Chapter 2

Compressed Sensing

Compressed sensing is a novel research area introduced by Donoho [18] and by Candes, Romberg, and Tao [9] in 2006, and since then has already become a key concept in various areas of applied mathematics, computer science, and electrical engineering. It surprisingly predicts that high-dimensional signals, which allow a sparse representation by a suitable basis or, more generally, a frame, can be recovered from what was previously considered highly incomplete linear measurements by using efficient algorithms. This chapter shows an introduction to and a survey of compressed sensing.

2.1   Introduction

In the real world, people often observe and save signals in digital devices for further analysis, e.g., medical image imaging, CT tomography. Using the discrete sampling and storage, the system are prospected to exactly reconstruct the original signals. If we assume that original signal is $1 \times N$ vector, the observation behaviour can be modelled as a sensing matrix A with a size of $m \times N$. Then the sampled data y can be described as

$$y = Ax, \tag{2.1}$$

Then an important question emerges:

- In order to reconstruct signal x, *at least* how many samples y are needed?

Traditional linear algebra theory states that unique solution of (2.1) exists if $m = N$. That is to say, we need at least N samples y. However, consider a case when the signal x is sparse, which indicates most of values in x are zero or relatively small, is there any method to reconstruct x using smaller numbers of observations y? ($m < N$)

Conventionally, if $m < N$, the the equation (2.1) is underdetermined, so various solutions exist which means it cannot uniquely recover the original signal x. However, if the x is sparse, then the answer maybe *Yes*: Using the theory of the compressed sensing (CS), surprisingly it allows a possible sparse information's reconstruction from a very few numbers of samples y ($m << N$), where the sparse information can naturally exist (such as x) or generated by orthogonal basis decomposition (such as $x = As$ where s is sparse, A is the basis e.g. wavelets decomposition). Consequently, the CS provides an possible method to reduce the number of observations theoretically. In conclusion, the task of CS can be abstracted as:

- Find the minimum number of observations for the sensing matrix.

- Design reconstruction algorithm to uniquely recover original signals via those observations.

2.2 Sparsity: A Reasonable Assumption

2.2.1 Definition of sparse signals

To state the problem mathematically precisely, let now $x \in R^{N \times 1}$ be our signal of interest. As prior information, we either assume that x itself is sparse, i.e., it has very few K non-zero coefficients in the sense that

$$\|x\|_0 := \{i : x_i \neq 0\} \tag{2.2}$$

is small, or that there exists an orthonormal basis or a frame Ψ such that $x = \Psi s$, where x is a linear combination of few k basis chosen from Ψ, and s is the corresponding coefficients of representing x in the domain constructed by the basis Ψ, k is termed as the sparsity.

For real world signals, a variety of representation systems can be used to provide sparse approximations. If this is not the case, more data sensitive methods such as dictionary learning algorithms, in which a suitable representation system is computed for a given set of test signals, are available This process is customarily termed dictionary learning. The most well-known adaptive learning algorithm is the K-SVD algorithm introduced by Aahron, Elad, and Bruckstein in [19]. In this chapter, we will assume throughout that x is already a sparse vector or can be sparse represented.

2.3 Model of Sparse Reconstruction

When the x is sparse, an intuitive solution of the underdetermined equation (2.1), equals the solution of the problem (P0), which aims to find the minimum numbers of supports in x as following:

$$P0: \quad min\|x\|_0 \quad s.t. \quad y = Ax \tag{2.3}$$

Then the following Theorem presents the sufficient condition to gain the unique solution:

Theorem 1. *Assume that $A \in R^{m \times N}$ is an average of the independent $2k$ column matrix. Then for any k-sparse vector x, the x can be exactly recovered by solving the problem P0 (2.3).*

Hence, according to Theorem 1, then there is a coding and decoding of (A, Δ) such that $x = \Delta(Ax)$, which means $2k$ times the number of observations is sufficient for reconstruction. However, solving the problem P0 is very difficult. In fact, P0 is an NP-complete problem [20].

