Generalized Orthogonal Matching Pursuit- A Review and Some New Results

Mrityunjoy Chakraborty

Department of Electronics and Electrical Communication Engineering Indian Institute of Technology, Kharagpur, INDIA

Table of Contents

Table of Contents

Table of Contents

Table of Contents

Table of Contents

Table of Contents

Table of Contents

Overview of Compressed Sensing (CS)

Overview of Compressed Sensing (CS)

- A powerful technique to represent signals at a sub-Nyquist sampling rate, provided the signal is known to be sparse in some domain.

Overview of Compressed Sensing (CS)

- A powerful technique to represent signals at a sub-Nyquist sampling rate, provided the signal is known to be sparse in some domain.
- It retains the capacity of perfect (or near perfect) reconstruction of the signal from fewer samples than provided by Nyquist rate sampling.

Overview of Compressed Sensing (CS)

- A powerful technique to represent signals at a sub-Nyquist sampling rate, provided the signal is known to be sparse in some domain.
- It retains the capacity of perfect (or near perfect) reconstruction of the signal from fewer samples than provided by Nyquist rate sampling.
- In last few years, the CS technique has attracted considerable attention from across a wide array of fields like

Overview of Compressed Sensing (CS)

- A powerful technique to represent signals at a sub-Nyquist sampling rate, provided the signal is known to be sparse in some domain.
- It retains the capacity of perfect (or near perfect) reconstruction of the signal from fewer samples than provided by Nyquist rate sampling.
- In last few years, the CS technique has attracted considerable attention from across a wide array of fields like
  1. applied mathematics,
  2. statistics, and
  3. engineering including signal processing areas like
     (i) MR imaging,
     (ii) speech processing,
     (iii) analog to digital conversion etc.

Basic CS Formulation:

- Let a real valued, bandlimited signal $u(t)$ be sampled following Nyquist sampling rate condition and over a finite observation interval, generating the observation vector: $\mathbf{u} = (u_1, u_2, \cdots, u_N)^T$.

Basic CS Formulation:

- Let a real valued, bandlimited signal $u(t)$ be sampled following Nyquist sampling rate condition and over a finite observation interval, generating the observation vector: $\mathbf{u} = (u_1, u_2, \cdots, u_N)^T$.

- Further, $\mathbf{u}$ is known to be sparse in some transform domain

  $\Rightarrow$ if $\mathbf{\Psi}$: The $N \times N$ transform matrix (usually unitary) and,
  $\mathbf{x}$: Transform coefficient vector ($\in \mathbb{R}^N$), so that
  $\mathbf{u} = \mathbf{\Psi}\mathbf{x}$,

Basic CS Formulation:

- Let a real valued, bandlimited signal $u(t)$ be sampled following Nyquist sampling rate condition and over a finite observation interval, generating the observation vector: $\mathbf{u} = (u_1, u_2, \cdots, u_N)^T$.

- Further, $\mathbf{u}$ is known to be sparse in some transform domain

  $\Rightarrow$ if $\mathbf{\Psi}$: The $N \times N$ transform matrix (usually unitary) and,
  $\mathbf{x}$: Transform coefficient vector ($\in \mathbb{R}^N$), so that

  $$\mathbf{u} = \mathbf{\Psi}\mathbf{x},$$

  then, $\mathbf{x}$ is K-sparse

  $\Rightarrow$ a maximum of K no. of terms in $\mathbf{x}$ are non-zero.

- Then, according to the CS theory, it is possible to recover $\mathbf{u}$ from a fewer no., say, M $(M < N)$ samples: $y_1, y_2, \cdots, y_M$, linearly related to $u_1, u_2, \cdots, u_N$ as
  $\mathbf{y} = \mathbf{Au} = \mathbf{\Phi x}$, $(\mathbf{\Phi} = \mathbf{A\Psi})$
  where,
  $\mathbf{y} = [y_1, y_2, \cdots, y_M]^T$, and
  $\mathbf{A}$: A $M \times N$ sensing matrix.

Overview of CS (Cont.)

- Then, according to the CS theory, it is possible to recover $\mathbf{u}$ from a fewer no., say, M ($M < N$) samples: $y_1, y_2, \cdots, y_M$, linearly related to $u_1, u_2, \cdots, u_N$ as
  $\mathbf{y} = \mathbf{Au} = \mathbf{\Phi x}$, $(\mathbf{\Phi} = \mathbf{A\Psi})$
  where,
  $\mathbf{y} = [y_1, y_2, \cdots, y_M]^T$, and
  $\mathbf{A}$: A $M \times N$ sensing matrix.
- Ideal approach to recover $\mathbf{x}$ is by l_0 minimization:

$$\min_{\mathbf{x} \in \mathbb{R}^N} ||\mathbf{x}||_0 \ \textit{subject to} \ \mathbf{y} = \mathbf{\Phi x}. \tag{1}$$

  - It provides the sparsest solution for $\mathbf{x}$.
  - Uniqueness of the K-sparse solution requires that every $2K$ columns of $\mathbf{\Phi}$ should be linearly independent.

