Introduction to Compressed Sensing

Mrityunjoy Chakraborty

Dept. of Electronics & Electrical Communication Engg.

Indian Institute of Technology, Kharagpur, INDIA

Introduction

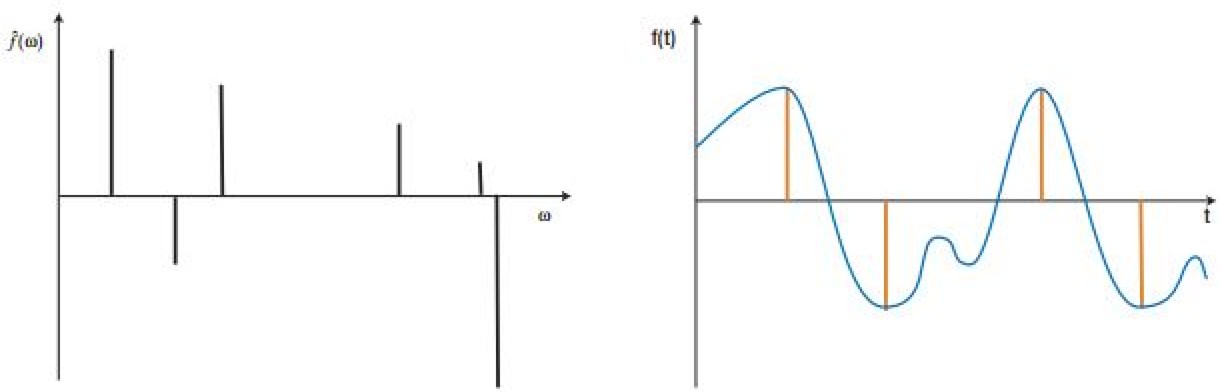
The field of Compressive Sensing(CS)

- A powerful method of exactly recovering signals at sub-Nyquist rate given that the signal has some sparse structure.

- It is a wide field with overlaps in several distinct areas of science and technology:
  - ∗ Signal Processing:
    - (i) MRI imaging
    - (ii) Speech processing
  - ∗ Applied mathematics:
    - (i) Applied harmonic analysis
    - (ii) Random matrix theory
    - (iii) Geometric functional analysis
  - ∗ Statistics

Data Acquisition

- For a signal bandlimited to B Hz, The Nquist rate demands at least $2B$ samples per second for perfect reconstruction.
    - Becomes pretty challenging for ADCs to deliver the high sampling rate in context of modern high bandwidth communication systems(e.g. radar).



- Can prior knowledge about sparse structure of the signal help perfect reconstruction from a sub – Nyquist sampling strategy?

Data Compression

- Many signals are sparse in transform domains, like Fourier, Wavelet etc.

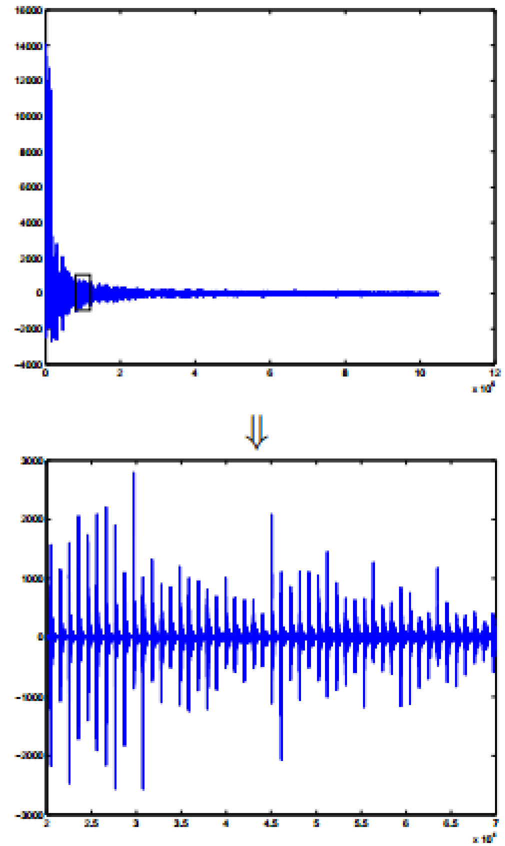
- Can we use the sparse structure in the transform domains to get compression even without the full acquisition (all signal coordinates)?

- Specifically, instead of taking samples of the actual vector $\mathbf{x} \in \mathbb{R}^N$, can we recover $\mathbf{x}$, from the linear measurements $\mathbf{y} = \mathbf{\Phi x}$, where, $\mathbf{x}$ is known to be sparse in some domain, that is there is some $\left(\text{known}\right)$ matrix $\mathbf{\Psi}$, such that $\mathbf{x} = \mathbf{\Psi z}$, such that $\mathbf{z}$ is sparse.

Sparsity in Wavelet domain

1 megapixel image

wavelet coefficients



zoom in

The problem of compressive sensing and its solutions

The l_0 "norm" optimization formulation

- A suitable optimization problem must be formulated that addresses these questions by seeking out an unknown vector which is *highly sparse*, i.e. with as few nonzero coordinates as possible.
- Mathematically, let the system of linear measurements be given by $\mathbf{y} = \mathbf{\Phi x}$, where $\mathbf{x} \in \mathbb{R}^N$, $\mathbf{\Phi} \in \mathbb{R}^{M \times N}$, $M << N$ and $\mathbf{x}$ is highly sparse.
- Then the optimization problem we seek to solve is solving a constrained l_0 "norm" minimization

