1 Motivation

In previous classes, we have explored the topic of compressive sensing. Traditional techniques for signal acquisition involve acquiring $N$ samples of a signal sampled at a rate faster than twice the Nyquist rate in order to guarantee perfect signal reconstruction. Many signals of practical interest, however, are sparse in some basis, meaning that in some basis they can be represented with $K \ll N$ samples. Rather than initially acquire $N$ samples and then throw away $\approx N - K$ samples during compression, we would like to be able to sample the compressed signal directly.

This can be accomplished via compressive sensing techniques to produce a compressed signal $y$ by taking $M \approx K$ linear projections of the desired signal, $x$, via a projection matrix $\Phi$, as follows:

$$y = \Phi x,$$

where $x$ is an $N \times 1$ (sparse) vector, $\Phi$ is an $M \times N$ matrix, and $y$ is an $M \times 1$ vector.

For signal reconstruction to be successful, $\Phi$ must satisfy the restricted isometry property or RIP:

$$(1 - \delta_K)\|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1 + \delta_K)\|x\|_2^2,$$

where $\| \cdot \|_2$ defines the $\ell_2$ norm. Finding matrices that satisfies this property is, in general, an NP-complete problem. However, random matrices chosen with i.i.d random entries (Gaussian, Bernouli, etc) will satisfy the RIP with high probability subject to $M = O(K \log(N/K)) \ll N$.

Several algorithms exist for performing the signal reconstruction problem. Some of these include:
• Convex Optimization:
  – Basis Pursuit (BP)
  – Basis Pursuit De-Noising (BPDN)
  – Second-Order Cone Programming

• Iterative Greedy Algorithms
  – Matching Pursuit (MP)
  – Orthogonal Matching Pursuit (OMP)
  – CoSaMP

Here, we will focus our attention on the CoSaMP (Compressive Sampling Matching Pursuit) algorithm. In section 2 we briefly discuss some notational conventions that will be used throughout this document. In section 3 we present the steps of CoSaMP algorithm. Finally, in Section 4 we derive several lemmas regarding the convergence properties of the CoSaMP algorithm. In particular, we will show that the error in the CoSaMP estimate of a given sparse vector decays exponentially in the number of algorithm iterations.

2 Notation

We define some basic notation used throughout this scribe to eliminate confusion.

Vectors:

- Bolded lower case letters (\( x, y, \text{ etc} \)) are used exclusively for vectors.

- Double bars around a vector with an subscript \( p \) (ex. \( \| x \|_p \)) indicate the ell-\( p \) norm of the vector \( x \). \( \| x \|_p = \sqrt[p]{\sum_i |x_i|^p} \)

Matrices:

- Matrices are denoted with upper-case, bold letter (\( \Phi, \text{ etc} \)).

- \( \Phi^* \) indicates the Hermitian transpose of \( \Phi \)

- \( \Phi^\dagger \) indicates the pseudo-inverse of \( \Phi \) such that \( \Phi^\dagger = (\Phi^* \Phi)^{-1} \Phi^* \)
Sets:

-Captial letters (ex: $T$, $S$) are used to indicate sets.

- $|T|$ indicate the *cardinality* of the set, or the number of elements that it contains.

- $T^C$ indicates the compliment of set $T$

- $T \setminus S$ is used to indicate the set $T$ with all elements in $S$ removed.

- We use the subscript notation $|T$ to show that a vector or matrix is being restricted to only certain elements or columns. For example, $x_{|T}$ indicates the vector $x$ is restricted only the elements given in $T$. $\Phi_{|T^C}$ indicates that the matrix $\Phi$ is restricted to the columns contained in $T^C$.

- We often use the notation $x_N$ to indicate the best $N$-point support set of the vector $x$, $\text{supp}(x)$. That is, the set of indices that best represent the vector $x$ by minimizing $\|x - x_{|\text{supp}(x)}\|_2$

### 3 CoSaMP Algorithm

The following table lists the steps of the CoSaMP algorithm:
1) Initialization:
\[ x_{-1} = 0 \] (\( x_J \) is the estimate of \( x \)) at the \( J^{th} \) iteration
\[ r = y \] (the current residual)

2) Loop until convergence
i) Compute the current error: (Note that for Gaussian \( \Phi \), \( \Phi^T\Phi \) is \( \sim \) diagonal)
\[ e = \Phi^*r. \]

ii) Compute the best 2K support set of the error (index set):
\[ \Omega = e_{2K}. \]

iii) Merge the the strongest support sets:
\[ T = \Omega \cup \text{supp}(x_{J-1}). \]

iv) Perform a Least-Squares Signal Estimation:
\[ b_T = \Phi_T^\dagger y, \quad b_{T^c} = 0. \]

v) Prune \( x_J \) and compute \( r \) for next round:
\[ x_J = b_k, \]
\[ r = y - \Phi x_J. \]

4 Performance Bounds on CoSaMP algorithm

In this section, we prove several important lemmas regarding the CoSaMP algorithm and use these to prove convergence bounds for the algorithm. Specifically, we will show that the error in the CoSaMP estimate decays exponentially until achieving a lower bound dominated primarily by the noise power present in the measured signal.

