Sparse Modeling and Compressive Sensing

Ivan Ralašić
University of Zagreb, Faculty Of Electrical Engineering and Computing
ivan.ralasic@fer.hr

***Abstract*—Sparse modeling and compressive sensing are novel methods for signal representation and acquisition. Sparse signal representations are manifestation of the parsimony principle also known as the Occam's razor which states that the simplest and most concise explanation of a natural phenomenon is in most cases the best one possible. Sparse structure appears to be an inherent property of many natural signals. Compressive sensing represents a signal acquisition technique that exploits underlying sparse signal structure and enables accurate signal recovery from an incomplete set of measurements. In this overview, we will cover theory of basis representation, sparse signal decomposition and sparse recovery problem formulations and bridge it with different practical applications of compressive sensing framework.**

***Index Terms*—sparse modeling, compressive sensing, basis representation, signal processing, signal acquisition, optimization.**

I. Introduction

FUNDAMENTAL problem in sparse modeling and compressive sensing is to obtain accurate recovery of an unobserved high-dimensional signal from a reduced number of measurements. Traditional approach to signal acquisition is based on the classical Shannon-Nyquist theorem [1], [2] stating that in order to preserve information about a signal, one must sample the signal at a rate which is at least twice the signal's bandwidth, defined as the highest frequency in the signal's spectrum. Traditional sample-and-compress framework is efficient and used in many real-world applications. However, the fact that we are able to compress the acquired data suggests that Shannon-Nyquist theorem is too pessimistic and it does not take advantage of any specific underlying structure that the signal may possess. In practice, the traditional sampling scheme produces tremendous number of samples and it must be followed by a compression step in order to store and process obtained information successfully. The compression step uses different basis representations to obtain concise signal representation and it essentially keeps only the significant basis coefficients while disregarding the others. The described procedure is known as transform coding and it takes advantage of underlying sparse signal structure. A natural question arises, can we combine the acquisition and compression into one-step process which will enables us to make signal acquisition more efficient.

Candés, Romberg and Tao considered signal reconstruction from an incomplete frequency samples in [3] and presented a novel kind of nonlinear sampling theorem. They state that exact signal recovery from an incomplete set of measurements is possible by solving a convex optimization problem under certain constraints. Donoho [4] defined compressed sensing on an example of arbitrary unknown vector $\boldsymbol{x} \in \mathbf{R}^n$ (digital image or signal). If $\boldsymbol{x}$ is known to be compressible by transform coding with a known transform, and we use nonlinear reconstruction procedure to obtain signal reconstruction, the number of measurements m can be dramatically smaller than the signal dimensionality n. This led to considerable research interest in the signal processing community and numerous papers followed [5]–[8].

Compressive sensing gained more popularity with the first practical applications. Compressed sensing MRI study by Lustig et al. [9] reviewed the requirements for successful compressive sensing reconstruction and described their natural fit to MRI. CS-MRI offered significant scan time reductions with benefits for patients and health care economics. In [10], authors present a new approach to building simpler, smaller, and cheaper digital cameras that can operate efficiently across a broader spectral range the conventional cameras. Their approach fuses a new camera architecture based on a digital micro-mirror device (DMD) with the new compressive sensing mathematical framework and instead of measuring pixel samples of the scene, it measures inner products between the scene and a set of random test functions. This leads to sub-Nyquist image acquisition that enables us to stably reconstruct an image from fewer measurements than the number of reconstructed pixels. Tropp et al. [11] proposed a new type of sampling system called a random demodulator that can be used to acquire sparse, band-limited signals. They demodulate the signal by multiplying it with a high-rate pseudo-noise sequence, which smears the tones across the entire spectrum. Then they apply low-pass anti-aliasing filter and capture the signal by sampling it at a relatively low rate. The major advantage of the random demodulator is that it bypasses the need for a high-rate ADC since demodulation is much easier to implement than high-rate sampling.

In this paper, we will provide an overview of basis representation fundamentals, sparse signal decomposition, sparse recovery problem formulations and different optimization procedures which lead to a definition of complete compressive sensing framework.