$$\min_{\mathbf{x}:\mathbf{y}=\mathbf{\Phi x}} \|\mathbf{x}\|_0$$

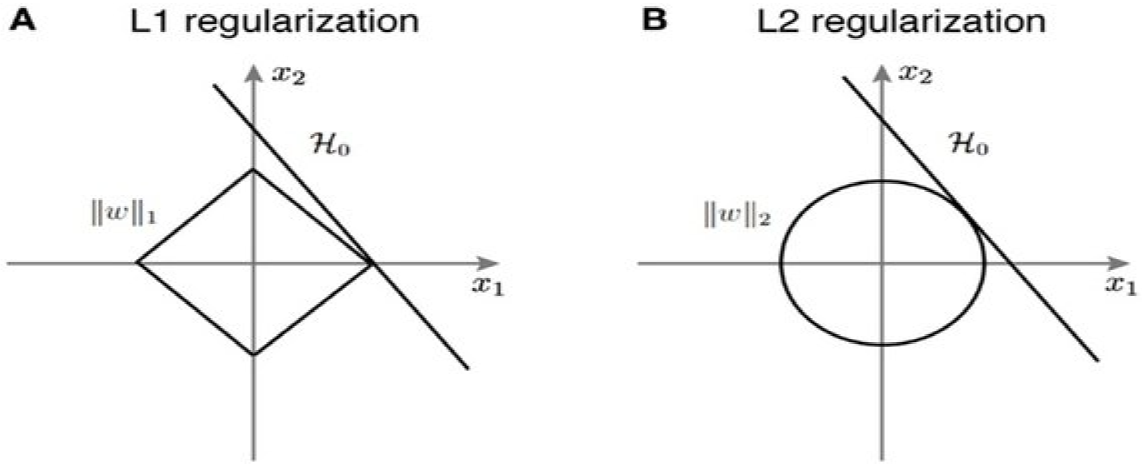
- However, this problem is a combinatorial one, and the complexity may be exponential. So, do we quit?

A relaxed convex optimization problem

- Turns out a slight convexification of the problem does the job, i.e. instead of minimizing l_0 "norm", minimize the l_1 norm

$$\min_{\mathbf{x}:\mathbf{y}=\Phi\mathbf{x}} \|\mathbf{x}\|_1$$

- The following diagram gives intuitive explanation to why l_1 optimization finds a sparse solution, while l_2 optimization does not



- Solution of this problem can recover unknown $\mathbf{x}$ with high probability, if $\mathbf{x}$ is $K-$sparse, and M satisfies,

$$M \geq cK \ln(N / K) << N$$

The Restricted Isometry Property

- In order to recover a high dimensional sparse vector $\mathbf{x}$, from a low dimensional measurement vector $\mathbf{y}$, obtained as $\mathbf{y} = \mathbf{\Phi x}$, the sensing matrix Φ must be "almost" orthonormal.
- This idea is captured by the *Restricted Isometry Property* (RIP):

    - A matrix $\mathbf{\Phi} \in \mathbb{R}^{M \times N}$ is said to satisfy RIP of order K, if $\forall$ $K-$sparse vector $\mathbf{x} \in \mathbb{R}^N$, $\exists \delta > 0$ such that

$$(1-\delta) \| \mathbf{x} \|_2^2 \le \| \mathbf{\Phi x} \|_2^2 \le (1+\delta) \| \mathbf{x} \|_2^2$$

    - The smallest such constant δ is denoted as δ_K

- In simple words, $\mathbf{\Phi}$ is an approximate isometry for all $K-$sparse vectors.

- RIP is fundamentally related to eigenvalues of a matrix

$$\delta_K = \max_{S \subset \{1,2,\cdots,N\}:|S|=K} \| \mathbf{I} - \mathbf{\Phi}_S^T \mathbf{\Phi}_S \|_{2\to 2}$$

- We can prove this as below
  - Let $\mathbf{x}$ be a $K-$sparse vector so that $\mathbf{\Phi x} = \mathbf{\Phi}_S \mathbf{x}_S,$ where S is the support of $\mathbf{x}$
  - Then, by definition of maximum and minimum eigenvalues of a matrix, $\lambda_{\min}(\Phi_S^T \Phi_S) = \arg\min_{\mathbf{u}} \dfrac{\mathbf{u}^T \mathbf{\Phi}_S^T \mathbf{\Phi}_S \mathbf{u}}{\| \mathbf{u} \|_2}, \lambda_{\max}(\Phi_S^T \Phi_S) = \arg\max_{\mathbf{u}} \dfrac{\mathbf{u}^T \mathbf{\Phi}_S^T \mathbf{\Phi}_S \mathbf{u}}{\| \mathbf{u} \|_2}$
  - Hence, from the definition of RIP, any $\delta,$ that satisfies the RIP of order $K,$ also satisfies, $\lambda_{\min}(\mathbf{\Phi}_S^T \mathbf{\Phi}_S) \geq (1-\delta), (1+\delta) \geq \lambda_{\max}(\mathbf{\Phi}_S^T \Phi_S) \Rightarrow \ \delta \geq \| \mathbf{I} - \mathbf{\Phi}_S^T \mathbf{\Phi}_S \|_{2\to 2}$

- Since this is true for any set S of indices with cardinality K, we can write, for any such δ satisfying the RIP property of order K, $\max_{S:|S|=K} \| \mathbf{I} - \mathbf{\Phi}_S^T \mathbf{\Phi}_S \|_{2\to 2} \leq \delta$
- Since, by definition, δ_K is the smallest such δ, we have $\delta_K = \max_{S:|S|=K} \| \mathbf{I} - \mathbf{\Phi}_S^T \mathbf{\Phi}_S \|_{2\to 2}$

- Another nice property of RIP is that if RIC is small, after transformation, orthogonal vectors remain *almost* orthogonal, as stated in the following form:

$$|\langle \mathbf{x}, \mathbf{y}\rangle| \le \delta_{|S_1|+|S_2|} \|\mathbf{x}\|_2 \|\mathbf{y}\|_2$$

