

# A SIMPLE, EFFICIENT AND NEAR OPTIMAL ALGORITHM FOR COMPRESSED SENSING

*T. Blumensath and M. E. Davies*

IDCOM & Joint Research Institute for Signal and Image Processing  
The University of Edinburgh  
King's Buildings, Mayfield Road  
Edinburgh EH9 3JL, UK

## ABSTRACT

When sampling signals below the Nyquist rate, efficient and accurate reconstruction is nevertheless possible, whenever the sampling system is well behaved and the signal is well approximated by a sparse vector. This statement has been formalised in the recently developed theory of compressed sensing, which developed conditions on the sampling system and proved the performance of several efficient algorithms for signal reconstruction under these conditions. In this paper, we prove that a very simple and efficient algorithm, known as Iterative Hard Thresholding, has near optimal performance guarantees rivaling those derived for other state of the art approaches.

**Index Terms**— Compressed Sensing, Iterative Hard Thresholding, Sparse Inverse Problem

## 1. INTRODUCTION

Compressed sensing [1], [2], [3], [4] is a technique to sample finite dimensional signals below the Nyquist rate. A signal  $f$  from an  $N < \infty$  dimensional Hilbert space is sampled using  $M$  linear measurements  $\{\langle f, \phi_n \rangle\}$ , where  $\langle \cdot, \cdot \rangle$  is the inner product and  $\phi_n$  are elements from the Hilbert space under consideration. Let  $\mathbf{x}$  be the vector of elements  $x_i$  such that  $f = \sum_{i=1}^N \psi_i x_i$  for some orthonormal basis  $\psi_i$  of the signal space so that  $f$  and  $\mathbf{x}$  are equivalent. Let  $\Phi \in \mathbb{R}^{M \times N}$  be the matrix with entries  $\langle \psi_i, \phi_j \rangle$  so that the observation can then be written as

$$\mathbf{y} = \Phi \mathbf{x} + \mathbf{e}, \quad (1)$$

where  $\mathbf{e}$  is observation noise.

In compressed sensing, we take fewer measurements than the dimension of the Hilbert space under consideration, that is,  $M < N$ . The problem of recovering  $\mathbf{x}$  from  $\mathbf{y}$  is therefore underdetermined. To overcome this problem, it is assumed that  $\mathbf{x}$  is well approximated by a  $K$ -sparse vector, that is, by a vector with only  $K$  non-zero elements, where  $K < M$ . Under this assumption, the problem of recovering  $\mathbf{x}$ , given the samples  $\mathbf{y}$  and the observation model  $\Phi$  can be posed as the following optimisation problem

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}: \|\mathbf{x}\|_0 \leq K} \|\mathbf{y} - \Phi \mathbf{x}\|_2, \quad (2)$$

where  $\|\mathbf{x}\|_0$  stands for the number of non-zero elements in the vector  $\mathbf{x}$ . We are therefore looking through all vectors  $\mathbf{x}$  with no more than

This research was supported by EPSRC grants D000246/2 and EP/F039697/1. MED acknowledges support of his position from the Scottish Funding Council and their support of the Joint Research Institute with the Heriot-Watt University as a component part of the Edinburgh Research Partnership.

$K$  non-zero elements to find that vector that minimises the squared error  $\|\mathbf{y} - \Phi \mathbf{x}\|_2$ . Whilst this is an NP-hard optimisation problem in general, several contributions have studied conditions on  $\Phi$  that allow this problem to be solved approximately using efficient algorithms. We state two of these results in section 3, before deriving similar results for an Iterative Hard Thresholding Algorithm in section 4. We here derive the main results only. More details are available in our technical report on arXiv [5].

## 2. THEORETICAL PRELIMINARIES

Two preliminary issues have to be discussed. Firstly, a condition on  $\Phi$  used in the results of this paper is introduced and, secondly, particular properties of best  $K$ -term approximations are stated.