II. Sparse Signal Recovery

A. Basis Representation Fundamentals

To introduce the notion of sparsity, we rely on a basis decomposition that results in a low-dimensional representation of the observed signal. Every signal $\boldsymbol{x} \in \mathbf{R}^N$ is representable in terms of N coefficients $\{s_i\}_{i=1}^N$ in a given basis $\{\psi_i\}_{i=1}^N$ for $\mathbf{R}^N$ as:

$$\boldsymbol{x} = \sum_{i=1}^{N} \boldsymbol{\psi}_i \boldsymbol{s}_i \tag{1}$$

Arranging the ψ_i as columns into the $N \times N$ matrix $\boldsymbol{\Psi}$ and the coefficients s_i into the $N \times 1$ coefficient vector $\boldsymbol{s}$, we can write that $\boldsymbol{x} = \boldsymbol{\Psi s}$, with $\boldsymbol{s} \in \mathbf{R}^N$. We say that signal $\boldsymbol{x}$ is K-sparse in the basis $\boldsymbol{\Psi}$ if there exists a vector $\boldsymbol{s} \in \mathbf{R}^N$ with only $K \ll N$ nonzero entries such that $\boldsymbol{x} = \boldsymbol{\Psi s}$. By a compressible representation, we mean that the coefficient's magnitudes, when sorted, have a fast power-law decay. Many natural signals are sparse or compressible when observed in an appropriate basis. If we use a frame $\boldsymbol{\Psi}$ containing N unit-norm column vectors of length L with $L < N$ (i.e., $\boldsymbol{\Psi} \in \mathbf{R}^{L \times N}$), then for any vector $\boldsymbol{x} \in \mathbf{R}^L$ there exist infinitely many decompositions $\boldsymbol{s} \in \mathbf{R}^N$ such that $\boldsymbol{x} = \boldsymbol{\Psi s}$. In a general setting, $\boldsymbol{\Psi}$ is called overcomplete sparsifying dictionary [12].

B. Motivating Example

As an illustrative example, we can consider the case where our overcomplete dictionary is the union of two particular orthobases: the identity (spike) basis and the Fourier (sine) basis $\boldsymbol{\Psi} = [\boldsymbol{I} \quad \boldsymbol{F}]$ where $\boldsymbol{I}$ is $n \times n$ identity matrix and $\boldsymbol{F}$ is $n \times n$ normalized discrete Fourier matrix with entries defined with:

$$\boldsymbol{F}(m,l) = \frac{1}{\sqrt{n}} e^{j2\pi(m-1)(l-1)/n} \tag{2}$$

Identity and Fourier basis are mutually fully incoherent in the sense that it takes n spikes to build up a single sinusoid and also it takes n sinusoids to build up a single spike.

Now we can create a signal which is a mixture of spikes and sinusoids. As we know that the first half of our matrix $\boldsymbol{\Psi}$ contains spike functions and the second half corresponds to sine functions, we can construct random sparsity pattern with sparsity K for the vector $\boldsymbol{s}$ (of size $2N$) such that some of the non-zero entries fall into the first half and some in the second half of the vector, and then compute $\boldsymbol{x} = \boldsymbol{\Psi s}$ to obtain a signal which is a mixture of impulses and sinusoids.

As we said before, there is an infinite number of ways to decompose signal $\boldsymbol{x}$ using atoms of our overcomplete dictionary. The most natural way would be to use certain basis functions which correspond to previously selected basis function indices and by doing so we get the sparsest possible representation.

Another way we can get representation for $\boldsymbol{x}$ is by applying $\boldsymbol{\Psi}^*$ and by dividing the result by 2. Since $\boldsymbol{\Psi\Psi}^* = 2\boldsymbol{I}$, we get reproducing formula for $\boldsymbol{x}$:

$$\boldsymbol{x} = \frac{1}{2}\boldsymbol{\Psi}(\boldsymbol{\Psi}^*\boldsymbol{x}) \tag{3}$$

When we apply $\boldsymbol{s} = \frac{1}{2}\boldsymbol{\Psi}^*\boldsymbol{x}$ we get result that corresponds to the minimum energy decomposition of our signal into a coefficient vector that represents $\boldsymbol{x}$. Minimum energy decomposition corresponds to l_2-norm minimization. Unfortunately, minimum energy decomposition almost never yields the sparsest possible solution. The reason for this is that a vector has minimum energy when its total energy is distributed over all the coefficients of the vector. l_2-norm minimization gives us a solution that is dense, but has small values diffused over all coefficients.