- Then, according to the CS theory, it is possible to recover $\mathbf{u}$ from a fewer no., say, M ($M < N$) samples: $y_1, y_2, \cdots, y_M$, linearly related to $u_1, u_2, \cdots, u_N$ as
  $\mathbf{y} = \mathbf{Au} = \mathbf{\Phi x}$, $(\mathbf{\Phi} = \mathbf{A\Psi})$
  where,
  $\mathbf{y} = [y_1, y_2, \cdots, y_M]^T$, and
  $\mathbf{A}$: A $M \times N$ sensing matrix.
- Ideal approach to recover $\mathbf{x}$ is by l_0 minimization:

$$\min_{\mathbf{x} \in \mathbb{R}^N} ||\mathbf{x}||_0 \ \textit{subject to} \ \mathbf{y} = \mathbf{\Phi x}. \tag{1}$$

  - It provides the sparsest solution for $\mathbf{x}$.
  - Uniqueness of the K-sparse solution requires that every $2K$ columns of $\mathbf{\Phi}$ should be linearly independent.

But, it is a non-convex problem and is NP-hard.

Overview of CS (Cont.)

- More practical approaches using l_1 norm (and above) can find the desired K-sparse solution.
  - But this requires $\mathbf{\Phi}$ to satisfy certain "Restricted Isometry Property (RIP)".

Overview of CS (Cont.)

- More practical approaches using l_1 norm (and above) can find the desired K-sparse solution.
  - But this requires $\mathbf{\Phi}$ to satisfy certain "Restricted Isometry Property (RIP)".
  - A matrix $\mathbf{\Phi}$ is said to satisfy the RIP of order K if there exists a "Restricted Isometry Constant" $\delta_K \in (0,\ 1)$ so that

$$(1-\delta_K)\,||\mathbf{x}||_2^2 \leqslant ||\mathbf{\Phi x}||_2^2 \leqslant (1+\delta_K)\,||\mathbf{x}||_2^2 \qquad (2)$$

for all K-sparse $\mathbf{x}$.

Overview of CS (Cont.)

- More practical approaches using l_1 norm (and above) can find the desired K-sparse solution.
  - But this requires $\mathbf{\Phi}$ to satisfy certain "Restricted Isometry Property (RIP)".
    - A matrix $\mathbf{\Phi}$ is said to satisfy the RIP of order K if there exists a "Restricted Isometry Constant" $\delta_K \in (0,\ 1)$ so that

$$(1-\delta_K)\,||\mathbf{x}||_2^2 \leqslant ||\mathbf{\Phi x}||_2^2 \leqslant (1+\delta_K)\,||\mathbf{x}||_2^2 \tag{2}$$

      for all K-sparse $\mathbf{x}$.
    - The constant δ_K is taken as the smallest number from $(0,\ 1)$ for which the RIP is satisfied.

- More practical approaches using l_1 norm (and above) can find the desired K-sparse solution.
  - But this requires $\mathbf{\Phi}$ to satisfy certain "Restricted Isometry Property (RIP)".
    - A matrix $\mathbf{\Phi}$ is said to satisfy the RIP of order K if there exists a "Restricted Isometry Constant" $\delta_K \in (0,\ 1)$ so that

$$(1-\delta_K)\,||\mathbf{x}||_2^2 \leqslant ||\mathbf{\Phi x}||_2^2 \leqslant (1+\delta_K)\,||\mathbf{x}||_2^2 \qquad (2)$$

      for all K-sparse $\mathbf{x}$.
    - The constant δ_K is taken as the smallest number from $(0,\ 1)$ for which the RIP is satisfied.
    - If $\mathbf{\Phi}$ satisfies RIP of order K, then it also satisfies RIP for any order L where $L < K$ and that $\delta_L \geq \delta_K$.

Overview of CS (Cont.)

- More practical approaches using l_1 norm (and above) can find the desired K-sparse solution.
  - But this requires $\mathbf{\Phi}$ to satisfy certain "Restricted Isometry Property (RIP)".
    - A matrix $\mathbf{\Phi}$ is said to satisfy the RIP of order K if there exists a "Restricted Isometry Constant" $\delta_K \in (0,\ 1)$ so that

$$(1-\delta_K)\,||\mathbf{x}||_2^2 \leqslant ||\mathbf{\Phi x}||_2^2 \leqslant (1+\delta_K)\,||\mathbf{x}||_2^2 \qquad (2)$$

      for all K-sparse $\mathbf{x}$.
    - The constant δ_K is taken as the smallest number from $(0,\ 1)$ for which the RIP is satisfied.
    - If $\mathbf{\Phi}$ satisfies RIP of order K, then it also satisfies RIP for any order L where $L < K$ and that $\delta_L \geq \delta_K$.
    - Simple choice of random matrices for $\mathbf{\Phi}$ can make it satisfy RIP with high probability.

CS Reconstruction Methods

CS Reconstruction Methods

Convex Relaxation:

CS Reconstruction Methods

Convex Relaxation:

- It replaces the l_0 norm in (1) by l_1 norm to reduce the problem to a convex problem.