4.1 Lemma 1

Suppose that \( \Phi \) obeys the RIP and has an isometry constant \( \delta_r \). Also, define \( T \) to be a set of \( r \) indices. Then:
∥\Phi_T^*u\parallel^2 \leq \sqrt{1 + \delta_r}\parallel u\parallel^2,  \hspace{1cm} (3) \\
\parallel \Phi_T^\dagger u\parallel^2 \leq \frac{1}{\sqrt{1 + \delta_r}}\parallel u\parallel^2,  \hspace{1cm} (4) \\
\sqrt{1 - \delta_r}\parallel u\parallel^2 \leq \parallel \Phi_T^*\Phi_T^* u\parallel^2 \leq \sqrt{1 + \delta_r}\parallel u\parallel^2,  \hspace{1cm} (5) \\
\frac{1}{\sqrt{1 - \delta_r}}\parallel u\parallel^2 \leq \parallel \Phi_T^*\Phi_T^* u\parallel^2 \leq \frac{1}{\sqrt{1 + \delta_r}}\parallel u\parallel^2.  \hspace{1cm} (6)

Proof: This follows directly from the RIP property which indicates that the singular values of \Phi|_T^* lie between \sqrt{1 + \delta_r} and \sqrt{1 - \delta_r}.

4.2 Lemma 2

Suppose that we have index sets \(S\) and \(T\) that are disjoint \((S \cap T = \emptyset)\). Now define \(R = S \cup T\) with \(|R| \leq r\).

\[\parallel \Phi_S^*\Phi_T\parallel^2 \leq \delta_r.\]  \hspace{1cm} (7)

Proof: This follows from the spectral norm inequality, which states that any given matrix \(M\) has a spectral norm greater than or equal to the spectral norm of any of its submatrices. For our case, if we define \(M = \Phi_R^*\Phi_R - I\) we can see that:

\[M = \begin{bmatrix} \Phi_S^*\Phi_T - I & \Phi_S^*\Phi_T \\ \Phi_T^*\Phi_S & \Phi_T^*\Phi_T - I \end{bmatrix}.\]  \hspace{1cm} (8)

Hence, \(\Phi_S^*\Phi_T\) is a submatrix of \(M\). By the spectral norm inequality:

\[\parallel \Phi_S^*\Phi_T\parallel \leq \parallel M\parallel.\]  \hspace{1cm} (9)

Applying the triangle inequality we also can see that:

\[\parallel M\parallel \leq \parallel \Phi_R^*\Phi_R\parallel + \parallel I\parallel.\]  \hspace{1cm} (10)

We combine the above with the results of Lemma 1 in order to obtain:
\[ \| \Phi^*_T \Phi_T \| \leq \| M \|, \] 
\[ \leq \max \{ (1 + \delta_r) - 1, 1 - (1 - \delta_r) \}, \] 
\[ = \delta_r. \]

which concludes the proof.

### 4.3 Lemma 3

Suppose that \( \Phi \) obeys the RIP and has isometry constant \( \delta_r \). Let \( T \) be a set of indices and \( x \) be a vector. Further suppose that \( r \geq |T \cup \text{supp}(x)| \) then:

\[ \| \Phi^*_T \Phi_T x \|_2 \leq \delta_r \| x \|_2. \] 

Proof: This follows directly from the spectral norm inequality with application of the results from Lemma 2:

\[ \| \Phi^*_T \Phi_T x \|_2 = \| \Phi^*_T (\Phi_T x) \|_2, \] 
\[ \leq \| \Phi_T^* \Phi_T \| \| x \|_2, \] 
\[ \leq \delta_r \| x \|_2. \]

which concludes the proof.

### 4.4 Lemma 4

Define:

\[
\begin{align*}
    s &= x - x_J \\
    r &= y - \Phi x_J = \Phi s + n \\
    e &= \Phi^* r \quad \text{and} \\
    \Omega &= e_{2K}
\end{align*}
\]

We will now show that:

\[ \| s \|_{\Omega^c} \leq 0.2223 \| r \|_2 + 2.34 \| n \|_2. \] 

Proof:
Defining \( R = \text{supp}(s) \) we can show that

\[
\|e - e_\Omega\|^2 \leq \|e - e_R\|^2, \tag{20}
\]

\[
\Rightarrow \sum_n (e(n) - e_\Omega(n))^2 \leq \sum_n (e(n) - e_R(n))^2, \tag{21}
\]

\[
\Rightarrow \sum_{n \notin \Omega} (e(n))^2 \leq \sum_{n \notin R} (e(n))^2, \tag{22}
\]

\[
\Rightarrow \sum_{n \in \Omega} (e(n))^2 \geq \sum_{n \in R} (e(n))^2, \tag{23}
\]

\[
\Rightarrow \|e_{\Omega \setminus R}\|^2 \geq \|e_{R \setminus \Omega}\|^2. \tag{24}
\]