### 2.1. Conditions on $\Phi$

The analysis of algorithms for compressed sensing relies heavily on a property<sup>1</sup> of  $\Phi$  known as the restricted isometry property (RIP). The *restricted isometry constant* is the smallest quantity  $\delta_K$  for which the following inequalities hold [1]

$$(1 - \delta_K) \|\mathbf{x}\|_2^2 \leq \|\Phi \mathbf{x}\|_2^2 \leq (1 + \delta_K) \|\mathbf{x}\|_2^2 \quad (3)$$

for all vectors  $\mathbf{x}$  with no more than  $K$  non-zero elements. If  $\delta_K > 0$ , the matrix  $\Phi$  is said to satisfy the Restricted Isometry Property (RIP) of order  $K$ .

Our results in this paper are based on a re-scaled matrix  $\hat{\Phi} = \frac{\Phi}{1 + \delta_K}$ , which satisfies the following asymmetric isometry property

$$(1 - \beta_K) \|\mathbf{x}\|_2^2 \leq \|\hat{\Phi} \mathbf{x}\|_2^2 \leq \|\mathbf{x}\|_2^2 \quad (4)$$

for all  $K$ -sparse  $\mathbf{x}$ , where  $\beta_K = 1 - \frac{1 - \delta_K}{1 + \delta_K}$ .

Throughout this paper, when referring to the RIP, we mean in general this slightly modified version for which we always use the letter  $\beta$ . If we need to refer to the standard RIP, for the non-normalised matrix  $\hat{\Phi}$ , we use the letter  $\delta$ .

### 2.2. Best $K$ term approximations

Compressed sensing relies on the signal  $\mathbf{x}$  to be well approximated by a  $K$ -sparse vector. We write  $\mathbf{x}_K$  to mean the best  $K$ -term approximation to the vector  $\mathbf{x}$ . The following results are stated here without proof, which follow those presented in [7].

<sup>1</sup>Note that there exist matrices with this property whenever  $M \geq cK \log(N/K)$  for some constant  $c$ . For example, if we generate matrices with  $M \geq 36/7\delta_K (\ln(eN/K) + K \ln(12/\delta_K) + \ln(2) + t)$  rows, by drawing the elements from an i.i.d. normal distribution, then it has a restricted isometry constant of  $\delta_K$  with probability of at least  $1 - e^{-t}$  [6].

**Lemma 1** (Needell and Tropp, Proposition 3.5 in [7]). *Suppose we have a matrix  $\Phi$  that satisfies the inequality  $\|\Phi \mathbf{x}_K\|_2 \leq \|\mathbf{x}_K\|_2$  for all  $K$ -sparse vectors. For such a matrix and for all vectors  $\mathbf{x}$ , the following inequality holds*

$$\|\Phi \mathbf{x}\|_2 \leq \|\mathbf{x}\|_2 + \frac{1}{\sqrt{K}} \|\mathbf{x}\|_1. \quad (5)$$

**Lemma 2** (Needell and Tropp, lemma 6.1 in [7]). *Given any  $\mathbf{x}$ , consider the (any) best  $K$ -term approximation to  $\mathbf{x}$ , denoted by  $\mathbf{x}_K$ . Define the error between the best  $K$ -term approximation and  $\mathbf{x}$  as  $\mathbf{x}_r = \mathbf{x} - \mathbf{x}_K$  and the ‘error’ term  $\tilde{\mathbf{e}} = \Phi \mathbf{x}_r + \mathbf{e}$ , such that*

$$\mathbf{y} = \Phi \mathbf{x} + \mathbf{e} = \Phi \mathbf{x}_K + \tilde{\mathbf{e}}. \quad (6)$$

*If  $\Phi$  satisfies the restricted isometry property for sparsity  $K$ , then the norm of the error  $\tilde{\mathbf{e}}$  can be bounded by*

$$\|\tilde{\mathbf{e}}\|_2 \leq \|\mathbf{x} - \mathbf{x}_K\|_2 + \frac{1}{\sqrt{K}} \|\mathbf{x} - \mathbf{x}_K\|_1 + \|\mathbf{e}\|_2 \quad (7)$$

### 3. BASIS PURSUIT, SUBSPACE PURSUIT AND COSAMP

Of all the methods proposed for the sparse recovery problem in compressed sensing, two stand out in terms of the performance bounds available. The first method is to solve the convex Basis Pursuit De-noising optimisation proposed in [8] and the second one is the Compressed Sensing Matching Pursuit (CoSaMP) algorithm [7]. Whilst a detailed discussion on these approaches has to be omitted, we here present the main theoretical results derived for these two approaches.

