A Short Introduction to Compressed Sensing

M.D.Baltzer

March 24, 2015

1 Introduction

Compressed sensing (CS) or compressive sampling is a new framework of signal acquisition with significant advances in the theory over the last 10 years. As a result of the drive for higher resolution and data capture rates, the amount of data generated world-wide exceeded the available storage capacity in ± 2007 [1] and continues to rise. Computer processing rates or clock times are increasing at a rate of Moores Law (approximately doubling every one and a half years), however our ability to convert analogue signals to digital in terms of the sampling rate is only doubling in the region of every 6 to 8 years [1]. It therefore makes imminent sense to reduce the amount of data acquired, especially if a significant portion of it is discarded in subsequent processing (e.g. JPEG compression) or the downstream transmission is unable to cope with the data rates. One way to reduce the quantity of data is through a dimensionality reduction especially of signals of known sparsity or compressibility. By sparse we mean a signal representation of length n with k non zero coefficients. A compressible signal on the other hand is one which can be well represented by a signal with only k of the n coefficients [2]. A classic example of such a signal is one of an image in which the pixel energy values are transformed either by a Fourier transform or discrete cosine transform (DCT). These coefficients if arranged in order of magnitude, decay at a power law rate and by retaining the k largest coefficients and discarding the rest, the siganl is well represented. The existing paradigm to signal acquisition is that one has to sample at twice the Nyquist rate (twice the highest frequency of a bandlimited signal) to recover the signal perfectly. Intuitively one would think that it must be possible with measurements proportional to the size of the compressed signal. In fact it can be shown that with as few as $O(klog(n/k))$ measurements of a k-sparse signal, the signal can be recovered exactly [2]. In principle the problem is one of a dimensionality reduction and recovery of the original signal which can be expressed mathematically by:

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

where $\mathbf{y}$ is the measured signal and $\mathbf{x}$ the signal to be recovered. If the measurement matrix A was full rank there would be no problem as there would be a unique solution to $\mathbf{x}$, however with only k measurements of a signal of length n the system is under-determined. To obtain a solution we need to turn to some sort of recovery algorithm with appropriate constraints. In this text we aim to discuss the properties of a desired sensing matrix A and some of the algorithms available to reconstruct the signal.

2 Background Information

In order to anaylse compressed sensing a number of mathematical concepts for vectors and matrices are required. In the rest of the text, vectors will be denoted by bold font $\mathbf{x}^n$ where n is the dimension of the vector, vector components will be denoted by x_i where i is the vector component and vector norms will be denoted by $\|x\|_p$ where p is the l_p norm given by:

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p} \tag{2}$$

The following concepts are used to describe the desired properties of the sensing matrix A:

- A basis for $\mathbb{R}^n$ is denoted by $\phi_{i=1}^n$ if the vectors ϕ_i span $\mathbb{R}^n$ and are linearly independent.

- The null space of a matrix A is defined by $\mathcal{N}(A) = \{\mathbf{x} | A\mathbf{x} = \mathbf{0} \ \forall \ \mathbf{x} \in \mathbb{R}^n\}$.

- The spark of a matrix is the minimum number of colums in A such that there are n linearly dependent colums. For example if there are 4 linearly dependent colums and there are no smaller sets of columns that are linearly dependent then $spark(A) = 4$.

- The l_0 norm of $\mathbf{x}$ written $\|\mathbf{x}\|_0$ is defined as $\|\mathbf{x}\|_0 = |\text{supp}(\mathbf{x})|$ where $\text{supp}(\mathbf{x}) = \{i \,|\, x_i \neq 0\}$

- A k sparse signal is one for which $\Sigma_k = \{\mathbf{x} \mid \|\mathbf{x}\|_0 \leq k\}$ where k is the number of non zero coefficients of the polynomial representing the signal. Normally k is very much smaller than n, the ambient dimension of the signal.

- The null space property (NSP) of order k of a matrix A is one for which there exists a constant C such that

$$\|h_\Lambda\|_2 \leq C \frac{\|h_{\Lambda^c}\|_1}{\sqrt{k}} \tag{3}$$

where $h \in \mathcal{N}(A)$. Λ is the set of indices of the n non zero coefficients such that $|\Lambda| \leq k$ and Λ^c are those not in Λ.