So could we find a more efficient decoding algorithms? The answer maybe *Yes* if some condition is satisfied. In this case, using the solution of the following problem P1, which is

termed as Basis Pursuit (BP) or l1-norm minimisation, we can uniquely reconstruct x.

$$P1: \quad min\|x\|_1 \quad s.t. \quad y = Ax \tag{2.4}$$

For solving the problem P1, various types of recovery algorithms exist, which can be sorted as convex optimisation, greedy, and combinatorial algorithms (section 2.5), and each one having its own advantages and disadvantages. Here in this section, we provide the widely used linear programming (LP) as an example to solve P1:

$$LP: \quad min\|t_1 + t_2 + \cdots + t_N\|_1 \quad t \in R^N \quad s.t. \quad y = Ax,$$
$$-t_j \le x \le t_j, \quad t_j \ge 0, \quad j = 1 \cdots N \tag{2.5}$$

2.3.1 Null Space Property

According to (2.4), the sparse reconstruction is achievable when P1 equals P0. So lets return to the question: when P1 equals P0? In order to develop this question, the null space of sensing matrix plays a important role, which is defined as $N(A) = \{x : Ax = 0\}$. Then the null space property and its corresponding theorem are given to answer that question, and presents us the sufficient and necessary condition for P1 = P0:

Definition 1. *The $m \times N$ sensing matrix A satisfies the null space property of order k provided for every index set S with supp(S) = k, we have*

$$\|h_S\|_1 < \|h_{S^c}\|_1, \quad \forall h \in N(A), \quad h \ne \Theta \tag{2.6}$$

Theorem 2. *Every k-sparse vector $\hat{x}$ is the unique solution of the P1 problem (2.4) if and only if A satisfies the null space property of order k [21].*

According to the Theorem 2, the equivalent condition for the existence of a unique sparse solution of (P1) can now be stated in terms of the null space property (NSP). Then many convex optimisation algorithm and greedy method can be applied to solve P1 problem (section 2.5.1).

2.3.2 Restricted Isometry Property

Although the NSP and the Theorem 2 presents us an equivalent condition to get unique reconstruction through solving P1, however, how to verify the sensing matrix satisfying NSP in (2.6) is computational complex, so NSP is not practically useful.

Then a more intuitive concept termed as Uniform Uncertainty Principle or Restricted Isometry Property (RIP) is introduced in [9]. It measures the degree to which each subset of k column vectors of the sensing matrix A is close to being an isometry.

Definition 2. *Let A be an $m \times n$ sensing matrix. Then A has the Restricted Isometry Property (RIP) of order k, if there exists a $\delta_k \in (0, 1)$ such that*

$$(1 - \delta_k)\|x\|_2 \leq \|Ax\|_2 \leq (1 + \delta_k)\|x\|_2 \quad for\ all\ k\ sparse\ x \tag{2.7}$$

Theorem 3. *A sufficient condition for the null space property (NSP) to hold and for all k-sparse signal is that RIP holds for 2k-sparse signals with*

$$\rho = \frac{\sqrt{2}\delta_{2k}}{1 - \delta_{2k}} \leq 1 \tag{2.8}$$

It follows that if $\delta_{2k} \leq \sqrt{2} - 1$ then NSP is satisfied, so that the unique reconstruction can be achieved by l1-norm minimisation (2.4).

According to the Theorem 3, an equivalent condition for the existence of a unique sparse solution of (P1) can now be stated in terms of RIP. If the sensing matrix satisfies the RIP, then for any k-sparse vector x, the Euclidian distance of projected x (i.e. Ax) and x still keep very close if the $\delta_{2k} \in (0, 1)$ is relatively small.

2.3.3 Robust Recovery

Back to our original task in compressed sensing framework in section 2.1, which aims at designing a efficient coding and decoding procedure, we now can use RIP-satisfied matrix for encoding, and l1-norm minimisation as decoding to recover k-sparse signals. However,

in the real world, signals sometimes are not strictly sparse signals, but usually approximate sparse. For such signals, is the framework of coding (RIP matrix projection) and decoding (l1-norm minimisation) still suitable? Interestingly, the answer is yes, it still can be better done: First, we define the concept of the best k-term approximation error:

$$\sigma_k(x)_p := min\|x - \hat{x}\|_p \tag{2.9}$$

Then, the error boundary can be presented in terms of the best k-term approximation as follows:

Theorem 4. *([9]) Let A be an $m \times N$ matrix which satisfies the RIP of order 2k with $\delta_{2k} \leq \sqrt{2} - 1$. Let $x \in RN$, and let $\hat{x}$ be a solution of the associated (P1) problem. Then*

$$\|x - \hat{x}\|_2 \leq C \cdot (\frac{\sigma_k(x)_1}{\sqrt{k}}) \tag{2.10}$$

2.3.4 Minimum Samples

Lets return to the question in section 2.1: In order to reconstruct signal x, *at least* how many samples y are needed? The [22] shows the following Theorem which presents a lower boundary:

Theorem 5. *Assume a matrix $\in R^{m \times N}$ for sensing and l1-norm minimisation for reconstruction, then for any 2k-sparse vector $x \in R^N$, then minimum required number of observations is:*

$$n \geq c_1 k log(\frac{N}{c_2 k}), \quad c_1 = 1/log9 \ and \ c_2 = 4 \tag{2.11}$$

Following this Theorem, any underdetermined sensing matrix requires at least $O(klog(\frac{N}{k}))$ samples for sparse reconstruction in (2.4). Then a new question emerges: Since the Theorem 3 presents a unique sparse solution of (P1) stated in terms of RIP, can this sensing and recovery approach (RIP-matrix sensing and l1-norm minimisation) reach the lower boundary of the minimum required samples that is $O(klog(\frac{N}{k}))$? The following section answers this question with respect to various of sensing matrix.

2.4   Sensing Matrix

Verification for RIP is not practical, for the RIP serves for every possible sparse vector, and [23] demonstrates that testing whether a matrix satisfies RIP is NP-hard. However, fortunately, we can often construct the matrix satisfying RIP by using randomness, which gives a practical way for creating the sensing matrix.

2.4.1   Random Matrix

We consider two types of random matrices: Gaussian random matrix and Bernoulli random matrix. The Gaussian random matrices refers to the matrix whose elements $a_{i,j}$ are independent random variables and subject to the distribution $N(0, 1)$. The Bernoulli matrix refers to the matrix whose elements $a_{i,j}$ has the same probability to be 1 or -1.

2.4.1.1   Probability Reconstruction

Theorem 6. *Assumed $m \times N$ matrix A is a Gaussian or Bernoulli random matrix. Then, if $k \leq C_1/log(N/k)$, matrix A satisfy the k-order RIP (2) with the probability bigger than*

$$1 - exp(-C_2 n) \tag{2.12}$$

, where constants C_1, C_2 depends only on the RIP constant δ.

The Theorem shows that the Gaussian or Bernoulli random matrices satisfies the $k = O(n/log(N/k))$ order RIP, indicating that these random matrices reach the lower boundary of the minimum required samples that is $O(klog(\frac{N}{k}))$ in (5). Therefore, these random matrices presents an efficient and optimal coding and decoding method, which involve random sensing matrix projection and l1-norm minimisation respectively, and apply the minimum amount of observations.

2.4.2 Deterministic Matrix

Although a random matrix (Gaussian or Bernoulli) has high probability to satisfies RIP and provide optimal coding and decoding method in terms of number of samples, in practice, it is more desirable to construct a deterministic matrix because the certainty in matrix is more attractive. For example, from the perspective of structural decoding algorithms, deterministic matrix helps to reduce memory cost and improve reconstruction speed. The current construction methods for deterministic matrix are mainly based on the property of matrix coherence:

Definition 3. *Assume a $m \times N$ sensing matrix A with normalised columns $A_1, \cdot, A_N$. The coherence of matrix A is defined as*

$$\mu(A) = max_{i \neq j} | < A_i, A_j > | \tag{2.13}$$

Then the lower bound of $\mu(A)$, termed as Welch bound, is given in [24] as:

$$\mu(A) \geq \sqrt{\frac{N - n}{(n - 1)N}} \tag{2.14}$$

Theorem 7. *Assume matrix A has the coherence $\mu(A)$, then A satisfies $\delta_k = (k - 1)\mu(A)$ order RIP [25].*

Combining the Theorem 7 with the Welch bound in definition 3, the sensing matrix A must satisfy the $k = \sqrt{2n}$ RIP. Hence, if we only investigate matrix coherence, we hardly get the lower boundary of the minimum required samples $O(klog(\frac{N}{k}))$ in (5). Thus, how to construct the deterministic RIP matrix to match this lower bound of required samples is still an open issue [26].