We have included these results here to show that, not only have both algorithms similar performance guarantees under similar conditions, but, importantly, the guarantees and conditions are also comparable to those derived in the next section for the Iterative Hard Thresholding algorithm.

To summarise the results of the next two subsections, both, Basis Pursuit De-noising and CoSaMP are able to estimate any vector  $\mathbf{x}$  from  $\mathbf{y} = \Phi \mathbf{x} + \mathbf{e}$  if  $\Phi$  has a bounded restricted isometry constant. The accuracy of this estimate depends on the size of the error  $\mathbf{e}$  as well as on the error between  $\mathbf{x}$  and its best  $K$  term approximation, that is, if  $\mathbf{x}$  is well approximated by a sparse vector, both algorithms are able to calculate a good approximation to  $\mathbf{x}$ .

#### 3.1. Basis Pursuit De-Noising

Instead of solving the non-convex optimisation problem (2), Basis Pursuit De-Noising [8] solves the convex optimisation problem

$$\mathbf{x}^* = \min_{\mathbf{x}} \|\mathbf{x}\|_1 \text{ such that } \|\mathbf{y} - \Phi \mathbf{x}\|_2 \leq \epsilon.$$

Due to the convexity of this problem, optimisation is much easier than optimisation of (2) and several efficient algorithms have been proposed over the years.

Candès [9] derived the following result for the quality of the Basis Pursuit De-Noising solution.

**Theorem 3** ([9]). *If  $\Phi$  has a restricted isometry constant of  $\delta_{2K} < 0.2$ , then the solution  $\mathbf{x}^*$  to the convex Basis Pursuit De-Noising optimisation problem with  $\epsilon = \|\mathbf{e}\|_2$  obeys*

$$\|\mathbf{x}^* - \mathbf{x}\|_2 \leq 8.5 \left[ \frac{\|\mathbf{x} - \mathbf{x}_K\|_1}{\sqrt{K}} + \|\mathbf{e}\|_2 \right]. \quad (8)$$

#### 3.2. Subspace Pursuit and CoSaMP

Subspace Pursuit [10] and CoSaMP [7] are very similar algorithms which share many of their theoretical properties. The following result, similar to the one derived for Basis Pursuit De-Noising, has been derived for CoSaMP in [7].

**Theorem 4** ([7]). *Assume  $\Phi$  has a restricted isometry constant of  $\delta_{4K} \leq 0.1$ , then, after at most  $6(K+1)$  iterations, CoSaMP produces an approximation  $\mathbf{x}^*$  that satisfies*

$$\|\mathbf{x}^* - \mathbf{x}\|_2 \leq 20 \left[ \|\mathbf{x} - \mathbf{x}_K\|_2 + \frac{\|\mathbf{x} - \mathbf{x}_K\|_1}{\sqrt{K}} + \|\mathbf{e}\|_2 \right]. \quad (9)$$

### 4. ITERATIVE HARD THRESHOLDING

The Iterative Hard Thresholding algorithm ( $IHT_K$ ) is the following iterative procedure. Let  $\mathbf{x}^{[0]} = \mathbf{0}$  and use the iteration

$$\mathbf{x}^{[n+1]} = H_K(\mathbf{x}^{[n]} + \Phi^T(\mathbf{y} - \Phi \mathbf{x}^{[n]})), \quad (10)$$

where  $H_K(\mathbf{a})$  is the non-linear operator that sets all but the largest (in magnitude)  $K$  elements of  $\mathbf{a}$  to zero. If there is no unique such set, a set can be selected either randomly or based on a predefined ordering of the elements. Inspired by the work in [11], we first formally studied convergence properties of this algorithm in [12]. In this paper we prove that this algorithm has similar performance guarantees to Basis Pursuit De-Noising and CoSaMP.

#### 4.1. Main Result

The main result is the following theorem that bounds the performance of the algorithm for arbitrary (not necessary sparse) signals  $\mathbf{x}$ .