CS Reconstruction Methods

Convex Relaxation:

- It replaces the l_0 norm in (1) by l_1 norm to reduce the problem to a convex problem.
- Three main directions under this category, namely the basis pursuit (BP) , the basis pursuit de-noising (BPDN) and the least absolute shrinkage and selection operator (LASSO):

1. *BP:* $$\min_{\mathbf{x}\in\mathbb{R}^N} ||\mathbf{x}||_1 \ \textit{subject to} \ \mathbf{y} = \mathbf{\Phi x} \tag{3}$$
2. *BPDN:* $$\min_{\mathbf{x}\in\mathbb{R}^N} \lambda\, ||\mathbf{x}||_1 + ||\mathbf{r}||_2^2 \ \textit{s.t} \ \mathbf{r} = \mathbf{y} - \mathbf{\Phi x} \tag{4}$$
3. *LASSO:* $$\min_{\mathbf{x}\in\mathbb{R}^N} ||\mathbf{y} - \mathbf{\Phi x}||_2^2 \ \textit{s.t} \ ||\mathbf{x}||_1 \leqslant \epsilon \tag{5}$$

- The BP problem can be solved by standard polynomial time algorithms of linear programming (LP) methods.

Overview of CS (Cont.)

- The BP problem can be solved by standard polynomial time algorithms of linear programming (LP) methods.
- The exact K-sparse signal reconstruction by BP algorithm based on RIP was first investigated by E. Candès et.al. (2006) with the following bound on δ_K

$$\delta_K + \delta_{2K} + \delta_{3K} < 1 \tag{6}$$

Overview of CS (Cont.)

- The BP problem can be solved by standard polynomial time algorithms of linear programming (LP) methods.
- The exact K-sparse signal reconstruction by BP algorithm based on RIP was first investigated by E. Candès et.al. (2006) with the following bound on δ_K

$$\delta_K + \delta_{2K} + \delta_{3K} < 1 \tag{6}$$

- Later the bound was refined by S. Foucart (2010) as $\delta_{2K} < 0.4652$.

Overview of CS (Cont.)

- The BP problem can be solved by standard polynomial time algorithms of linear programming (LP) methods.
- The exact K-sparse signal reconstruction by BP algorithm based on RIP was first investigated by E. Candès et.al. (2006) with the following bound on δ_K

$$\delta_K + \delta_{2K} + \delta_{3K} < 1 \tag{6}$$

- Later the bound was refined by S. Foucart (2010) as $\delta_{2K} < 0.4652$.
- The BPDN and LASSO problems can be solved by efficient quadratic programming (QP) like primal-dual interior method.

- The BP problem can be solved by standard polynomial time algorithms of linear programming (LP) methods.
- The exact K-sparse signal reconstruction by BP algorithm based on RIP was first investigated by E. Candès et.al. (2006) with the following bound on δ_K

$$\delta_K + \delta_{2K} + \delta_{3K} < 1 \tag{6}$$

- Later the bound was refined by S. Foucart (2010) as $\delta_{2K} < 0.4652$.
- The BPDN and LASSO problems can be solved by efficient quadratic programming (QP) like primal-dual interior method.
- However, the regularization parameters λ and ϵ play a crucial role in the performance of these algorithms.

Greedy Pursuits:

- This approach recovers the K-sparse signal by iteratively constructing the support set of the sparse signal (index of non-zero elements in the sparse vector).

Overview of CS (Cont.)

Greedy Pursuits:

- This approach recovers the K-sparse signal by iteratively constructing the support set of the sparse signal (index of non-zero elements in the sparse vector).
- At each iteration, it updates its support set by appending the index of one or more columns (called atoms) of the matrix $\mathbf{\Phi}$ (often called dictionary) by some greedy principles based on the correlation between current residual of observation vector and the atoms.

Overview of CS (Cont.)

Greedy Pursuits:

- Few examples of greedy algorithms:
  1. Orthogonal Matching Pursuit (OMP): $\delta_{K+1} < \frac{1}{\sqrt{2K}}$
  2. Compressive Sampling Matching Pursuit(CoSaMP):$\delta_{4K} < 0.1$
  3. Subspace Pursuit(SP): $\delta_{3K} < 0.165$
  4. Iterative Hard Thresholding (IHT): $\delta_{3K} < \frac{1}{\sqrt{32}} \approx 0.177$.
  5. Generalized Orthogonal Matching Pursuit (gOMP): $\delta_{NK} < \frac{\sqrt{N}}{\sqrt{K}+3\sqrt{N}}$ where $N(\geq 1)$ is the number of atoms selected by the gOMP algorithm in each iteration [1].

– OMP reconstructs the K-sparse signal in K steps by selecting one atom in each iteration.

Overview of CS (Cont.)

- OMP reconstructs the K-sparse signal in K steps by selecting one atom in each iteration.
- CoSaMP and SP select a fixed number of atoms ($2K$ in CoSaMP and K in SP, for K-sparse signal) in each iteration while keeping the provision of rejecting a previously selected atom.

Overview of CS (Cont.)

– OMP reconstructs the K-sparse signal in K steps by selecting one atom in each iteration.

– CoSaMP and SP select a fixed number of atoms ($2K$ in CoSaMP and K in SP, for K-sparse signal) in each iteration while keeping the provision of rejecting a previously selected atom.

– IHT uses gradient descent followed by a hard thresholding that sets all but the K largest (in magnitude) elements in a vector to zero.

Generalized Orthogonal Matching Pursuit (gOMP)

Generalized Orthogonal Matching Pursuit (gOMP)

- The generalized orthogonal matching pursuit (gOMP)is a generalization of orthogonal matching pursuit (OMP).

Generalized Orthogonal Matching Pursuit (gOMP)

- The generalized orthogonal matching pursuit (gOMP)is a generalization of orthogonal matching pursuit (OMP).