2.4.3 Partial Random Structural Matrix

Since there exists problem in constructing a deterministic matrix, and since there is always not enough freedom for building random Gaussian matrix subject to physical or other constraints in applications, at the moment, a new type of matrix structure is investigate as it

has a combination features of partial randomness and partial deterministic design, which is termed as the partial random structural matrix. It becomes popular since most applications have constraints in design sensing matrix, although many those matrix cannot reach the optimal boundary of minimum required samples $O(klog(\frac{N}{k}))$ in (5).

2.4.3.1   Random Partial Fourier Matrix

When we randomly choose m rows in the Fourier matrix Ψ to generate a $m \times N$ matrix, the matrix becomes the random partial Fourier matrix Ψ_m. Part of random Fourier matrix has a strong application background. For example, the randomly observed partial frequency information can be fully recovered by l1-norm minimisation [27]. And according to the CS theory which has a $m \times N$ sensing matrix where $m << N$, the sampled points m is far less than original required number N under Nyquist rate. In addition, we just apply the random sample behaviour so that the low-rate CS sampling is achieved. In [28], the number of minimum samples is proven as $n = O(k(logN)^4)$.

2.4.3.2   Rademacher Matrix

The Rademacher random variable takes only values -1 and +1 with the same likelihood. In hardware design, it can be easily implemented, so the [15] use this sequence in a pre-mixer for random projection, and successfully designs a low rate analog to digital converter based on the CS framework, and terms this architecture as the random demodulator (details in the next chapter). From the theoretical aspect, the required minimum samples $m = O(klog^6N)$. In practise, however, simulations show that the random demodulator requires just $O(klog(N/k))$ samples to stably reconstruct the signal. This result reaches the optimal boundary of minimum required samples $O(klog(\frac{N}{k}))$ in (5).

2.4.3.3 Random Circulant Matrix

In many cases where signal convolution is involved, the circulant matrix always exists whose each row is shifted compared to its neighbours. This is because that any n-point circular convolution of two sequences $x \in R^N$ and $h \in R^N$ is $h \otimes x = Hx$, where H is the corresponding $N \times N$ circulant matrix to original impulse response h. In CS framework, the randomness is embedded into each elements of the matrix, e.g. Rademacher random variable is often implemented for constructing each row of circulant matrix, which forms a partial $m \times N$ random circulant matrix. Then paper [29] proves that a partial random circulant matrix satisfies k order RIP when the number m of samples $m \simeq (slogN)^{3/2}$, where N is the length of the pulse. Also, the required number of samples do not reach the optimal bound in (5). In practise, these matrices are so important in signal processing applications and wireless communications where convolutions widely occur.

2.4.3.4 Specific Signal Processing Matrix

Many tasks signal processing do not really require full signal information but rather to solve maybe inference or estimation problems. In these case, fully sensing may not be sufficient so that specific sensing matrix need to designed. For instance, in [30], a compressive filtering matrix is introduced, which aims at sensing and directly extracting useful information without recovery. This area relates to the Compressive Signal Processing (CSP, details in chapter 5), and how to measure the minimum required number of samples is still an open issue.

2.5 Reconstruction Algorithms

In this section, we will provide a brief overview of the different types of algorithms typically used for sparse recovery. Convex optimisation algorithms require very few measurements but are computationally more complex. On the other extreme are combinatorial

algorithms, which are very fast, and often sub-linear, but require many measurements that are sometimes difficult to obtain. Greedy algorithms are in some sense a good compromise between those extremes concerning computational complexity and the required number of measurements.

2.5.1 Convex Optimisation

In section 2.3.2 we already stated the convex optimisation problem (P1) as (2.4). If the measurements are affected by noise, a conic constraint is required and the minimisation problem needs to be changed to basis pursuit de-noising (BPDN) [31] problem, which allows some measurement mismatch of $\epsilon > 0$:

$$\min \|x\|_1 \quad s.t. \quad \|y - Ax\|_2 \le \epsilon \tag{2.15}$$

for a carefully chosen $\epsilon > 0$. For a particular parameter $\lambda > 0$, the problem is equivalent to the unconstrained version given by

$$\min \|Ax - y\|_2 + \lambda \|x\|_1 \tag{2.16}$$

Developed convex optimisation algorithms specifically adapted to the compressed sensing include interior-point methods [9], projected gradient methods [32], and iterative thresholding [33].