**Theorem 5.** *Let  $\mathbf{x}$  be any arbitrary vector, which is observed through a linear measurement system  $\Phi$  that satisfies the restricted isometry property with  $\beta_{3K} < 1/\sqrt{32}$ . Given the noisy observation  $\mathbf{y} = \Phi \mathbf{x} + \mathbf{e}$ , then after at most*

$$\left\lceil \log_2 \left( \frac{\|\mathbf{x}_K\|_2}{\|\mathbf{x} - \mathbf{x}_K\|_2 + \|\mathbf{x} - \mathbf{x}_K\|_1/\sqrt{K} + \|\mathbf{e}\|_2} \right) \right\rceil \quad (11)$$

iterations,  $IHT_K$  calculates an estimate  $\mathbf{x}^*$  that satisfies

$$\|\mathbf{x}^* - \mathbf{x}\|_2 \leq 6 \left[ \|\mathbf{x} - \mathbf{x}_K\|_2 + \frac{\|\mathbf{x} - \mathbf{x}_K\|_1}{\sqrt{K}} + \|\mathbf{e}\|_2 \right], \quad (12)$$

where  $\mathbf{x}_K$  is the best  $K$ -term approximation to  $\mathbf{x}$ .

#### 4.2. Comparison to CoSaMP and Basis Pursuit De-Noising

Several remarks are in order. Firstly, for  $IHT_K$ , we require  $\beta_{3K} \leq 0.175$ , whilst for CoSaMP,  $\delta_{4K} \leq 0.1$  is required. Now,  $\beta$  and  $\delta$  are related as follows  $\frac{\beta}{2-\beta} = \delta$ , therefore,  $IHT_K$  requires  $\delta_{3K} \leq 0.0959 \approx 0.1$ . As  $\delta(k)$  is an increasing function of  $(k)$ , the condition on  $\delta_{4K}$  in the theorem for CoSaMP is somewhat stricter than the condition on  $\delta_{3K}$  in the  $IHT_K$  result.

Secondly, the convergence of  $IHT_K$  is linear so that the number of iterations required depends logarithmically on the ‘signal to noise ratio’  $\frac{\|\mathbf{x}_K\|_2}{\|\mathbf{x} - \mathbf{x}_K\|_2 + \|\mathbf{x} - \mathbf{x}_K\|_1/\sqrt{K} + \|\mathbf{e}\|_2}$ . To achieve the iteration bound of  $6(K+1)$ , given for CoSaMP, the algorithm requires the solution to an inverse problem in each iteration, which is costly in

terms of the required computations.  $IHT_K$  does not require the solution to an inverse problem, instead, each iteration only requires the application of the operators  $\Phi$  and  $\Phi^T$  once each. In many practical applications,  $\Phi$  is chosen such that these multiplications can be computed efficiently<sup>2</sup>. CoSaMP can also be implemented using a partial solution to the required inverse problems in which case the iteration count guarantees become similar to the ones derived here for  $IHT_K$ .

Finally, for both, CoSaMP and  $IHT_K$ , the results are in terms of the error  $\left[\|\mathbf{x} - \mathbf{x}_K\|_2 + \|\mathbf{x} - \mathbf{x}_K\|_1/\sqrt{K} + \|\mathbf{e}\|_2\right]$ . Interestingly, for Basis Pursuit De-Noising, the error only depends on  $\|\mathbf{x} - \mathbf{x}_K\|_1/\sqrt{K} + \|\mathbf{e}\|_2$ , but not on  $\|\mathbf{x} - \mathbf{x}_K\|_2$ . As CoSaMP and  $IHT_K$  do calculate  $K$ -sparse approximation, it is clear that the best attainable error will have to depend on the term  $\|\mathbf{x} - \mathbf{x}_K\|_2$ . As Basis Pursuit De-Noising does not necessarily give a sparse result, this dependency does not seem to exist.