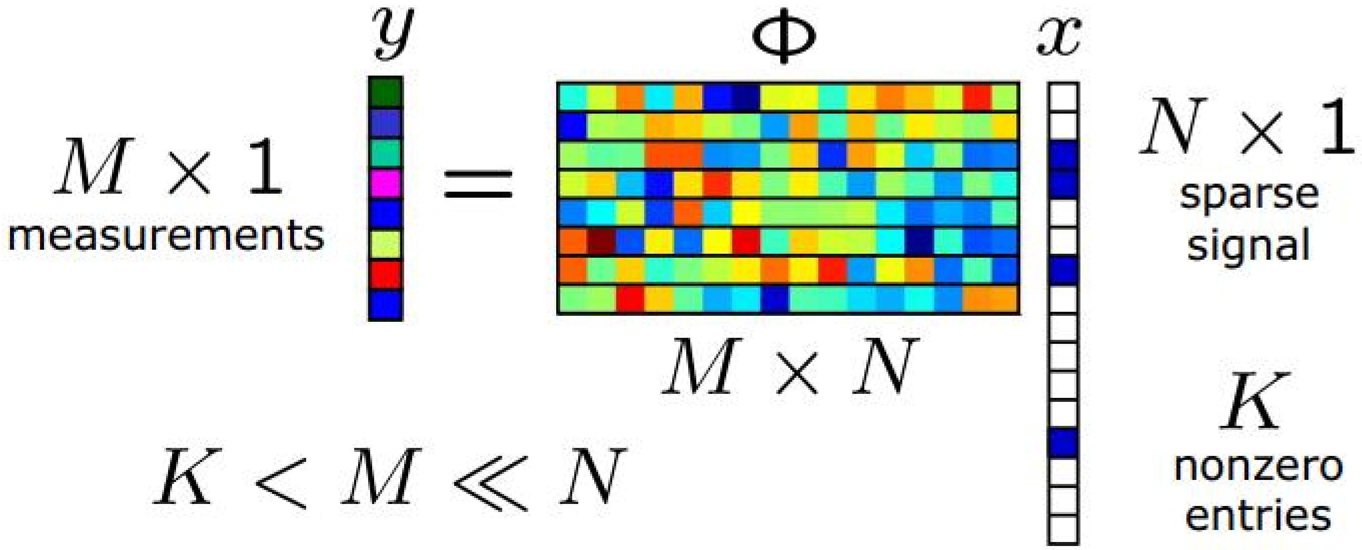
where $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$, with supports S_1, S_2 such that $S_1 \cap S_2 = \varnothing$

- To prove this, note that, $|\langle \mathbf{x}, \mathbf{y}\rangle| = 0$, since $S_1 \cap S_2 = \varnothing$, which allows us to write

$$\begin{aligned}
|\langle \mathbf{x}, \mathbf{y}\rangle| &= |\mathbf{x}^T(\mathbf{I} - \mathbf{\Phi}^T\mathbf{\Phi})\mathbf{y}| \\
&= |\mathbf{x}^T_{S_1\cup S_2}(\mathbf{I}_{S_1\cup S_2} - \mathbf{\Phi}^T_{S_1\cup S_2}\mathbf{\Phi}_{S_1\cup S_2})\mathbf{y}_{S_1\cup S_2}| \\
&\le \|\mathbf{x}_{S_1\cup S_2}\|_2 \|(\mathbf{I}_{S_1\cup S_2} - \mathbf{\Phi}^T_{S_1\cup S_2}\mathbf{\Phi}_{S_1\cup S_2})\mathbf{y}_{S_1\cup S_2}\|_2 \\
&\quad \left(\text{due to Cauchy} - \text{Scwartz}\right) \\
&\le \|\mathbf{x}_{S_1\cup S_2}\|_2 \|\mathbf{I}_{S_1\cup S_2} - \mathbf{\Phi}^T_{S_1\cup S_2}\mathbf{\Phi}_{S_1\cup S_2}\|_{2\to 2} \|\mathbf{y}_{S_1\cup S_2}\|_2 \\
&\le \delta_{|S_1|+|S_2|} \|\mathbf{x}\|_2 \|\mathbf{y}\|_2
\end{aligned}$$

How to find good sensing matrix ?

- A unique minimizer of the l_0 minimization problem is guaranteed if every $2K$ columns of the sensing matrix is linearly independent, equivalently, $\delta_{2K} \in (0,1)$; but how to find it ?
- How to design a sensing matrix such that $\delta_{2K} \in (0,1)$?



- An easy answer is *random matrices*, i.e., matrices with elements independent and identically distributed according to some distribution

- Fantastic examples are :
  - Gaussian sensing matrices, i.e., elements are i.i.d. Gaussian
  - Bernoulli sensing matrices with elements i.i.d. 0,1 with probabilities $1-p,\ p$

Recovery algorithms

l_1 minimization algorithms

- Basis pursuit:

$$\min_{\mathbf{x}\in\mathbb{R}^N} \|\mathbf{x}\|_1$$
$$\text{s.t.} \quad \mathbf{y} = \mathbf{\Phi}\mathbf{x}$$

- Quadratically constrained basis pursuit or Basis pursuit denoising (BPDN):

$$\min_{\mathbf{x}\in\mathbb{R}^N} \|\mathbf{x}\|_1$$
$$\text{s.t.} \quad \|\mathbf{y} - \mathbf{\Phi}\mathbf{x}\|_2 \leq \epsilon$$