- In contrast to the OMP algorithm, the gOMP algorithm reconstructs the K-sparse signal in K steps by selecting $N(> 1)$ atoms in each iteration.

Table: gOMP algorithm

Input: measurement $\mathbf{y} \in \mathbb{R}^m$, sensing matrix $\mathbf{\Phi}^{m \times n}$ **Initialization**: counter k=0, residue $\mathbf{r}^0 = \mathbf{y}$, estimated support set $\Lambda^k = \emptyset$
While k<K and $\|\|\mathbf{r}^k\|\|_2 > 0$ *Identification*: h^{k+1}=Set of indices corresponding to the N largest entries in $\|\mathbf{\Phi}^t \mathbf{r}^k\|$. $(NK \leq m)$ *Augment*: $\Lambda^{k+1} = \Lambda^k \cup \{h^{k+1}\}$ *Estimate*: $\mathbf{x}_{\Lambda^{k+1}} = \arg\min_{\mathbf{z}} \|\|y - \mathbf{\Phi}_{\Lambda^{k+1}} \mathbf{z}\|\|_2$ *Update*: $\mathbf{r}^{k+1} = y - \mathbf{\Phi}_{\Lambda^{k+1}} \mathbf{x}_{\Lambda^{k+1}}$ k=k+1 **End While** **Output**: $\mathbf{x} = \arg\min_{\mathbf{u}: supp(\mathbf{u}) = \Lambda^k} \|\|y - \mathbf{\Phi}\mathbf{u}\|\|_2$

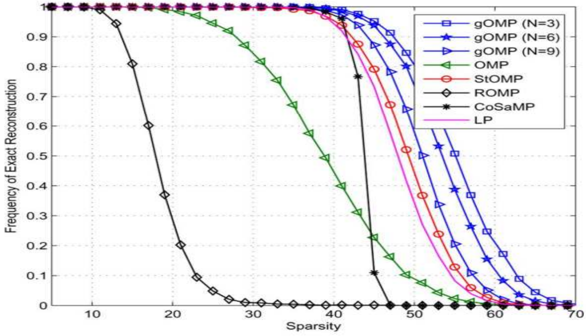

Figure: Reconstruction performance for K-sparse Gaussian signal vector using Gaussian dictionary(128×256) as a function of sparsity K

Courtesy: J. Wang, S. Kwon & B. Shim (2012) [1]

Analysis of the gOMP algorithm by J. Wang, S. Kwon & B. Shim (2012):

Analysis of the gOMP algorithm by J. Wang, S. Kwon & B. Shim (2012):

- In gOMP algorithm, convergence in a maximum of K steps is established by ensuring that in each iteration, at least one of the N new atoms chosen belongs to $\mathbf{\Phi}_T$, i.e., it has an index belonging to the true support set T.

Analysis of the gOMP algorithm by J. Wang, S. Kwon & B. Shim (2012):

- In gOMP algorithm, convergence in a maximum of K steps is established by ensuring that in each iteration, at least one of the N new atoms chosen belongs to $\mathbf{\Phi}_T$, i.e., it has an index belonging to the true support set T.
- Let $\beta^{k+1},\ k = 0, 1, \cdots, K-1$ denote the largest (in magnitude) correlation between $\mathbf{r}^k$ and the atoms of $\mathbf{\Phi}_T$ at the k-th step of iteration, i.e., $\beta^{k+1} = max\{|\phi_i^t \mathbf{r}^k| \, | i \in T,\ k = 0, 1, \cdots, K-1\}$.

Analysis of the gOMP algorithm by J. Wang, S. Kwon & B. Shim (2012):

- In gOMP algorithm, convergence in a maximum of K steps is established by ensuring that in each iteration, at least one of the N new atoms chosen belongs to $\mathbf{\Phi}_T$, i.e., it has an index belonging to the true support set T.
- Let $\beta^{k+1},\ k = 0, 1, \cdots, K-1$ denote the largest (in magnitude) correlation between $\mathbf{r}^k$ and the atoms of $\mathbf{\Phi}_T$ at the k-th step of iteration, i.e., $\beta^{k+1} = max\{|\phi_i^t \mathbf{r}^k|\ | i \in T,\ k = 0, 1, \cdots, K-1\}$.
- Let the N largest (in magnitude) correlations between $\mathbf{r}^k$ and the atoms of $\mathbf{\Phi}$ not belonging to $\mathbf{\Phi}_T$ be given by $\alpha_i^{k+1},\ i = 1, 2, \cdots, N$, arranged in descending order as $\alpha_1^{k+1} > \alpha_2^{k+1} ... > \alpha_N^{k+1}$.

Analysis of the gOMP algorithm by J. Wang, S. Kwon & B. Shim (2012):

Then, it is shown in [1] that

Analysis of the gOMP algorithm by J. Wang, S. Kwon & B. Shim (2012):

Then, it is shown in [1] that

- $$\alpha_N^{k+1} < \frac{1-3\delta_{NK}}{1-\delta_{NK}} \frac{||\mathbf{x}_{T-\Lambda^k}||_2}{\sqrt{N}} \tag{7}$$

Analysis of the gOMP algorithm by J. Wang, S. Kwon & B. Shim (2012):

Then, it is shown in [1] that

- $$\alpha_N^{k+1} < \frac{1-3\delta_{NK}}{1-\delta_{NK}} \frac{||\mathbf{x}_{T-\Lambda^k}||_2}{\sqrt{N}} \tag{7}$$

and

- $$\beta^{k+1} > \frac{\delta_{NK}}{1-\delta_{NK}} \frac{||\mathbf{x}_{T-\Lambda^k}||_2}{\sqrt{K-l}}, \tag{8}$$

  where $l = |T \cap \Lambda^k|$.