In our example, we synthetically produced a signal with only K coefficient representation in our overcomplete basis $\boldsymbol{\Psi}$ and we want to find a decomposition that yields a signal decomposition that is K sparse. Since our goal of finding sparsest possible representation of our signal $\boldsymbol{x}$ over some basis $\boldsymbol{\Psi}$ is equivalent to finding the solution with the smallest number of nonzero elements in the basis coefficient vector $\boldsymbol{s}$ we will use l_0-pseudo-norm to find our solution. Sparse signal recovery can be formulated as finding minimum-cardinality solution to a constrained optimization problem. In the noiseless case, our constraint is simply $\boldsymbol{x} = \boldsymbol{\Psi s}$, while in the noisy case (assuming Gaussian noise), the solution must satisfy $\|\boldsymbol{x} - \boldsymbol{x}^*\|_2 \leq \epsilon$ where $\boldsymbol{x}^* = \boldsymbol{\Psi s}$ is the hypothetical noiseless representation and the actual representation is ϵ-close to it in l_2-norm. The objective function is the cardinality of $\boldsymbol{s}$ (number of non-zeros) which is often denoted $\|\boldsymbol{x}\|_0$ and called l_0-norm of $\boldsymbol{s}$ [13].

C. Sparse Signal Recovery

We will use the following notation in this section to formalize the sparse signal recovery: $\boldsymbol{x} = (x_1, ..., x_N) \in \mathbf{R}^N$ is an unobserved sparse signal, $\boldsymbol{y} = (y_1, ..., y_M) \in \mathbf{R}^M$ is a vector of measurements (observations), and $\boldsymbol{A} = \{a_{i,j}\} \in \mathbf{R}^{M \times N}$ is a design matrix.

The simplest problem we are going to start with is the noiseless signal recovery from a set of linear measurements, i.e., solving for $\boldsymbol{x}$ the system of linear equations:

$$\boldsymbol{y} = \boldsymbol{Ax} \tag{4}$$

It is usually assumed that $\boldsymbol{A}$ is a full-rank matrix, and thus for any $\boldsymbol{y} \in \mathbf{R}^M$, the above system of linear equations has a solution. Note that when the number of unknown variables, i.e., dimensionality of the signal, exceeds the number of observations. When $M < N$, the above system is underdetermined, and can have infinitely many solutions. In order to recover the signal $\boldsymbol{x}$, it is necessary to further constrain, or regularize the problem. This is usually done by introducing an objective function, or regularizer $R(\boldsymbol{x})$ to existing loss function. Regularizer encodes additional properties of the signal, with lower values corresponding to more desirable solutions. Signal recovery is then formulated as a constrained optimization problem:

$$\min_{\boldsymbol{x} \in \mathbf{R}^N} R(\boldsymbol{x}) \quad s.t. \quad \boldsymbol{y} = \boldsymbol{Ax} \tag{5}$$

Since we want to exploit underlying sparse structure of the observed signal, $R(\boldsymbol{x})$ can be defined as the number of nonzero elements, or the cardinality of $\boldsymbol{x}$, also called the l_0-norm.

In general, l_q-norms for particular values of q, denoted $\|\boldsymbol{x}\|_q$, or more precisely, their q-th power $\|\boldsymbol{x}\|_q^q$, are frequently used as regularizers $R(\boldsymbol{x})$ in constrained optimization problems.

For $q \geq 1$, the l_q norm, also called just q-norm of a vector $\boldsymbol{x} \in \mathbf{R}^N$ is defined as:

$$\|\boldsymbol{x}\|_q = (\sum_{i=1}^{N} |x_i|^q)^{\frac{1}{q}} \tag{6}$$

We can now observe the relation between cardinality and $\|l_q\|$-norms. The function $\|\boldsymbol{x}\|_0$ referred to as l_0-pseudo-norm of $\boldsymbol{x}$ is defined as a limit of $\|\boldsymbol{x}\|_q^q$ when $q \to 0$:

$$\|\boldsymbol{x}\|_0 = \lim_{q\to 0} \|\boldsymbol{x}\|_q^q = \lim_{q\to 0} \sum_{i=1}^{p} |x_i|^q = \sum_{i=1}^{p} \lim_{q\to 0} |x_i|^q \quad (7)$$