- Dantzig Selector:

$$\min_{\mathbf{x}\in\mathbb{R}^N} \|\mathbf{x}\|_1$$
$$\text{s.t.} \quad \|\mathbf{\Phi}^T(\mathbf{y} - \mathbf{\Phi}\mathbf{x})\|_\infty \leq \tau$$

Greedy algorithms

Some of the most important greedy algorithms for sparse recovery are:

- Matching pursuit (MP), Orthogonal matching pursuit (OMP), Orthogonal least squares (OLS)
- Compressive sampling matching pursuit (CoSaMP), Subspace Pursuit (SP)
- Iterative Hard Thresholding (IHT), Hard Thresholding Pursuit (HTP)

Matching Pursuit

- Given that $\mathbf{y} \in \mathbb{R}^M$, the goal is to, iteratively, find the best linear representation of $\mathbf{y}$ in the dictionary $\{\phi_1, \cdots, \phi_N\}$
  - In other words, find $\mathbf{x}$, iteratively, such that $\|\mathbf{y} - \mathbf{\Phi x}\|_2$ is minimum where $\mathbf{\Phi} = [\phi_1 \cdots \phi_N]$

- In matching pursuit
  - Initialize the residual $\mathbf{r} = \mathbf{y}$
  - Find the atom most correlated to the residual, i.e., find ϕ_i such that $i = \arg\max_{1 \le j \le N} |\rho_j|$, where $\rho_j = \langle \phi_\mathbf{j}, \mathbf{r} \rangle$
  - Update the residual: $\mathbf{r} \leftarrow \mathbf{r} - \rho_i \phi_i$, and return to step 1

Orthogonal matching pursuit

- Same as matching pursuit, except that the dictionary representation is known to be $K-$sparse
- The Orthogonal matching pursuit goes as below :
  - Initialize residual $\mathbf{r}=\mathbf{y}$, and the temporary support $\Lambda=\varnothing$
  - Find the atom most correlated to the residual $\mathbf{r}$, i.e., find ϕ_i such that $i=\arg\max_{1\le j\le N}|\rho_j|$, where $\rho_j=\langle\phi_{\mathbf{j}},\mathbf{r}\rangle$
  - Enlarge the temporary support by augmenting this new index, i.e. $\Lambda\leftarrow\Lambda\cup\{i\}$
  - Find the best $K-$sparse representation of $\mathbf{y}$ with the atoms from the dictionary supported on Λ, i.e. find $\mathbf{x}$, such that $\mathbf{x}_\Lambda=\mathbf{\Phi}_\Lambda^\dagger\mathbf{y}$, $\mathbf{x}_{(\Lambda)^c}=\mathbf{0}$
  - Update residual $\mathbf{r}\leftarrow\mathbf{y}-\mathbf{\Phi}_\Lambda\mathbf{x}$ and return to step 1

- Many types of conditions have been found for the sensing matrix $\mathbf{\Phi}$, to ensure perfect recovery of the $K-$sparse vector $\mathbf{x}$ from the measurement $\mathbf{y} = \mathbf{\Phi x}$ vector in K iterations

- RIP based recovery conditions:
    - Davenport and Wakin [1] found the condition $\delta_{K+1} < \frac{1}{3\sqrt{K}}$
    - Wang et.al [2] improved the condition to $\delta_{K+1} < \frac{1}{\sqrt{K}+1}$
    - To date the best condition is established by chang et.al [3], which is $\delta_{K+1} < \frac{\sqrt{4K+1}-1}{2K}$

Another type of recovery conditions are given by the *worst* – *case coherence* μ, and the average – case coherence ν

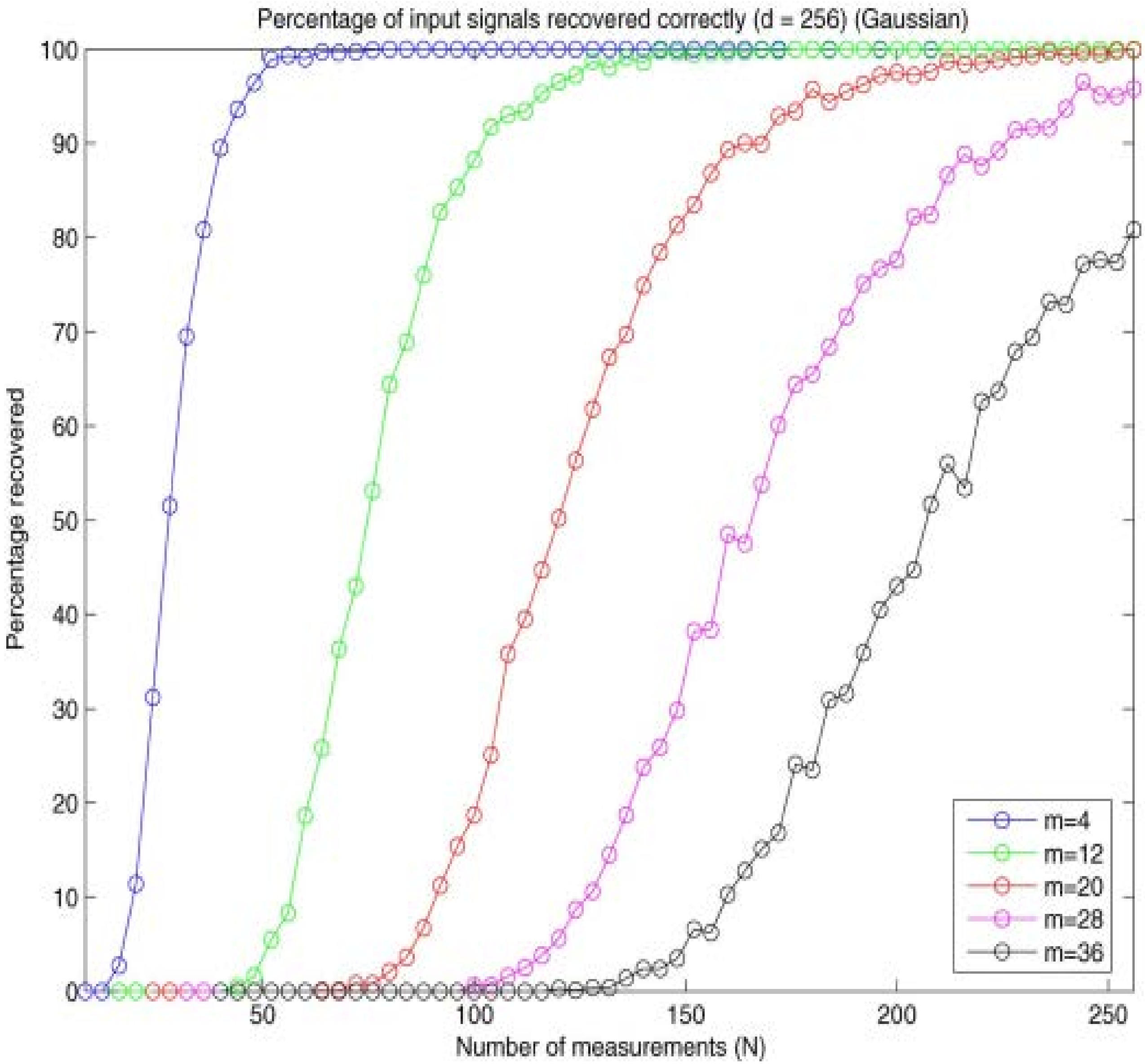
- Worst case coherence is defined as the maximum absolute cross – correlation among the columns of the sensing matrix, in other words $\mu := \max_{i \neq j} |\langle \phi_i, \phi_j \rangle|$
- Average case coherence is defined as the maximum among all the absolute values of row averages (excluding the diagonal in that row) of the Gram matrix $\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}$, in other words,