### 4.3. Optimality of Main Result

In fact, the results in this paper are optimal in the sense that we cannot do substantially better, even if an oracle would give us the support set of the  $K$  largest coefficients in  $\mathbf{x}$ . To see this, let  $\Phi_K$  be the sub-matrix of  $\Phi$  containing only those columns associated with the  $K$  largest coefficients in  $\mathbf{x}$ . The best  $K$ -term approximation to  $\mathbf{x}$  can then be written as  $\Phi_K^\dagger \mathbf{y}$ , where the  $\dagger$  signifies the pseudo inverse. The error for such an oracle estimate would be

$$\begin{aligned} \|\mathbf{x} - \Phi_K^\dagger \mathbf{y}\|_2 &\leq \|\mathbf{x} - \mathbf{x}_K - \Phi_K^\dagger \Phi(\mathbf{x} - \mathbf{x}_K) - \Phi_K^\dagger \mathbf{e}\|_2 \\ &\leq \|\mathbf{x} - \mathbf{x}_K\|_2 + \|\Phi_K^\dagger \Phi(\mathbf{x} - \mathbf{x}_K)\|_2 + \|\Phi_K^\dagger \mathbf{e}\|_2 \\ &\leq \|\mathbf{x} - \mathbf{x}_K\|_2 + \|\Phi_K^\dagger\|_2 \|\Phi(\mathbf{x} - \mathbf{x}_K)\|_2 \\ &\quad + \|\Phi_K^\dagger\|_2 \|\mathbf{e}\|_2 \\ &\leq \left(1 + \frac{1}{\sqrt{1 - \beta_K}}\right) \|\mathbf{x} - \mathbf{x}_K\|_2 \\ &\quad + \frac{1}{\sqrt{1 - \beta_K}} \frac{\|\mathbf{x} - \mathbf{x}_K\|_1}{\sqrt{K}} \\ &\quad + \frac{1}{\sqrt{1 - \beta_K}} \|\mathbf{e}\|_2, \end{aligned} \quad (13)$$

where the bound on  $\|\Phi_K^\dagger\|_2$  can be found in, for example, [7] and where the bound on  $\|\Phi(\mathbf{x} - \mathbf{x}_K)\|_2$  is due to lemma 1.

## 5. PROOF OF THE MAIN RESULT

Let us introduce the following notation.

- $\mathbf{r}^{[n]} = \mathbf{x}_K - \mathbf{x}^{[n]}$ ,
- $\mathbf{a}^{[n+1]} = \mathbf{x}^{[n]} + \Phi^T(\mathbf{y} - \Phi \mathbf{x}^{[n]}) = \mathbf{x}^{[n]} + \Phi^T(\Phi \mathbf{x}_K + \tilde{\mathbf{e}} - \Phi \mathbf{x}^{[n]})$ ,
- $\mathbf{y}^{[n+1]} = H_K(\mathbf{a}^{[n+1]})$ ,
- $\Gamma_K^* = \text{supp}\{\mathbf{x}_K\}$ ,
- $\Gamma^n = \text{supp}\{\mathbf{x}^{[n]}\}$ ,
- $B^{n+1} = \Gamma_K^* \cup \Gamma^{n+1}$

Of particular importance are the following sets.  $\Gamma_K^*$  are the non-zero elements in the best  $K$  term approximation to  $\mathbf{x}$ , in effect, this is the set we are striving to identify. The set  $\Gamma^n$  contains the indices for the  $K$  non-zero elements in iteration  $n$  and finally,  $B^n$  is the union of these two sets.

<sup>2</sup>For example, using the fast Fourier transform.

### 5.1. Preliminary lemmata

The results in this paper require the following lemmata. The first set of tools bounds different operator norms involving sub-matrices of  $\Phi$  using the restricted isometry property. These results are known and have been derived in, for example, [7]. We have to note that the restricted isometry constant bounds (from above and below) the largest and smallest singular values of all sub-matrices of  $\Phi$  with  $K$  columns. This directly implies the following inequalities. For all index sets  $\Gamma$  with  $K = |\Gamma|$  and all  $\Phi$  for which the RIP holds for sparsity  $K$  we have the bounds

$$\|\Phi_\Gamma^T \mathbf{y}\|_2 \leq \|\mathbf{y}\|_2, \quad (14)$$

$$(1 - \beta_{|\Gamma|}) \|\mathbf{x}_\Gamma\|_2 \leq \|\Phi_\Gamma^T \Phi_\Gamma \mathbf{x}_\Gamma\|_2 \leq \|\mathbf{x}_\Gamma\|_2. \quad (15)$$

We here use the notation  $\Phi_\Gamma$  and  $\mathbf{x}_\Gamma$  to refer to the sub-matrices and sub-vectors of  $\Phi$  and  $\mathbf{x}$  containing only those columns (elements) with indices in the set  $\Gamma$ .