Analysis of the gOMP algorithm by J. Wang, S. Kwon & B. Shim (2012):

Then, it is shown in [1] that

- $$\alpha_N^{k+1} < \frac{1-3\delta_{NK}}{1-\delta_{NK}} \frac{||\mathbf{x}_{T-\Lambda^k}||_2}{\sqrt{N}} \tag{7}$$

and

- $$\beta^{k+1} > \frac{\delta_{NK}}{1-\delta_{NK}} \frac{||\mathbf{x}_{T-\Lambda^k}||_2}{\sqrt{K-l}}, \tag{8}$$

  where $l = |T \cap \Lambda^k|$.
- Therefore, the sufficient condition to ensure convergence in a maximum of K steps is then obtained by setting the RHS of (8) greater than that of (7) as $\delta_{NK} < \frac{\sqrt{N}}{\sqrt{K}+3\sqrt{N}}$ where $N(\geq 1)$

Our (i.e. with R. L. Das & S. Satpathi) contribution:

Our (i.e. with R. L. Das & S. Satpathi) contribution:

- We first retain the upper bound of α_N^{k+1} as given in (2), while the lower bound of β^{k+1} given in (3) is refined which eventually results in a lesser restrictive upper bound on δ_{NK} as shown in Theorem 1.

Our (i.e. with R. L. Das & S. Satpathi) contribution:

- We first retain the upper bound of α_N^{k+1} as given in (2), while the lower bound of β^{k+1} given in (3) is refined which eventually results in a lesser restrictive upper bound on δ_{NK} as shown in Theorem 1.
- Subsequently, we refine both the upper bound of α_N^{k+1} and the lower bound of β^{k+1}, which leads to an improved upper bound on δ_{NK+1} as shown Theorem 2.

The derivation uses the following Lemma, which is an extension of the Lemma 3.2 of [11].

Lemma 1

Given $\mathbf{u} \in \mathbb{R}^n$, $I_1,\ I_2 \subset Z$ *where* $I_2 = supp(\mathbf{u})$ *and* $I_1 \cap I_2 = \emptyset$

$$(1 - \frac{\delta_{|I_1|+|I_2|}}{1 - \delta_{|I_1|+|I_2|}})||\mathbf{u}||_2^2 \leq ||\mathbf{A}_{I_1}\mathbf{u}||_2^2 \leq (1 + \delta_{|I_1|+|I_2|})||\mathbf{u}||_2^2$$

Additionally, we use certain properties of the RIP constant, given by Lemma 1 in [7],[8] which are reproduced below.

Lemma 2

For $I,\ J \subset Z, I \cap J = \emptyset$, $\mathbf{q} \in \mathbb{R}^{|I|}$ *and* $\mathbf{p} \in \mathbb{R}^{|J|}$

(a) $\delta_{K_1} \leq \delta_{K_2}\ \forall\, K_1 < K_2$ *(monotonicity)*

(b) $(1-\delta_{|I|})||\mathbf{q}||_2 \leq ||\mathbf{\Phi}_I^t \mathbf{\Phi}_I \mathbf{q}||_2 \leq (1+\delta_{|I|})||\mathbf{q}||_2$

(c) $\langle \mathbf{\Phi}_I \mathbf{q}, \mathbf{\Phi}_J \mathbf{p} \rangle \leq \delta_{|I|+|J|}||\mathbf{p}||_2||\mathbf{q}||_2$ *with equality holding if either of* $\mathbf{p}$ *and* $\mathbf{q}$ *is zero. Also,* $||\mathbf{\Phi}_I^t \mathbf{\Phi}_J \mathbf{p}||_2 \leq \delta_{|I|+|J|}||\mathbf{p}||_2$ *with equality holding if* $\mathbf{p}$ *is zero.*

The Theorem 1:

Theorem 1

The gOMP can recover $\mathbf{x}$ *exactly when* $\mathbf{\Phi}$ *satisfies RIP of order* NK *with*

$$\delta_{NK} < \frac{\sqrt{N}}{\sqrt{K} + 2\sqrt{N}}$$

The Theorem 1:

Theorem 1

The gOMP can recover $\mathbf{x}$ *exactly when* $\mathbf{\Phi}$ *satisfies RIP of order* NK *with*

$$\delta_{NK} < \frac{\sqrt{N}}{\sqrt{K} + 2\sqrt{N}}$$

Proof.

We have shown in [12] that

$$\beta^{k+1} > \frac{1}{\sqrt{K}}\left(1 - \frac{\delta_{NK}}{1-\delta_{NK}}\right)||\mathbf{x}_{T-\Lambda^k}||_2 \tag{9}$$

Setting the RHS of (9) greater than the RHS of (7), the result follows trivially. □

For Complete Proof of Theorem 1

The Theorem 2:

Theorem 2

The gOMP algorithm can recover $\mathbf{x}$ *exactly when* $\mathbf{\Phi}$ *satisfies RIP of order* $NK + 1$ *with*

$$\delta_{NK+1} < \frac{\sqrt{N}}{\sqrt{K} + \sqrt{N}}$$

Proof.