$$\nu := \frac{1}{N-1} \max_i \left| \sum_{j:i \neq j} \langle \phi_i, \phi_j \rangle \right|$$

- Tropp. [4] gives recovery condition in terms of μ as $\mu < \frac{1}{2K-1}$
- Chi and Calderbank. [5] give conditions $\mu < \frac{1}{240 \log N}$, and

$$\nu < \frac{\mu}{\sqrt{M}}, \text{with } N \geq 128$$

- Tropp and Gilbert. [6] have shown that OMP can indeed recover a $K-$sparse vector with very high probability if an "uncorrelated" (that is the mutual correlation between the columns of the matrix is very low with high probability) sensing matrices are used : specifically, if $\delta \in (0, 0.36)$, and if an "admissible" sensing matrix $\mathbf{\Phi}$ is chosen with dimension $M \times N$, with $M \geq C\ln(N\delta)$, for some constant C, then, OMP can recover the original, $K-$sparse vector $\mathbf{x}$ from the measurements $\mathbf{y} = \mathbf{\Phi x}$, with probability exceeding $1-\delta$


Percentage of input signals recovered correctly (d = 256) (Gaussian)
Percentage recovered
Number of measurements (N)
0
10
20
30
40
50
60
70
80
90
100
50
100
150
200
250
m=4
m=12
m=20
m=28
m=36