For each x_i, when $q \to 0$, $|x_i|^q \to I(x_i)$, the indicator function, which is 0 at $x = 0$ and 1 otherwise. Thus, $\|\boldsymbol{x}\|_0 = \sum_{i=1}^{p} I(x_i)$, which gives exactly the number of nonzero elements of vector $\boldsymbol{x}$. Using the cardinality function, we can now write the problem of sparse signal recovery from noiseless linear measurements as:

$$\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_0 \quad s.t. \quad \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \quad (8)$$

The above optimization problem is NP-hard and no known algorithm can solve it efficiently in polynomial time. Therefore, approximations have to be introduced. Under appropriate conditions the optimal solution can be recovered efficiently by certain approximate techniques.

First approach to approximation is a heuristic-based search such as greedy search. In greedy search method, one can start with a zero vector and keep adding nonzero coefficients one by one, selecting at each step the coefficient that leads to the best improvement in the objective function (greedy coordinate descent). In general, such heuristic search methods are not guaranteed to find the global optimum. However, in practice, they are simple to implement, computationally efficient and under certain conditions they are even guaranteed to recover the optimal solution.

An alternative approximation technique is the relaxation approach based on replacing an intractable objective function or constraint by a tractable one. In other words, one can either solve the exact problem approximately, or solve an approximate problem exactly. In the following section, we will discuss l_q-norm based relaxations, and show that the l_1-norm occupies a unique position among them, combining convexity with sparsity. A convex optimization problem is minimization of a convex function over a convex set of feasible solutions defined by the constraints. Convex problems are easier to solve than general optimization problems because of the important property that any local minima of a convex function is also a global one [14].

D. Convex Relaxations of Sparse Recovery Problem

We will focus on different l_q-norms as possible relaxations of l_0-norm. These functions are convex for $q \geq 1$ and non-convex for $q < 1$. For example, l_2-norm (Euclidean norm) is a natural first choice as a relaxation of l_0-norm. Our sparse recovery problem using l_2-norm writes:

$$\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_2^2 \quad s.t. \quad \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \quad (9)$$

Using l_2-norm as an objective has several advantages some of which are its convexity and thus its property that it has a unique minimum, and finally it solution is available in a closed form. The closed form solution to this problem is also known as pseudo-inverse solution of $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ when $\boldsymbol{A}$ has more columns than rows (we assume that $\boldsymbol{A}$ is full-rank, i.e. all of its rows are linearly independent). However, despite its convenient properties, l_2-norm has a serious drawback when it comes to sparse recovery. As we already stated before, the optimal solution obtained by pseudo-inverse is practically never sparse. To understand why the l_2-norm does not promote the solution sparsity while the l_1-norm does, and to understand the convexity and sparsity-inducing properties of l_q-norms in general, let us consider the geometry of a sparse recovery problem, where $\|\boldsymbol{x}\|_q^q$ replaces the original cardinality objective $\|\boldsymbol{x}\|_0$:

$$\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_q^q \quad s.t. \quad \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \quad (10)$$

Sets of vectors with same value of the function $f(\boldsymbol{x})$, i.e. $f(\boldsymbol{x}) = const$, are called the level sets of $f(\boldsymbol{x})$. For example, the level sets of $\|\boldsymbol{x}\|_q^q$ function are vector sets with same l_q-norm. A set of vectors satisfying $\|\boldsymbol{x}\|_q^q \leq r^q$ is called an l_q-ball of radius r; its "surface" (set boundary) is the corresponding level set $\|\boldsymbol{x}\|_q^q = r^q$. Note that the corresponding l_q-balls bounded by the level sets are convex for $q \geq 1$ (line segments between a pair of its points belong to the ball), and non-convex for $0 < q < 1$ (line segments between a pair of its points do not always belong to the ball).

From a geometric point of view, solving the optimization problem in Eq. 10 is equivalent to "blowing up" l_q-balls with the center at the origin, i.e., increasing their radius, starting from 0, until they touch the hyperplane $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$. The resulting point is the minimum l_q-norm vector that is also a feasible point, i.e. it is the optimal solution of sparse recovery problem.