Note that $\mathbf{r}^k = \mathbf{P}^{\perp}_{\Lambda^k}\mathbf{y} = \mathbf{P}^{\perp}_{\Lambda^k}\mathbf{\Phi}_{T-\Lambda^k}\mathbf{x}_{T-\Lambda^k}$. We can then write,

$$\mathbf{r}^k = \mathbf{\Phi}_{T\cup\Lambda^k}\mathbf{x}''_{T\cup\Lambda^k} \tag{10}$$

where

$$\mathbf{x}''_{T\cup\Lambda^k} = \begin{bmatrix} x_{T-\Lambda^k} \\ -z_{\Lambda^k} \end{bmatrix} \tag{11}$$

and $\mathbf{P}_{\Lambda^k}\mathbf{\Phi}_{T-\Lambda^k}\mathbf{x}_{T-\Lambda^k} = \mathbf{\Phi}_{\Lambda^k}\mathbf{z}_{\Lambda^k}$ for some $\mathbf{z}_{\Lambda^k} \in \mathbb{R}^{|\Lambda^k|} \equiv \mathbb{R}^{Nk}$. Then, it is shown in [12] that $\alpha_N^{k+1} < \frac{1}{\sqrt{N}}\delta_{NK+1}||\mathbf{x}''_{T\cup\Lambda^k}||_2$ and $\beta^{k+1} > \frac{1}{\sqrt{K}}(1-\delta_{NK})||\mathbf{x}''_{T\cup\Lambda^k}||_2$. Setting the RHS of (11) greater than that of (10), the result is obtained trivially. □

For Complete Proof of Theorem 2 · Go to references

First note that $\beta^{k+1} = ||\mathbf{\Phi}_T^t \mathbf{r}^k||_\infty,\ k = 0, 1, \cdots, K-1$ and that $\mathbf{r}^k = \mathbf{P}_{\Lambda^k}^\perp \mathbf{y}$ is orthogonal to each column of $\mathbf{\Phi}_{\Lambda^k}$, which also means that $\mathbf{P}_{\Lambda^k}^\perp \mathbf{y} = \mathbf{P}_{\Lambda^k}^\perp \mathbf{\Phi}_T \mathbf{x}_T$ $= \mathbf{P}_{\Lambda^k}^\perp (\mathbf{\Phi}_{T-\Lambda^k} \mathbf{x}_{T-\Lambda^k} + \mathbf{\Phi}_{T\cap\Lambda^k} \mathbf{x}_{T\cap\Lambda^k}) = \mathbf{P}_{\Lambda^k}^\perp \mathbf{\Phi}_{T-\Lambda^k} \mathbf{x}_{T-\Lambda^k}$. It is then possible to write,

$$\begin{aligned}
\beta^{k+1} &= ||\mathbf{\Phi}_T^t \mathbf{r}^k||_\infty > \frac{1}{\sqrt{K}} ||\mathbf{\Phi}_T^t \mathbf{r}^k||_2 \text{ (as } |T| = K) \\
&= \frac{1}{\sqrt{K}} ||\mathbf{\Phi}_{T-\Lambda^k}^t \mathbf{r}^k||_2 = \frac{1}{\sqrt{K}} ||\mathbf{\Phi}_{T-\Lambda^k}^t \mathbf{P}_{\Lambda^k}^\perp \mathbf{y}||_2 \\
&= \frac{1}{\sqrt{K}} ||\mathbf{\Phi}_{T-\Lambda^k}^t (\mathbf{P}_{\Lambda^k}^\perp)^t \mathbf{P}_{\Lambda^k}^\perp \mathbf{y}||_2 \text{ (as } \mathbf{P} = \mathbf{P}^t \& \mathbf{P} = \mathbf{P}^2) \\
&= \frac{1}{\sqrt{K}} ||(\mathbf{P}_{\Lambda^k}^\perp \mathbf{\Phi}_{T-\Lambda^k})^t \mathbf{P}_{\Lambda^k}^\perp \mathbf{\Phi}_T \mathbf{x}_T||_2 \\
&= \frac{1}{\sqrt{K}} ||(\mathbf{P}_{\Lambda^k}^\perp \mathbf{\Phi}_{T-\Lambda^k})^t \mathbf{P}_{\Lambda^k}^\perp \mathbf{\Phi}_{T-\Lambda^k} \mathbf{x}_{T-\Lambda^k}||_2 \qquad (12)
\end{aligned}$$

Proof of Theorem 1 in detail I

Next we define a vector $\mathbf{x}'$, where $x'_i = x_i$ if $i \in T - \Lambda^k$ and $x'_i = 0$ otherwise. It is easy to see that $\mathbf{\Phi x}' = \mathbf{\Phi}_{T-\Lambda^k}\mathbf{x}_{T-\Lambda^k}$ and thus,

$$\mathbf{A}_{\Lambda^k}\mathbf{x}' = \mathbf{P}^{\perp}_{\Lambda^k}\mathbf{\Phi x}' = \mathbf{P}^{\perp}_{\Lambda^k}\mathbf{\Phi}_{T-\Lambda^k}\mathbf{x}_{T-\Lambda^k}. \tag{13}$$