We also need the following two lemmata.

**Lemma 6.** *If a matrix  $\Phi$  satisfies RIP for sparsity  $K$ , then for all sets  $\Gamma$  with  $|\Gamma| = K$  we have the bound*

$$\|(\mathbf{I} - \Phi_\Gamma^T \Phi_\Gamma) \mathbf{x}_\Gamma\|_2 \leq \beta_K \|\mathbf{x}_\Gamma\|_2. \quad (16)$$

This result has been established as a by-product in the proof of Proposition 3.2 in [7], though again for the slightly different definition of the restricted isometry property.

The next inequality can also be found in [7].

**Lemma 7.** *If a matrix  $\Phi$  satisfies RIP for sparsity  $K$ , then for any two disjoint sets  $\Gamma$  and  $\Lambda$  (i.e.  $\Gamma \cap \Lambda = \emptyset$ ) for which  $|\Gamma \cup \Lambda| = K$  we have the following inequality*

$$\|\Phi_\Gamma^T \Phi_\Lambda \mathbf{x}_\Lambda\|_2 \leq \beta_K \|\mathbf{x}_\Lambda\|_2. \quad (17)$$

The next result can be found in, for example, [9].

**Lemma 8.** *For any two orthogonal vectors  $\mathbf{r}_1$  and  $\mathbf{r}_2$ , the following holds*

$$\|\mathbf{r}_1\|_2 + \|\mathbf{r}_2\|_2 \leq \sqrt{2} \|\mathbf{r}_1 + \mathbf{r}_2\|_2. \quad (18)$$

*Proof.* Note that  $\|\mathbf{r}_1\|_2 + \|\mathbf{r}_2\|_2$  is the one norm of the vector  $[\|\mathbf{r}_1\|_2 \|\mathbf{r}_2\|_2]^T$ . Using standard norm inequalities, we have

$$\left\| \begin{bmatrix} \|\mathbf{r}_1\|_2 \\ \|\mathbf{r}_2\|_2 \end{bmatrix} \right\|_1 \leq \sqrt{2} \left\| \begin{bmatrix} \|\mathbf{r}_1\|_2 \\ \|\mathbf{r}_2\|_2 \end{bmatrix} \right\|_2. \quad (19)$$

Due to orthogonality,  $\|\mathbf{r}_1\|_2^2 + \|\mathbf{r}_2\|_2^2 = \|\mathbf{r}_1 + \mathbf{r}_2\|_2^2$ , which completes the proof.  $\square$

### 5.2. Proof of the error bound in theorem 5

*Proof.* We want to bound the error

$$\|\mathbf{x} - \mathbf{x}^{[n+1]}\|_2 \leq \|\mathbf{x} - \mathbf{x}_K\|_2 + \|\mathbf{x}_K - \mathbf{x}^{[n+1]}\|_2. \quad (20)$$

We now bound the right hand term using the triangle inequality

$$\|\mathbf{x}_K - \mathbf{x}^{[n+1]}\|_2 \leq \|\mathbf{x}_{K, B^{n+1}} - \mathbf{a}_{B^{n+1}}^{[n+1]}\|_2 + \|\mathbf{x}_{B^{n+1}}^{[n+1]} - \mathbf{a}_{B^{n+1}}^{[n+1]}\|_2.$$

Crucially, as indicated by the subscripts above, the error  $\mathbf{x}_K - \mathbf{x}^{[n+1]}$  is supported on the set  $B^{n+1} = \Gamma_K^* \cup \Gamma^{n+1}$ . Furthermore,  $\mathbf{x}^{[n+1]}$

is the best  $K$ -term approximation to  $\mathbf{a}_{B^{n+1}}^{[n+1]}$  and is therefore a better approximation than  $\mathbf{x}_K$  so that  $\|\mathbf{x}^{[n+1]} - \mathbf{a}_{B^{n+1}}^{[n+1]}\|_2 \leq \|\mathbf{x}_K - \mathbf{a}_{B^{n+1}}^{[n+1]}\|_2$ . We therefore have

$$\|\mathbf{x}_K - \mathbf{x}^{[n+1]}\|_2 \leq 2\|\mathbf{x}_{K, B^{n+1}} - \mathbf{a}_{B^{n+1}}^{[n+1]}\|_2. \quad (21)$$