Note that when $q \leq 1$, l_q-balls have sharp "corners" on the coordinate axis, corresponding to sparse vectors, since some of their coordinates are zero, but l_q-balls for $q > 1$ do not have this property. Thus, for $q \leq 1$, l_q-balls are likely to meet the hyperplane $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ at the corners, thus producing sparse solutions, while for $q > 1$ the intersection practically never occurs at the axes, and thus solutions are not sparse.

Within the family of $\|\boldsymbol{x}\|_q^q$ functions, only those with $q \geq 1$ are convex, but only those with $0 < q \leq 1$ are sparsity-enforcing. The only function within that family that has both useful properties is therefore $\|\boldsymbol{x}\|_1$, i.e.the l_1-norm [14]. This unique combination of sparsity and convexity is the reason for the widespread use of l_1-norms in the modern sparse signal recovery field. Optimization problem using l_1 norm writes:

$$\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_1 \quad s.t. \quad \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \quad (11)$$

III. Compressive Sensing

Compressive sensing offers a framework for simultaneous sensing and compression of finite-dimensional vectors, that relies on dimensionality reduction. In compressive sensing, the signal is not measured via standard point samples but rather through the projection by a measurement matrix $\boldsymbol{\Phi}$:

$$\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x} = \boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{s} = \boldsymbol{A}\boldsymbol{s} \quad (12)$$

where $\boldsymbol{\Phi}$ is an $N \times M$ measurement matrix and $\boldsymbol{y} \in \mathbf{R}^M$ is a set of M measurements or samples where M can be much smaller than the original dimensionality of the signal, hence

the name compressive sensing. We introduce $\boldsymbol{A} = \boldsymbol{\Phi\Psi}$ and refer to it as the design matrix for compressive sensing. The central problem of compressive sensing is reconstruction of high-dimensional sparse signal representation $\boldsymbol{x}$ from a low-dimensional linear observation $\boldsymbol{y}$, also called the measurement vector. Ideally, the measurement matrix $\boldsymbol{\Phi}$ is designed to reduce the number of measurements M as much as possible while allowing for recovery of a wide class of signals. However, the fact that $M < N$ renders the matrix rank-deficient, meaning that it has a nonempty null-space, which in turn implies that for any particular signal $\boldsymbol{x}_0 \in \mathbf{R}$, an infinite number of signals will yield the same measurements $\boldsymbol{y}_0 = \boldsymbol{\Phi x}_0 = \boldsymbol{\Phi x}$ for chosen measurement matrix $\boldsymbol{\Phi}$. The motivation behind the design of the matrix $\boldsymbol{\Phi}$ is to allow for distinct signals $(x,\, x')$ within a class of signals of interest to be uniquely identifiable using sparse optimization techniques from their measurements $(y,\, y')$, even though $M \ll N$ [12].

A. Uniqueness of Compressive Sensing Recovery Problem

In this section we will discuss when the solutions of the l_0- and l_1- norm minimization problems are unique. The main design criteria for matrix $\boldsymbol{A}$ is to enable the unique identification of a signal of interest $\boldsymbol{x}$ from its measurements $\boldsymbol{y} = \boldsymbol{Ax}$. Clearly, when we consider the class of K-sparse signals Σ_K, the number of measurements has to be $M > K$ for any matrix design, since the identification problem has K unknowns.

We will now determine properties of A that guarantee that distinct signals $\boldsymbol{x},\, \boldsymbol{x'} \in \Sigma_K, \boldsymbol{x} \neq \boldsymbol{x'}$, lead to different measurement vectors $\boldsymbol{Ax} \neq \boldsymbol{Ax'}$. In other words, we want each vector $\boldsymbol{y} \in \mathbf{R}^M$ to be matched to at most one vector $\boldsymbol{x} \in \Sigma_K$ such that $\boldsymbol{y} = \boldsymbol{Ax}$.