[It is also easy to observe that $\mathbf{P}^{\perp}_{\Lambda^k}\mathbf{\Phi x}' = \mathbf{P}^{\perp}_{\Lambda^k}\mathbf{\Phi}_T\mathbf{x}_T = \mathbf{r}^k$, since $\mathbf{P}^{\perp}_{\Lambda^k}\phi_i = \mathbf{0}$ for $i \in \Lambda^k$.] We are now in a position to apply Lemma 1 on $\mathbf{A}_{\Lambda^k}\mathbf{x}'$, taking $I_1 = \Lambda^k$ and $I_2 = supp(\mathbf{x}') = T - \Lambda^k$ and noting that $I_1 \cap I_2 = \emptyset, |I_1| + |I_2| = Nk + K - l$, to obtain

$$\begin{aligned} ||\mathbf{A}_{\Lambda^k}\mathbf{x}'||_2^2 &\geq (1 - \frac{\delta_{Nk+K-l}}{1-\delta_{Nk+K-l}})||\mathbf{x}'||_2^2 \\ &\overset{L1a}{>} (1 - \frac{\delta_{NK}}{1-\delta_{NK}})||\mathbf{x}_{T-\Lambda^k}||_2^2, \end{aligned} \tag{14}$$

where $Nk + K - l < NK$ follows from the fact that $k \leq l$ and $k < K$. Moreover,

$$\begin{aligned}
||\mathbf{A}_{\Lambda^k}\mathbf{x}'||_2^2 &\overset{(13)}{=} ||\mathbf{P}_{\Lambda^k}^{\perp}\mathbf{\Phi}_{T-\Lambda^k}\mathbf{x}_{T-\Lambda^k}||_2^2 \\
&= \langle \mathbf{P}_{\Lambda^k}^{\perp}\mathbf{\Phi}_{T-\Lambda^k}\mathbf{x}_{T-\Lambda^k}, \mathbf{P}_{\Lambda^k}^{\perp}\mathbf{\Phi}_{T-\Lambda^k}\mathbf{x}_{T-\Lambda^k} \rangle \\
&= \langle (\mathbf{P}_{\Lambda^k}^{\perp}\mathbf{\Phi}_{T-\Lambda^k})^t P_{\Lambda^k}^{\perp}\mathbf{\Phi}_{T-\Lambda^k}\mathbf{x}_{T-\Lambda^k}, \mathbf{x}_{T-\Lambda^k} \rangle \\
&\leq ||(\mathbf{P}_{\Lambda^k}^{\perp}\mathbf{\Phi}_{T-\Lambda^k})^t P_{\Lambda^k}^{\perp}\mathbf{\Phi}_{T-\Lambda^k}\mathbf{x}_{T-\Lambda^k}||_2 ||\mathbf{x}_{T-\Lambda^k}||_2 \qquad (15)
\end{aligned}$$

Combining (14) and (15) we get $||(\mathbf{P}_{\Lambda^k}^{\perp}\mathbf{\Phi}_{T-\Lambda^k})^t$ $P_{\Lambda^k}^{\perp}\mathbf{\Phi}_{T-\Lambda^k}\mathbf{x}_{T-\Lambda^k}||_2 > (1 - \frac{\delta_{NK}}{1-\delta_{NK}})||\mathbf{x}_{T-\Lambda^k}||_2$.
From (12), it then follows that

$$\beta^{k+1} > \frac{1}{\sqrt{K}}(1 - \frac{\delta_{NK}}{1-\delta_{NK}})||\mathbf{x}_{T-\Lambda^k}||_2 \qquad (16)$$

Setting the RHS of (16) greater than the RHS of (7), the result follows trivially. Return to Theorem 2

Proof of Theorem 2 in detail

First, as seen earlier, $\mathbf{r}^k = \mathbf{P}^{\perp}_{\Lambda^k}\mathbf{y} = \mathbf{P}^{\perp}_{\Lambda^k}\mathbf{\Phi}_{T-\Lambda^k}\mathbf{x}_{T-\Lambda^k}$. We can then write,

$$\begin{aligned}
\mathbf{r}^k &= \mathbf{\Phi}_{T-\Lambda^k}\mathbf{x}_{T-\Lambda^k} - \mathbf{P}_{\Lambda^k}\mathbf{\Phi}_{T-\Lambda^k}\mathbf{x}_{T-\Lambda^k} \\
&= \mathbf{\Phi}_{T-\Lambda^k}\mathbf{x}_{T-\Lambda^k} - \mathbf{\Phi}_{\Lambda^k}\mathbf{z}_{\Lambda^k} \\
&= \mathbf{\Phi}_{T\cup\Lambda^k}\mathbf{x}''_{T\cup\Lambda^k} \qquad (17)
\end{aligned}$$

where we use the fact that $\mathbf{P}_{\Lambda^k}\mathbf{\Phi}_{T-\Lambda^k}\mathbf{x}_{T-\Lambda^k}$ belongs to the $span(\mathbf{\Phi}_{\Lambda^k})$ and thus can be written as $\mathbf{\Phi}_{\Lambda^k}\mathbf{z}_{\Lambda^k}$ for some $\mathbf{z}_{\Lambda^k} \in \mathbb{R}^{|\Lambda^k|} \equiv \mathbb{R}^{Nk}$. The vector $\mathbf{x}''_{T\cup\Lambda^k}$ is then given as,

$$\mathbf{x}''_{T\cup\Lambda^k} = \begin{bmatrix} x_{T-\Lambda^k} \\ -z_{\Lambda^k} \end{bmatrix} \qquad (18)$$