We can now expand  $\mathbf{a}_{B^{n+1}}^{[n+1]}$  and find that

$$\begin{aligned} & \|\mathbf{x}_K - \mathbf{x}^{[n+1]}\|_2 \\ & \leq 2\|\mathbf{x}_{K, B^{n+1}} - \mathbf{x}_{B^{n+1}}^{[n]} - \Phi_{B^{n+1}}^T \Phi \mathbf{r}^{[n]} - \Phi_{B^{n+1}}^T \tilde{\mathbf{e}}\|_2 \\ & \leq 2\|\mathbf{r}_{B^{n+1}}^{[n]} - \Phi_{B^{n+1}}^T \Phi \mathbf{r}^{[n]}\|_2 + 2\|\Phi_{B^{n+1}}^T \tilde{\mathbf{e}}\|_2 \\ & = 2\|(\mathbf{I} - \Phi_{B^{n+1}}^T \Phi_{B^{n+1}}) \mathbf{r}_{B^{n+1}}^{[n]} \\ & \quad - \Phi_{B^{n+1}}^T \Phi_{B^n \setminus B^{n+1}} \mathbf{r}_{B^n \setminus B^{n+1}}^{[n]}\|_2 \\ & \quad + 2\|\Phi_{B^{n+1}}^T \tilde{\mathbf{e}}\|_2 \\ & \leq 2\|(\mathbf{I} - \Phi_{B^{n+1}}^T \Phi_{B^{n+1}}) \mathbf{r}_{B^{n+1}}^{[n]}\|_2 \\ & \quad + 2\|(\Phi_{B^{n+1}}^T \Phi_{B^n \setminus B^{n+1}}) \mathbf{r}_{B^n \setminus B^{n+1}}^{[n]}\|_2 \\ & \quad + 2\|\Phi_{B^{n+1}}^T \tilde{\mathbf{e}}\|_2 \end{aligned}$$

As the set  $B^n \setminus B^{n+1}$  is disjoint from  $B^{n+1}$  and as  $|B^n \cup B^{n+1}| \leq 3K$ , we can make use of (14), (16) and (17) and the fact that  $\beta_{2K} \leq \beta_{3K}$

$$\|\mathbf{x}_K - \mathbf{x}^{[n+1]}\|_2 \quad (22)$$

$$\leq 2\beta_{2K} \|\mathbf{r}_{B^{n+1}}^{[n]}\|_2 + 2\beta_{3K} \|\mathbf{r}_{B^n \setminus B^{n+1}}^{[n]}\|_2 + 2\|\tilde{\mathbf{e}}\|_2 \quad (23)$$

$$\leq 2\beta_{3K} \left( \|\mathbf{r}_{B^{n+1}}^{[n]}\|_2 + \|\mathbf{r}_{B^n \setminus B^{n+1}}^{[n]}\|_2 \right) + 2\|\tilde{\mathbf{e}}\|_2 \quad (24)$$

Furthermore  $\mathbf{r}_{B^{n+1}}^{[n]}$  and  $\mathbf{r}_{B^n \setminus B^{n+1}}^{[n]}$  are orthogonal so that lemma 8 gives

$$\|\mathbf{x}_K - \mathbf{x}^{[n+1]}\|_2 \leq \sqrt{8}\beta_{3K} \|\mathbf{r}^{[n]}\|_2 + 2\|\tilde{\mathbf{e}}\|_2. \quad (25)$$

