData( n1, xa3 );
% Set up fittype and options.
ft = fittype( 'sin1' );
opts = fitoptions( 'Method', 'NonlinearLeastSquares' );
opts.Display = 'Off';
opts.Lower = [-Inf 0 -Inf];
% Fit model to data.
[fitresult, gof] = fit( xData, yData, ft, opts );
% Plot fit with data.
plot(fitresult,'--b',xData,yData,'ob');
```

```
%%constants
Fs = 1000;
T = 1/Fs;
L = 1500;
t = (0:L-1)*T;

%%generate signal
S = 0.5*sin(2*pi*50*t) + 2*sin(2*pi*120*t);
X = S + randn(size(t));

%%plot the original signal
plot(1000*t(1:50),X(1:50))
title('(a) Signal with Zero-Mean Random Noise')
xlabel('t (milliseconds)')
ylabel('X(t)')

%%fft and plot
Y = fft(X);
P2 = abs(Y/L);
P1 = P2(1:L/2+1);
f = Fs*(0:(L/2))/L;
figure
stem(f,P1) ;
title('(b) fft Spectrum of X(t)')
xlabel('f (Hz)')
ylabel('fft coefficients')
```