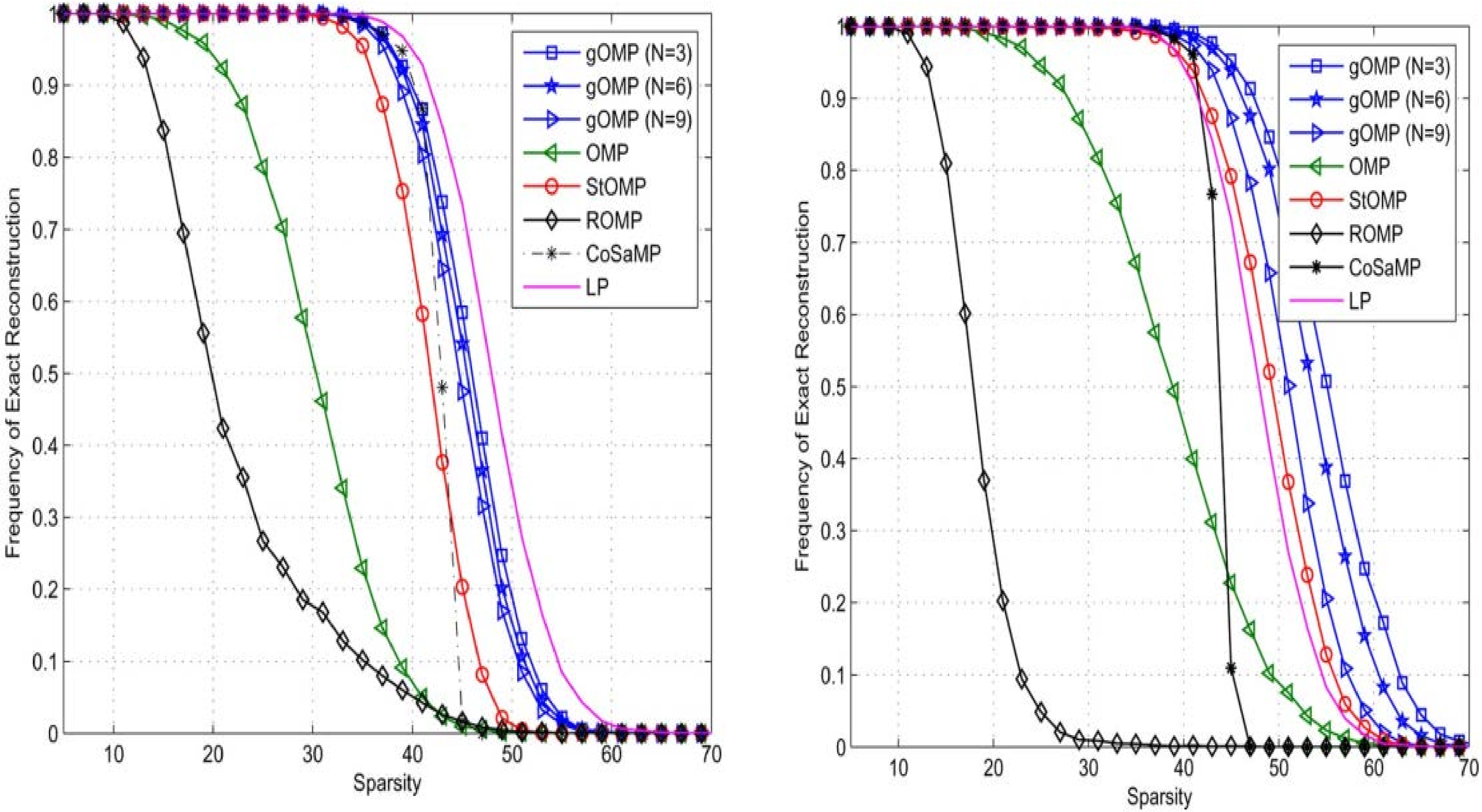
OMP with more than K iterations

- Recently a variant of OMP has been studied where OMP is run for more than K iterations, where K is the sparsity of the unknown vector
- Allowing the algorithm to run for more iterations improve the recovery condition
  - Recovery conditions found by Zhang: OMP can recover a K sparse vector with $30K$ iterations if $\delta_{31K} < \frac{1}{3}$
  - Recovery conditions found by Livshitz: OMP reconstructs a K sparse signals in $\lfloor \alpha\sqrt{K} \rfloor$, if $\delta_{\alpha\sqrt{K}} = \frac{\beta}{\sqrt{K}}$ for proper choices of $\alpha, \beta (\alpha \sim 2 \cdot 10^6, \beta \sim 10^{-6})$

- Sahoo and Makur $\left[7\right]$ has shown that if OMP is allowed to run for $K+\lfloor \alpha K \rfloor$ iterations $(\alpha \in [0,1])$, the algorithm can recover a K sparse vector with high probability with only $\mathcal{O}\left(K \ln \frac{n}{\lfloor \alpha K \rfloor + 1} \right)$ measurements, pretty close to the number of measurements required for success for Basis pursuit, that is $\mathcal{O}\left(K \ln \frac{n}{K} \right)$

Generalized orthogonal matching pursuit

- Wang. et.al [8] proposed a generalized orthogonal matching pursuit algorithm (gOMP) where at the augmentation step, instead of augmenting one index, $N (N \geq 1)$ indices are added, which are chosen according to decreasing order of absolute correlation with the residual vector.
- Recovery conditions for this algorithm are given as
  - $\delta_{KN} < \frac{\sqrt{N}}{\sqrt{K} + 3\sqrt{N}}$ [8]
  - $\delta_{KN} < \frac{\sqrt{N}}{\sqrt{K} + 2\sqrt{N}}, \ \delta_{NK+1} < \frac{\sqrt{N}}{\sqrt{K} + \sqrt{N}}$ [9]



Courtesy of Wang et.al. [8].

Orthogonal least squares

- OLS has the same functional structure as OMP
- The key difference is in the identification step :
  - Recall that OMP searches for a new index by finding the largest among the absolute correlations $|\langle \phi_{\mathbf{i}}, \mathbf{r}^k \rangle|$
  - OLS searches for an index such that inclusion of the corresponding column will minimize the projection error, i.e., find the index i such that $\|\mathrm{P}^{\perp}_{T^k \cup \{i\}} \mathbf{r}^k\|^2$ is minimized where i is searched over all indices in $\{1, 2, \cdots, n\} \setminus T^k$

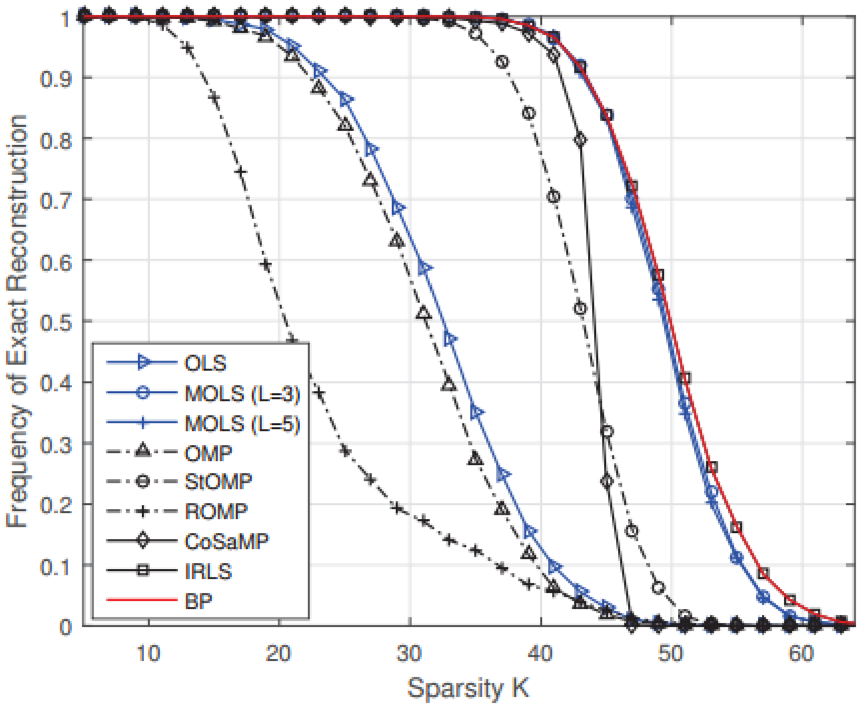
- There seems to be not much work on OLS in the literature.
- Soussen et.al [10] has numerically shown that OLS has uniformly higher recovery probability compared to OMP.
- Mukhopadhyay et.al [11] has tried to characterize the recovery performance of OLS in terms of recovery probability and explained why OLS has higher recovery probability, compared to OMP, in correlated dictionaries.