If we choose  $\beta_{3K} < \frac{1}{\sqrt{32}}$  we have the bound

$$\|\mathbf{x}_K - \mathbf{x}^{[n+1]}\|_2 = \|\mathbf{r}^{[n+1]}\|_2 < 0.5\|\mathbf{r}^{[n]}\|_2 + 2\|\tilde{\mathbf{e}}\|_2, \quad (26)$$

which we iterate, so that

$$\|\mathbf{x}_K - \mathbf{x}^{[k]}\|_2 < 2^{-k} \|\mathbf{x}_K\|_2 + 4\|\tilde{\mathbf{e}}\|_2 \quad (27)$$

and

$$\begin{aligned} \|\mathbf{x} - \mathbf{x}^{[n+1]}\|_2 & \leq 2^{-k} \|\mathbf{x}_K\|_2 + \|(\mathbf{x} - \mathbf{x}_K)\|_2 + 4\|\tilde{\mathbf{e}}\|_2 \\ & \leq 2^{-k} \|\mathbf{x}_K\|_2 + 5\|(\mathbf{x} - \mathbf{x}_K)\|_2 \\ & \quad + 4\frac{1}{\sqrt{K}} \|(\mathbf{x} - \mathbf{x}_K)\|_1 + 4\|\mathbf{e}\|_2 \end{aligned}$$

We are therefore guaranteed to reduce the error to below any multiple  $c$  of  $\|(\mathbf{x} - \mathbf{x}_K)\|_2 + \|(\mathbf{x} - \mathbf{x}_K)\|_1/\sqrt{K} + \|\mathbf{e}\|_2$ , as long as  $c > 5$ . For example,  $c = 6$  implies that we require that

$$2^{-k} \|\mathbf{x}_K\|_2 \leq \|(\mathbf{x} - \mathbf{x}_K)\|_2 + \|(\mathbf{x} - \mathbf{x}_K)\|_1/\sqrt{K} + \|\mathbf{e}\|_2, \quad (28)$$

i.e. that

$$2^k \geq \frac{\|\mathbf{x}_K\|_2}{\|(\mathbf{x} - \mathbf{x}_K)\|_2 + \|(\mathbf{x} - \mathbf{x}_K)\|_1/\sqrt{K} + \|\mathbf{e}\|_2}, \quad (29)$$

which in turn implies the second part of the theorem.  $\square$

## 6. CONCLUSIONS

Finite signals that are well approximated with sparse vector representations can be sampled below the Nyquist rate. If the sampling system satisfies the so called restricted isometry property and has a small restricted isometry constant, Compressed Sensing theory states that algorithms such as Basis Pursuit De-noising and CoSaMP can be used to recover the signal from the observations.

In this paper, we have shown that under similar conditions on the restricted isometry constant, a very simple and efficient Iterative Hard Thresholding algorithm has similar uniform performance guarantees than Basis Pursuit De-noising and CoSaMP. In particular, all algorithms are guaranteed to recover a vector  $\mathbf{x}$  from noisy measurements  $\mathbf{y} = \Phi \mathbf{x} + \mathbf{e}$  with similar error guarantees. In fact, this type of performance is optimal up to the constants.

## 7. REFERENCES

- [1] E. 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, Feb 2006.
- [2] E. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.
- [3] E. Candès and J. Romberg, "Quantitative robust uncertainty principles and optimally sparse decompositions," *Foundations of Comput. Math.*, vol. 6, no. 2, pp. 227 – 254, 2006.
- [4] D. Donoho, "Compressed sensing," *IEEE Trans. on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [5] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," Tech. Rep. arXiv:0805.0510v1, The University of Edinburgh, 2008.
- [6] T. Blumensath and M. E. Davies, "Sampling theorems for signals from the union of linear subspaces," *submitted to IEEE Transactions on Information Theory*, 2008.
- [7] D. Needell and J. Tropp, "COSAMP: Iterative signal recovery from incomplete and inaccurate samples.," *to appear in Appl. Comp. Harmonic Anal.*, 2008.
- [8] S. S. Chen, D. L. Donoho, and Michael A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal of Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.
- [9] E. Candès, "The restricted isometry property and its implications for compressed sensing," *Compte Rendus de l'Academie des Sciences, Paris, Serie I*, vol. 346, pp. 589592, 2008.
- [10] W. Dai and O. Milenkovic, "Subspace pursuit for compressed sensing: Closing the gap between performance and complexity," *submitted*, 2008.
- [11] N. G. Kingsbury and T. H. Reeves, "Iterative image coding with overcomplete complex wavelet transforms," in *Proc. Conf. on Visual Communications and Image Processing*, 2003.
- [12] T. Blumensath and M.E. Davies, "Iterative thresholding for sparse approximations," *to appear in Journal of Fourier Analysis and Applications, special issue on sparsity*, 2008.