Expanding the LHS of (25):

\[
\|e_{\Omega \setminus R}\|^2 = \|\Phi^*_R (\Phi s + n)\|^2, \tag{26}
\]

\[
\leq \|\Phi^*_R \Phi s\|^2 + \|\Phi^*_R n\|^2, \tag{27}
\]

\[
\leq \delta_{4K} \|s\|^2 + \sqrt{1 + \delta_{2K}} \|n\|^2. \tag{28}
\]

Expanding the RHS (25):

\[
\|e_{R \setminus \Omega}\|^2 = \|\Phi^*_R (\Phi s + n)\|^2, \tag{29}
\]

\[
\geq \|\Phi^*_{R \setminus \Omega} \Phi s\|^2 - \|\Phi^*_{R \setminus \Omega} n\|^2, \tag{30}
\]

\[
= \|\Phi^*_{R \setminus \Omega} \Phi s_{|R \setminus \Omega} \|^2 - \|\Phi^*_{R \setminus \Omega} \Phi s_{|(R \setminus \Omega)^c}\|^2 - \|\Phi^*_{R \setminus \Omega} n\|^2, \tag{31}
\]

\[
\geq (1 - \delta_{2K}) \|s_{|R \setminus \Omega}\|^2 - \delta_{2K} \|s\|^2 - \sqrt{1 + \delta_{2K}} \|n\|^2. \tag{32}
\]

Substituting the expansions of both the LHS and RHS into the initial inequality we find that:

\[
2.34 = \frac{2\sqrt{1 + \delta_{2K}}}{1 - \delta_{2K}}, \tag{33}
\]

\[
.2233 = \frac{\delta_{2K} + \delta_{4K}}{1 - \delta_{4K}}, \tag{34}
\]

\[
\delta_{2K} \leq \delta_{4K} \leq .1. \tag{35}
\]
4.5 Lemma 5

Define:

\[ b|_T = \Phi^\dagger|_T y \]
\[ b|_{TC} = 0 \]

then:

\[ \|x - b\|_2 \leq 1.112\|x|_{TC}\|_2 + 1.06\|n\|_2. \] (36)

Proof:

\[
\|x - b\|_2 \leq \|x|_{TC}\|_2 + \|x|_T - b|_T\|_2,
\]
(37)
\[
= \|x|_{TC}\|_2 + \|x|_T - \Phi^\dagger|_T (\Phi x|_T + \Phi x|_{TC} + n)\|_2,
\]
(38)
\[
= \Phi^\dagger|_T (\Phi x|_{TC} + n)\|_2,
\]
(39)
\[
\leq \|\Phi^\dagger|_T \Phi x|_{TC}\|_2 + \|\Phi^\dagger|_T n\|_2.
\]
(40)

We now recall that \( \Phi^\dagger|_T = \left( \Phi^*|_T \Phi|_T \right)^{-1} \Phi^*|_T \) and apply the results of Lemma 1 to arrive at:

\[
\|x - b\|_2 \leq \frac{\delta_{4K}}{1 - \delta_{3K}}\|x|_{TC}\|_2 + \frac{\|n\|_2}{\sqrt{1 - \delta_{3K}}}\|n\|_2,
\]
(41)
\[
= 1.112\|x|_{TC}\|_2 + 1.06\|n\|_2.
\]
(42)

4.6 Lemma 6

We wish to show that:

\[
\|x - x_J\|_2 \leq 2\|x - b\|_2.
\]
(43)

Proof:

We prove this lemma through application of the triangle inequality:
\[ \|x - x_J\|_2 = \|x - b + b - x_J\|_2, \quad (44) \]
\[ \leq \|x - b\|_2 + \|x_J - b\|_2, \quad (45) \]
\[ \leq 2\|x - b\|_2. \quad (46) \]

### 4.7 Final Result on Convergence

We now continue from the results of Lemma 6, substituting in our previous results, to show our final result:

\[ \|x - x_J\|_2 \leq 2\|x - b\|_2, \quad (47) \]
\[ \leq 2.224\|s_{|T^C}\|_2 + 2.12\|\mathbf{n}\|, \quad (48) \]
\[ \leq 2.224(.2223\|s\|_2 + 2.34\|\mathbf{n}\|) + 2.12\|\mathbf{n}\|, \quad (49) \]
\[ \leq .5\|s\|_2 + 7.5\|\mathbf{n}\|_2, \quad (50) \]
\[ = .5\|x - x_{J-1}\|_2 + 7.5\|\mathbf{n}\|_2. \quad (51) \]

What this result tells us is that the error in our approximation of \(x\) decays exponentially each iteration until finally reaching a limit bound by the noise power present in our signal. This makes the CoSaMP algorithm a very powerful tool in sparse signal reconstruction.