A key relevant property of the matrix in this context is its spark [15]. Given an $M \times N$ matrix $\boldsymbol{A}$, its spark $spark(\boldsymbol{A})$, is defined as the minimal number of linearly dependent columns. Spark is closely related to the Kruskal's rank $krank(\boldsymbol{A})$ defined as the maximal number k such that every subset of k columns of the matrix $\boldsymbol{A}$ is linearly independent [16]. We can now write the relation between *spark* and *krank* as:

$$spark(\boldsymbol{A}) = krank(\boldsymbol{A}) + 1 \quad \text{and} \quad rank(\boldsymbol{A}) \geq krank(\boldsymbol{A}) \tag{13}$$

By definition, the vectors in the null-space of the matrix $\boldsymbol{Ax} = 0$ must satisfy $\|x\|_0 \geq spark(\boldsymbol{A})$, since these vectors combine linearly columns from $\boldsymbol{A}$ to give the zero vector, and at least *spark* such columns are necessary by definition. Sparse recovery solution uniqueness via spark can be stated as: if $spark(\boldsymbol{A}) > 2K$, then for each measurement vector $\boldsymbol{y} \in \mathbf{R}^M$ there exists at most one signal $\boldsymbol{x} \in \Sigma_K$ such that $\boldsymbol{y} = \boldsymbol{Ax}$ [14]. The singleton bound yields that the highest spark of an matrix $A \in \mathbf{R}^{M \times N}$ with $M < N$ is less than or equal to $M + 1$ and using the before stated theorems we get the requirement $M \geq 2K$.

While *spark* is useful notion for proving the exact recovery of a sparse optimization problem, it is NP-hard to compute since one must verify that all sets of columns of a certain size are linearly independent. Thus, it is preferable to use properties of $\boldsymbol{A}$ which are easily computable to provide recovery guarantees.

The coherence $\mu(\boldsymbol{A})$ of a matrix is the largest absolute inner product between any two columns of $\boldsymbol{A}$:

$$\mu(\boldsymbol{A}) = \max_{1 \leq i \neq j \leq N} \frac{\langle a_i, a_j \rangle}{\|a_i\|_2 \|a_j\|_2} \tag{14}$$

For any matrix $\boldsymbol{A}$,

$$spark(\boldsymbol{A}) \geq 1 + \frac{1}{\mu(\boldsymbol{A})} \tag{15}$$

Quite simple way to read the coherence is from the absolute value Gram matrix [17]. Gram matrix is defined as $\boldsymbol{G} = \boldsymbol{A'A}$ where we are considering conjugate transpose of the matrix $\boldsymbol{A}$. To read the coherence from Gram matrix, we reject the diagonal elements since they correspond to the inner product of an atom with itself (for a properly normalized dictionary they should be 1 anyway). Since $\boldsymbol{G}$ is symmetric we need to look only upper triangular half of it to read off the coherence. The value of coherence $\mu(\boldsymbol{A})$ is equal to largest value in upper triangular part of matrix $\boldsymbol{A}$ with diagonal excluded. It can be shown that $\mu(\boldsymbol{A}) \in [\sqrt{\frac{N-M}{M(N-1)}}, 1]$. The lower bound is known as the Welch bound. Note that when $N >> M$, the lower bound is approximately $\mu(\boldsymbol{A}) \geq \frac{1}{\sqrt{M}}$ [15]. In our example with overcomplete dictionary with spikes and sines basis, the coherence exactly corresponds to the Welch bound. That confirms our statement that spikes and sines are mutually completely non-coherent. In [18], a method to design sensing matrices with minimum coherence to a given sparsifying orthogonal basis was proposed. They provided a mathematical proof of the optimality in terms of coherence minimization for the proposed sensing matrices.