2.5.2 Combinatorial Algorithms

These methods apply group testing to highly structured samples of the original signal. From the various types of algorithms, the HHS pursuit [34] and haining pursuit [35] are popular algorithms, which require many measurements but achieve extremely rapid computational speed for sparse signal reconstruction. They are sensitive to noise as they strongly rely on the group testing of highly structured samples.

2.5.3 Greedy Algorithms

Greedy algorithms iteratively approximate the coefficients and the support of the original signal. They have the advantage of being very fast and easy to implement. Often the theoretical performance guarantees are very similar to, for instance, l1 minimisation results. Greedy algorithms are usually looking for the following approximate solution of the problem (P0):

$$min \quad \|x\|_0 \quad s.t. \quad \|Ax - y\|_2 \leq \epsilon. \tag{2.17}$$

The basic idea is to choose the least column in A in order to form an approximate representation of y. The most commonly used as a greedy algorithm for OMP algorithm [36], which is a widely used for sparse reconstruction and develops the process of matching pursuit. As shown in Algorithm 1, it progressively manages to find the support of the unknown sparse signal: Give an acquisition system $y = Ax$ where x is K sparse vector and a $m \times N$ sensing matrix A, each iteration of OMP selects one support λ of the vector x which contributes the most to the observation y. This selection method is based on testing the correlation values between the current columns of A and residue r. The OMP iteration would not stop before the residue r becomes relatively small. The OMP algorithm also has many developed versions, such as StOMP [18], regularised OMP (ROMP) [37], and compressive sampling MP (CoSaMp) [38].

Algorithm 1: Orthogonal Matching Pursuit (OMP)

---

input : measurement $A \in R^{m \times N}$, observation y.

output: recovery $\hat{s}$.

1   $r_0 \leftarrow y; \Lambda_0 \leftarrow \Theta; i \leftarrow 0; ;$

2   **while** $r_i \geq theshold$ **do**

3      $h_i \leftarrow A^T r_i \quad //match;$

4      $\Lambda_{i+1} \leftarrow \Lambda_i \cup \{\arg\max_{1 \leq j \leq N} |h_i(j)|\} \quad //identity;$

5      $\hat{s}_{i+1} \leftarrow \arg\min_s \|y - A_{\Lambda_{i+1}} \hat{s}_i\|_2 \quad //determine;$

6      $r_{i+1} \leftarrow y - A\hat{s}_{i+1} \quad //update;$

7   **end**

Bibliography

[1] J. M. Moura, "What is signal processing?[president's message]," *Signal Processing Magazine, IEEE*, vol. 26, no. 6, pp. 6–6, 2009.

[2] A. V. Oppenheim, R. W. Schafer, J. R. Buck *et al.*, *Discrete-time signal processing*. Prentice-hall Englewood Cliffs, 1989, vol. 2.

[3] J. D. Broesch, *Digital Signal Processing: Instant Access: Instant Access*. Newnes, 2008.

[4] H. Mamaghanian, N. Khaled, D. Atienza, and P. Vandergheynst, "Compressed sensing for real-time energy-efficient ecg compression on wireless body sensor nodes," *Biomedical Engineering, IEEE Transactions on*, vol. 58, no. 9, pp. 2456–2466, 2011.

[5] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, "Compressed sensing mri," *Signal Processing Magazine, IEEE*, vol. 25, no. 2, pp. 72–82, 2008.

[6] K. Danckaert, K. Masselos, F. Cathoor, H. J. De Man, and C. Goutis, "Strategy for power-efficient design of parallel systems," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, vol. 7, no. 2, pp. 258–265, 1999.

[7] E. J. Cands and M. B. Wakin, "An introduction to compressive sampling," *Signal Processing Magazine, IEEE*, vol. 25, no. 2, pp. 21–30, 2008, 1stV.

[8] C. K. Chui, *An introduction to wavelets*. Academic press, 1992, vol. 1.

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

[10] R. G. Baraniuk, "Compressive sensing," *IEEE signal processing magazine*, vol. 24, no. 4, 2007.

[11] W. Bajwa, J. Haupt, A. Sayeed, and R. Nowak, "Compressive wireless sensing," in *Proceedings of the 5th international conference on Information processing in sensor networks*. ACM, 2006, pp. 134–142.