Multiple Orthogonal least squares

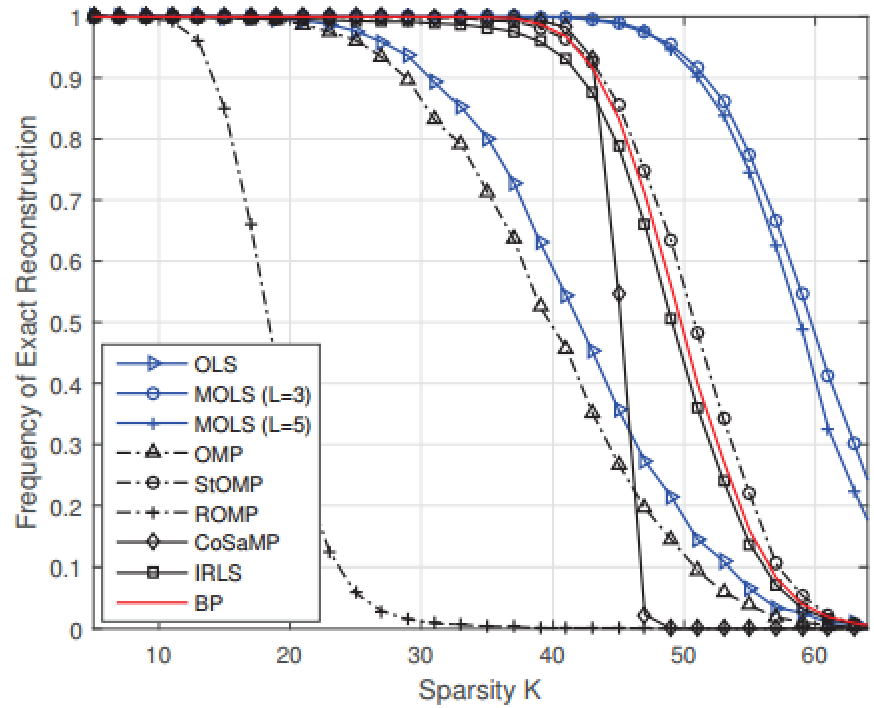
- MOLS is a generalization of OLS, proposed by Wang et.al [12].
- The generalization is realized in identification step, where instead of choosing one new index, a set of L indices $(L \geq 1)$ is chosen such that the sum of projection errors by individually appending an atom from that set is minimized, i.e.

$$\sum_{i \in S} \| \mathrm{P}^{\perp}_{T^k \cup \{i\}} \mathbf{r}^k \|^2 \text{ is minimized.}$$

- A recovery condition has been found by them: $\delta_{LK} < \dfrac{\sqrt{L}}{\sqrt{K} + 2\sqrt{L}}$.



(b) Sparse 2-PAM signal



(a) Sparse Gaussian signal.

Courtesy of Wang et.al. [8].

Compressive sampling matching pursuit (CoSaMP)

Input: Measurement vector $\mathbf{y} \in \mathbb{R}^m$, sensing matrix $\mathbf{\Phi} \in \mathbb{R}^{m \times n}$,

sparsity level K, initial estimate $\mathbf{x}^0$, ϵ

Initialize: counter $k = 0$

While $(\|\mathbf{y} - \mathbf{\Phi}\mathbf{x}^k\|_2 > \epsilon)$

Identification: $h^{k+1} = supp\left(H_{2K}\left(\mathbf{\Phi}^T(\mathbf{y} - \mathbf{\Phi}\mathbf{x}^k)\right)\right)$

Augmentation: $U^{k+1} = S^k \cup h^{k+1}$ where $S^k = supp(\mathbf{x}^k)$

Estimation: $\mathbf{u}^{k+1} = \arg\min_{\mathbf{u}:\mathbf{u}\in\mathbb{R}^n,\, supp(\mathbf{u})\subset U^{k+1}} \|\mathbf{y} - \mathbf{\Phi}\mathbf{u}\|_2$

Update: $\mathbf{x}^{k+1} = H_{2K}(\mathbf{u}^{k+1})$

$k = k + 1$

End While

Output: $\hat{\mathbf{x}} = \mathbf{x}^{k-1}$

- Needell and Tropp. [13] proposed CoSaMP as a sparse signal recovery algorithm. They proved the following recovery condition :
    - Let the measurement model be given by $\mathbf{y} = \mathbf{\Phi x} + \mathbf{e}$ where $\mathbf{x}$ is $K-$sparse and where $\mathbf{e}$ is the measurement noise vector. Then, for each iteration $k \geq 0,$ the signal approximation $\mathbf{x}^{\boldsymbol{k}}$ satisfies:

$$\|\mathbf{x} - \mathbf{x}_2^{k+1} \leq 0.5 \| \mathbf{x} - \mathbf{x}^k \|_2 + 10 \| \mathbf{e} \|_2 \|,$$

$$\|\mathbf{x} - \mathbf{x}^k \|_2 \leq 2^{-k} \| \mathbf{x} \|_2 + 20 \| \mathbf{e} \|_2$$