- A measure of the compressiblility of a signal can be given by the error in quantifing a signal $\mathbf{x}$ by a signal $\hat{\mathbf{x}}$ usch that:

$$\sigma_k((x)_p = \min_{\hat{\mathbf{x}} \in \Sigma_k} \|\mathbf{x} - \hat{\mathbf{x}}\|_p \tag{4}$$

- The restricted isometry property (RIP). A matrix A satisfies the RIP of order k if there exists a $\delta_k \in (0, 1)$ such that

$$(1 - \delta_k)\|\mathbf{x}\|_2^2 \leq \|A\mathbf{x}\|_2^2 \leq (1 + \delta_k)\|\mathbf{x}\|_2^2 \tag{5}$$

- The coherence of a matrix $\mu(A)$, is the largest absolute inner product between any two colums of A :

$$\mu(A) = \max_{1 \leq i < j \leq n} \frac{\langle a_i, a_j \rangle}{\|a_i\|_2 \|a_j\|_2} \tag{6}$$

3   Signal Recovery

In our discussion of a sensing matrix we will assume that the signal is already discretized by some means or another and that we are dealing with a sparse representation of the signal at hand. The construction of the sensing matrix will be disucssed in more detail later in the text. For now we have been given a suitable sensing matrix with which to work with. This matrix is composed of the coefficients of stacked functions so that multiplication of a signal vector $\mathbf{x}$ with A is an inner product of the test functions with the vector of interest. In principle we are comparing the signal against a test test signal so that the measurement $\mathbf{y}$ can be written as

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

In order to reconstruct the signal vector $\mathbf{x}$ from the measurement we can find the closest match to it by finding the vector $\hat{\mathbf{x}}$ that matches the measurement $\mathbf{y}$. Naively we could choose to minimize the l_2 norm as in solving a least squares problem. This can be written mathematically as

$$\hat{\mathbf{x}} \quad \text{such that} \quad \|\mathbf{y} - A\hat{\mathbf{x}}\|_2^2 \quad \text{is minimized} \tag{7}$$

The solution to this equation is well known and is given by $\hat{\mathbf{x}} = A^*(AA^*)^{-1}\mathbf{y}$ where $A^*(AA^*)^{-1}$ is the pseudoinverse of A denoted $A^\dagger$ [3]. This is fine if we wanted to minimize the energy of the signal but this may not be optimal for all types of signals, a classical case being in MRI where most coefficients of the transform are the same size. Then, by taking the k largest coefficients and setting the rest to zero and finding the least squares solution the
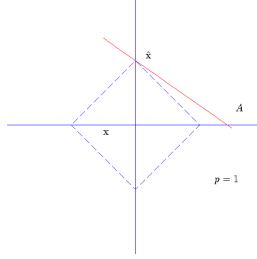
Figure 1: Approximation of a point $\mathbf{x}$ by a one-dimensional subspace using the l_1 norm

algorithm fails to find a satisfactory $\mathbf{x}$ [4]. Instead if we tried to solve the optimization problem:

$$\hat{\mathbf{x}} = \arg_{\mathbf{z}} \min \|\mathbf{z}\|_0 \quad \text{subject to} \quad z \in \mathcal{B}(\mathbf{y}) \tag{8}$$

where $\mathcal{B}(\mathbf{y}) = \{z \mid A\mathbf{z} = \mathbf{y}\}$ and the notation $\|\mathbf{z}\|_0$ refers to the zero norm which is defined by $\|\mathbf{x}\|_0 = |\text{supp}(x)|$, where $\text{supp}(x) = \{i \mid x_i \neq o\}$, we should find the sparsest $\mathbf{x}$. The problem with this approach is that it is NP-hard i.e. it cannot be solved in polynomial time. This is because to find the solution we need to check all $_k^n C$ combinations of k columns of A. The next natural step may be to minimize the function with the l_1 norm. We can write this as:

$$\hat{\mathbf{x}} = \arg_{\mathbf{z}} \min \|\mathbf{z}\|_1 \quad \text{subject to} \quad \mathbf{z} \in \mathcal{B}(\mathbf{y}) \tag{9}$$

If $\mathcal{B}(\mathbf{y}) = \{\mathbf{z} \mid A\mathbf{z} = \mathbf{y}\}$ then the problem can be posed as a linear problem [2]. This is because the ojective function $\|\mathbf{z}\|_1$ is convex. The solution to the linear program is known to promote sparsity, so the solution should be similar to that obtained through (8) i.e. a sparse vector. We see this in $\mathbb{R}^2$ when we approximate the closest point in A ($\hat{\mathbf{x}} \in A$) to $\mathbf{x}$ using the l_1 and l_2 norms (figures:1 and 2) that we obtain a sparse vecor with the l_1 norm in contrast to the l_2 norm which spreads out the error more evenly among the coefficients. This observation generalizes to higher dimensions [2].