Let W be the set of N incorrect indices corresponding to α_i^{k+1}'s for $i = 1, 2, \cdots, N$ (clearly, $W \subset (T \cup \Lambda^k)^c$ and $|W| = N$). So,

$$\begin{aligned}
\alpha_N^{k+1} &= min(|\langle \Phi_i, \mathbf{r}^k \rangle| \; |i \in W) \\
&\leqslant \frac{||\mathbf{\Phi}_W^t \mathbf{r}^k||_2}{\sqrt{N}} \quad (\text{ as } |W| = N) \\
&\overset{(17)}{=} \frac{1}{\sqrt{N}} ||\mathbf{\Phi}_W^t \mathbf{\Phi}_{T\cup\Lambda^k} \mathbf{x}''_{T\cup\Lambda^k}||_2 \\
&\overset{L2c}{\leqslant} \frac{1}{\sqrt{N}} \delta_{N+Nk+K-l} ||\mathbf{x}''_{T\cup\Lambda^k}||_2 \\
&\overset{L2a}{<} \frac{1}{\sqrt{N}} \delta_{NK+1} ||\mathbf{x}''_{T\cup\Lambda^k}||_2 \qquad (19)
\end{aligned}$$

where $N + Nk + K - l < NK + 1$ follows from the fact that $l \geq k$ and $k \leq K - 1$.

Similarly,

$$\beta^{k+1} = ||\mathbf{\Phi}_T^t \mathbf{r}^k||_\infty \geq \frac{1}{\sqrt{K}} ||\mathbf{\Phi}_T^t \mathbf{r}^k||_2 \quad (\text{ as } |T| = K)$$

$$= \frac{1}{\sqrt{K}} || \begin{bmatrix} \mathbf{\Phi}_T & \mathbf{\Phi}_{\Lambda^k - T} \end{bmatrix}^t \mathbf{r}^k ||_2 \tag{20}$$

$$= \frac{1}{\sqrt{K}} ||\mathbf{\Phi}_{T \cup \Lambda^k}^t \mathbf{\Phi}_{T \cup \Lambda^k} \mathbf{x}''_{T \cup \Lambda^k}||_2$$

$$\stackrel{L2b}{>} \frac{1}{\sqrt{K}} (1 - \delta_{Nk+K-l}) ||\mathbf{x}''_{T \cup \Lambda^k}||_2$$

$$\stackrel{L12a}{>} \frac{1}{\sqrt{K}} (1 - \delta_{NK}) ||\mathbf{x}''_{T \cup \Lambda^k}||_2. \tag{21}$$

Setting the RHS of (21) greater than that of (19), the result is obtained trivially.

References

References I

1. J. Wang, S. Kwon, and B. Shim, "Generalized orthogonal matching pursuit," *IEEE Trans. Signal Processing*, vol. 60, no. 12, pp. 6202-6216, Dec. 2012.
2. D.L. Donoho, "Compressed sensing",*IEEE Trans. Information Theory*, vol. 52, no. 4, pp. 1289-1306, Apr., 2006.
3. R. Baraniuk, "Compressive sensing",*IEEE Signal Processing Magazine,* vol. 25, pp. 21-30, March, 2007.
4. J.A. Tropp, and A.C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit", *IEEE Trans. Information Theory*, vol. 53, no. 12, pp. 4655-4666, Dec., 2007.

5. J. A. Tropp and S. J. Wright, "Computational methods for sparse solution of linear inverse problems," *Proc. IEEE*, vol. 98, no. 6, pp. 948-958, June, 2010.
6. S. Chen, S. A. Billings, and W . Luo, "Orthogonal least squares methods and their application to non-linear system identification", *Int. J. Contr.*, vol. 50, no. 5, pp. 1873-1896, 1989.
7. D. Needell and J. Tropp, "CoSaMP : Iterative Signal Recovery from Incomplete and Inaccurate Samples", *Appl. Comput. Harmon. Anal.*, vol. 26, pp. 301-321, 2009.
8. W. Dai and O. Milenkovic, "Subspace Pursuit for Compressive Sensing Signal Reconstruction", *IEEE Trans. Information Theory*, vol. 55, no. 5, pp. 2230-2249, 2009.

References III

9. M. Elad, *Sparse and Redundant Representations*, Springer, 2010.
10. J. Wang and B. Shim, “On the recovery limit of sparse signals using orthogonal matching pursuit,” *IEEE Trans. Signal Processing*, vol. 60, no. 9, pp. 4973-4976, sept. 2012.
11. M. A. Davenport and M. B. Wakin, “ Analysis of Orthogonal Matching Pursuit Using the Restricted Isometry Property”, *IEEE Trans. Information Theory*, vol. 56, no. 9, pp. 4395-4401, Sept., 2010.
12. S. Satpathi, R. L. Das and M. Chakraborty, “Improved Bounds on RIP for Generalized Orthogonal Matching Pursuit”, *arXiv:1302.0490*, 2013.

Thanks!