The prior properties of the CS design matrix provide guarantees of uniqueness when the measurement vector $\boldsymbol{y}$ is obtained without error. There can be two sources of error in measurements: inaccuracies due to noise at sensing stage (in the form of additive noise $\boldsymbol{y} = \boldsymbol{Ax} + noise$) and inaccuracies due to mismatches between the design matrix used during recovery and that implemented during acquisition (in the form of multiplicative noise $\boldsymbol{A'} = \boldsymbol{A} + \boldsymbol{A_{noise}}$). Under these sources of error, it is no longer possible to guarantee uniqueness, but it is desirable for the measurement process to be tolerant to both types of error. To be more formal, we would like the distance between the measurement vectors for two sparse signals $\boldsymbol{y} = \boldsymbol{Ax}$ and $\boldsymbol{y'} = \boldsymbol{Ax'}$ to be proportional to the distance between the original signal vectors $\boldsymbol{x}$ and $\boldsymbol{x'}$. Such a property allows us to guarantee that for small enough noise, two sparse vectors that are far apart from each other cannot lead to the same noisy measurement vector. This behavior has been formalized into the restricted isometry property (*RIP*) [19]–[22]. A matrix $\boldsymbol{A}$ has the (K, δ)-restricted isometry property((K, δ)-RIP if, for all $\boldsymbol{x} \in \Sigma_K$. In words, the (K, δ)-RIP ensures that all sub-matrices of $\boldsymbol{A}$ of size $M \times K$ are close to an isometry, and therefore are distance-preserving. This property suffices to prove that the recovery is stable to presence of additive noise and the RIP also leads to stability with respect to the multiplicative noise introduced by the CS matrix mismatches $\boldsymbol{A_{noise}}$.

IV. Conclusion

The conclusion goes here.

References

[1] C. E. Shannon, "Communication in the presence of noise," *Proc. Institute of Radio Engineers*, vol. 37, no. 1, pp. 10–21, 1949.

[2] H. Nyquist, "Certain topics in telegraph transmission theory," *Transactions of the American Institute of Electrical Engineers*, vol. 47, no. 2, pp. 617–644, 1928.

[3] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[4] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[5] E. J. Candès, "Compressive sampling," *Proceedings of the International Congress of Mathematicians*, pp. 1433–1452, 2006.

[6] E. Candès and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Problems*, vol. 23, no. 3, pp. 969–985, 2007.

[7] R. Baraniuk, "Compressive Sensing [Lecture Notes]," *IEEE Signal Processing Magazine*, vol. 24, no. July, pp. 118–121, 2007.

[8] E. Candes and M. Wakin, "An Introduction To Compressive Sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.

[9] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, "Compressed sensing MRI: A look at how CS can improve on current imaging techniques," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 72–82, 2008.

[10] M. F. Duarte, M. A. Davenport, D. Takbar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE signal processing magazine*, vol. 25, no. 2, pp. 83–91, 2008.

[11] J. A. Tropp, J. N. Laska, M. F. Duarte, J. K. Romberg, and R. G. Baraniuk, "Beyond Nyquist: Efficient sampling of sparse bandlimited signals," *IEEE Transactions on Information Theory*, vol. 56, no. 1, pp. 520–544, 2010.

[12] M. F. Duarte and Y. C. Eldar, "Structured Compressed Sensing: From Theory to Applications," vol. 59, no. 170, pp. 4053–4085, jun 2011.

[13] J. Romberg, "Tsinghua Course on Sparse Approximation - lecture notes," 2013, Accessed on 11.10.2017. [Online]. Available: http://jrom.ece.gatech.edu/tsinghua-oct13/

[14] I. Rish and G. Grabarnik, *Sparse Modeling: Theory, Algorithms, and Applications*, 2015.

[15] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via l1 minimization," *Proceedings of the National Academy of Sciences*, vol. 100, no. 5, pp. 2197–2202, 2003.

[16] J. B. Kruskal, "Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics," *Linear algebra and its applications*, vol. 18, no. 2, pp. 95–138, 1977.

[17] M. Elad, "Sparse and redundant representations: From theory to applications in signal and image processing," *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, pp. 1–376, 2010.

[18] M. P. Pereira, L. Lovisolo, E. A. B. Da Silva, and M. L. R. De Campos, "On the design of maximally incoherent sensing matrices for compressed sensing using orthogonal bases and its extension for biorthogonal bases case," *Digital Signal Processing: A Review Journal*, vol. 27, no. 1, pp. 12–22, 2014.

[19] J. Moreira, "What is... a rip matrix?" pp. 1–4, 2014.

[20] J. Romberg, "Mathematics of Compressive Sensing : Random matrices are restricted isometries," 2013.

[21] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, vol. 346, no. 9, pp. 589 – 592, 2008.

[22] J. D. Blanchard, C. Cartis, and J. Tanner, "Compressed sensing: How sharp is the restricted isometry property?" *SIAM Review*, vol. 53, no. 1, pp. 105–125, 2011.