4   The sensing Matrix

Earlier we mentioned that a sparse signal could be recoverd exactly with as little as $O(k \log(n/k))$ measurements. It is easy to show that with 2k measuremens the signal can be recovered uniquely from $A\mathbf{x} \in \mathbb{C}^{2k}$.

Lemma 4.1. *[4] suppose that the $m \times n$ measurement matrix A is such that every set of $2k$ columns of A are linearly independent. Then a k-sparse vector $\mathbf{x} \in \mathbb{C}^n$ can be reconstructed uniquely from $A\mathbf{x} \in \mathbb{C}^{2k}$.*
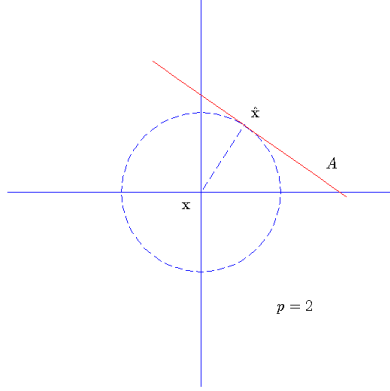
4

Figure 2: Approximation of a point $\mathbf{x}$ by a one-dimensional subspace using the l_2 norm

The proof is made by contradiction as follows:

Proof. [4] Suppose the unique construction failed, then there exists two k-sparse vectors $\mathbf{x}$, $\mathbf{x}' \in \mathbb{C}^n$ such that $A\mathbf{x} = A\mathbf{x}'$ which implies $A(\mathbf{x} - \mathbf{x}') = 0$. Since $(\mathbf{x} - \mathbf{x}')$ is $2k$ sparse there must be $2k$ columns that are linearly dependent, which is a contradiction. $\square$

To obtainer a tighter bound on the number of measurements we need to define some of the desired properties of our measurement (sensing) matrix. It is immediately evident that if we wish to recover all sparse signals then for any two vectors $\mathbf{x}$,$\mathbf{x}' \in \Sigma_k$ we cannot have $A\mathbf{x} = A\mathbf{x}'$. This means that the $\mathcal{N}(A)$ cannot contain any vectors in Σ_{2k}, refer section 2. Furthermore we would like our matrix to satisfy the ristricted isometry property 2. According to a theorem from Candés and Tao (2005) [4], if the matrix obeys the RIP property then, if every collection of $4k$ columns are almost orthogonal in that the $4k$ singular values range between 0.9 and 1.1, then any given k-sparse signal $\mathbf{x}$ can be recovered from $A\mathbf{x}$ by basis pursuit (9). With this in mind we set out to design our sensing matrix. One way would be to deterministically construct a matrix that obeys the RIP property of order k., however these tend to lead to an unacceptably high requirement on m [2]. As it happens a matrix constructed with independent and identically distributed (i.i.d.) entries chosen according to a Gaussian or any subgaussian disribution has the desired properties with high probability [2]. The number of measurements required to meet these criteria is $m = \mathcal{O}(klog(n/k)/\delta_{2k}^2)$. The choice of random matrix has a numberof benefits over a deterministic construction because in practice we are interested in signals that are sparse with respect to some basis Φ. In this case we require the matrix $A\Phi$ to meet

the RIP requirement which it does if A is random. On the other hand if if A is constructed deterministically, then Π would have to be taken into account in the construction to ensure that the resultant product meets the RIP requirements.

5   Algorithms for recovery

If the problem is posed as a covex optimization problem then there are accurate and efficient numerical solvers available. In section 3 it was mentioned that equation eqreflineareq could be posed and solved as a linear program. In cases where the problem is posed differently e.g. $\mathcal{B}(\mathbf{y}) = \{\mathbf{z} \mid \|A\mathbf{z} - \mathbf{y}\|_2 \leq \epsilon\}$ then the minimization problem becomes a convex problem with a conic constraint [2].

References

[1] Baraniuk, R.,2012. Compressed Sensing. ECE lecture series. Retrieved on 20/3/2015 from https://www.youtube.com/watch?v=RvMgVv-xZhQ.

[2] Kutyniok, G., Eldar, Y.C., Duarte, M.F., Davenport, M.A.,1993. Introduction to compressed sensing. Preprint 93. Retrieved 24/2/2015 from https://www.dfg-spp1324.de/download/preprints/preprint093.pdf

[3] Trefethen, L.N., 1997. Numerical linear algebra. Society for Industrial and Applied Mathematics, Philadelphia.

[4] Tao, Terence, 2008. Compressed sensing. NTNU's Onsager Lecture, retrieved 19/03/2015 from https://www.youtube.com/watch?v=IFgeXzturqk&list=PLC94A02A1218B24DF&index=2