[12] R. G. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE Signal Processing Magazine*, 2008.

[13] S. Foucart and H. Rauhut, *A mathematical introduction to compressive sensing*. Springer, 2013.

[14] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse mri: The application of compressed sensing for rapid mr imaging," *Magnetic resonance in medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.

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

[16] M. Mishali and Y. C. Eldar, "Expected rip: Conditioning of the modulated wideband converter," in *Information Theory Workshop, 2009. ITW 2009. IEEE*. IEEE, 2009, pp. 343–347.

[17] J. N. Laska, J. P. Slavinsky, and R. G. Baraniuk, "The polyphase random demodulator for wideband compressive sensing," in *Signals, Systems and Computers (ASILO-*

MAR), 2011 Conference Record of the Forty Fifth Asilomar Conference on. IEEE, 2011, pp. 515–519.

[18] D. L. Donoho, Y. Tsaig, I. Drori, and J.-L. Starck, "Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit," *Information Theory, IEEE Transactions on*, vol. 58, no. 2, pp. 1094–1121, 2006.

[19] M. Aharon, M. Elad, and A. Bruckstein, "¡ img src="/images/tex/484. gif" alt="\ rm k"¿-svd: An algorithm for designing overcomplete dictionaries for sparse representation," *Signal Processing, IEEE Transactions on*, vol. 54, no. 11, pp. 4311–4322, 2006.

[20] G. Davis, S. Mallat, and M. Avellaneda, "Adaptive greedy approximations," *Constructive approximation*, vol. 13, no. 1, pp. 57–98, 1997.

[21] R. A. DeVore, "Deterministic constructions of compressed sensing matrices," *Journal of Complexity*, vol. 23, no. 4, pp. 918–925, 2007.

[22] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *Information Theory, IEEE Transactions on*, vol. 49, no. 12, pp. 3320–3325, 2003.

[23] A. S. Bandeira, E. Dobriban, D. G. Mixon, and W. F. Sawin, "Certifying the restricted isometry property is hard," *arXiv preprint arXiv:1204.1580*, 2012.

[24] L. Welch, "Lower bounds on the maximum cross correlation of signals (corresp.)," *IEEE Transactions on Information theory*, pp. 397–399, 1974.

[25] A. Weil, "On some exponential sums," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 34, no. 5, p. 204, 1948.

[26] Z. Xu, "Deterministic sampling of sparse trigonometric polynomials," *Journal of Complexity*, vol. 27, no. 2, pp. 133–140, 2011.

[27] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *Information Theory, IEEE Transactions on*, vol. 52, no. 12, pp. 5406–5425, 2006.

[28] M. Rudelson and R. Vershynin, "On sparse reconstruction from fourier and gaussian measurements," *Communications on Pure and Applied Mathematics*, vol. 61, no. 8, pp. 1025–1045, 2008.

[29] H. Rauhut, J. Romberg, and J. A. Tropp, "Restricted isometries for partial random circulant matrices," *Applied and Computational Harmonic Analysis*, vol. 32, no. 2, pp. 242–254, 2012.

[30] M. A. Davenport, P. T. Boufounos, M. B. Wakin, and R. G. Baraniuk, "Signal processing with compressive measurements," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 4, no. 2, pp. 445–460, 2010.

[31] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM journal on scientific computing*, vol. 20, no. 1, pp. 33–61, 1998.

[32] M. A. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 1, no. 4, pp. 586–597, 2007.

[33] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Applied and Computational Harmonic Analysis*, vol. 27, no. 3, pp. 265–274, 2009.

[34] A. C. Gilbert, M. J. Strauss, J. A. Tropp, and R. Vershynin, "One sketch for all: fast algorithms for compressed sensing," in *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*.   ACM, 2007, pp. 237–246.

[35] ——, "Algorithmic linear dimension reduction in the l_1 norm for sparse vectors," *arXiv preprint cs/0608079*, 2006.

[36] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *Signal Processing, IEEE Transactions on*, vol. 41, no. 12, pp. 3397–3415, 1993.

[37] D. Needell and R. Vershynin, "Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit," *Foundations of computational mathematics*, vol. 9, no. 3, pp. 317–334, 2009.

[38] D. Needell and J. A. Tropp, "Cosamp: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, 2009.