- Foucart [14] later improved the recovery condition to $\|\mathbf{x} - \mathbf{x}^k \|_2 \leq \rho^k \| \mathbf{x} - \mathbf{x}^0 \|_2 + \tau \| \mathbf{e} \|_2$ where ρ and τ depend on δ_{4K}, and $\delta_{4K} < \sqrt{\dfrac{5}{4+\sqrt{73}}} \approx 0.3847.$

- Satpathi and Chakraborty [15] showed that the number of iterations for the convergence of the CoSaMP algorithm is $\lceil cK \rceil$, where $c = \frac{\log(4/\rho_{4K}^2)}{\log(1/\rho_{4K}^2)}$, where $\rho_{4K} = \sqrt{\frac{2\delta_{4K}^2(1+\delta_{4K}^2)}{1-\delta_{4K}^2}}$

Subspace Pursuit (SP)

Input : Measurement vector $\mathbf{y} \in \mathbb{R}^m$, sensing matrix $\mathbf{\Phi} \in \mathbb{R}^{m \times n}$,

sparsity level K, initial estimate $\mathbf{x}^0$, ϵ

Initialize : counter $k = 0$

While $(\|\mathbf{y}_2 - \mathbf{\Phi}\mathbf{x}^k\|_2 > \epsilon)$

$$Identification:\ h^{k+1} = supp\left(H_K\left(\mathbf{\Phi}^T(\mathbf{y} - \mathbf{\Phi}\mathbf{x}^k)\right)\right)$$

$$Augmentation:\ U^{k+1} = S^k \cup h^{k+1} \text{ where } S^k = supp(\mathbf{x}^k)$$

$$Estimation: \mathbf{u}^{k+1} = \arg\min_{\mathbf{u}:\mathbf{u}\in\mathbb{R}^n,\, supp(\mathbf{u})\subset U^k} \|\mathbf{y} - \mathbf{\Phi}\mathbf{u}\|_2$$

$$Update: S^{k+1} = supp(\mathbf{u}^{k+1})$$

$$\mathbf{x}^{k+1} = \arg\min_{\mathbf{u}:\mathbf{u}\in\mathbb{R}^n,\, supp(\mathbf{u})\subset S^{k+1}} \|\mathbf{y} - \mathbf{\Phi}\mathbf{u}\|_2$$

$$k = k + 1$$

End While

Output : $\hat{\mathbf{x}} = \mathbf{x}^{k-1}$

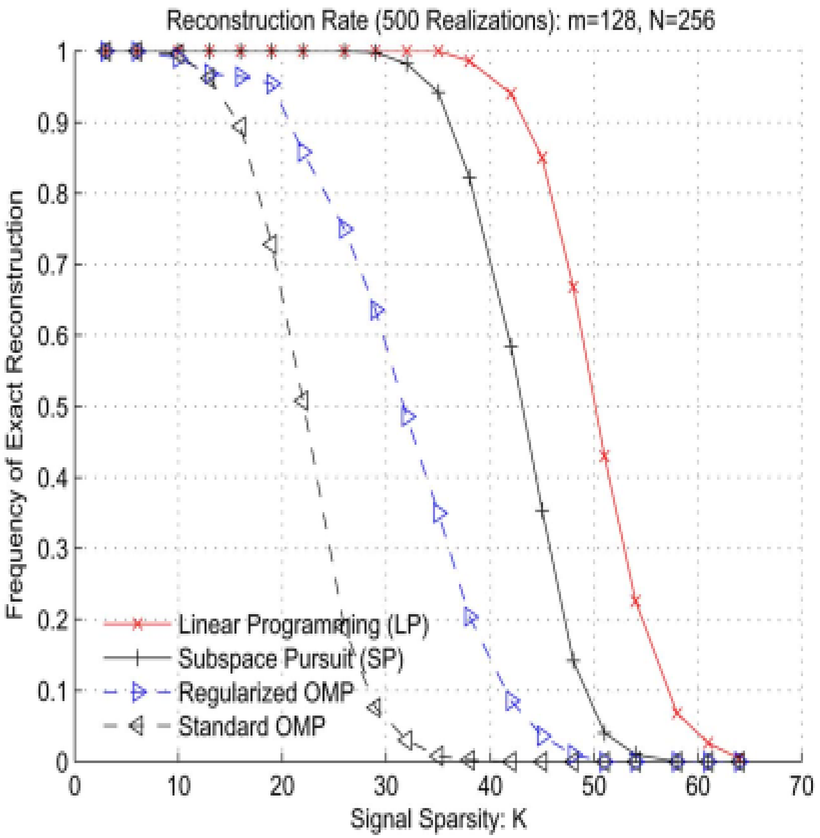
- Dai and Milenkovic [16] proposed SP almost at the same time Needel and Tropp proposed CoSaMP.
  - SP is quite similar to CoSaMP with the difference that SP has to compute two orthogonal projections, while CoSaMP requires to compute only one projection.
  - Dai and Milenkovic showed that if $\mathbf{x}$ is a K sparse unknown vector and $\mathbf{y} = \mathbf{\Phi x}$ is the measurement vector with the sensing matrix $\mathbf{\Phi}$ satisfying $\delta_{3K} < 0.165$, then, SP converges to the unknown vector $\mathbf{x}$ in a finite number of steps.
  - Dai and Milenkovic also found an upper bound for the number of iterations necessary for SP to converge as

$$n_{\text{it}} \le \min\left\{ \frac{\log \rho_{\min}}{\log c_K} + 1,\ \frac{125K}{-\log c_K} \right\}, \text{where}$$

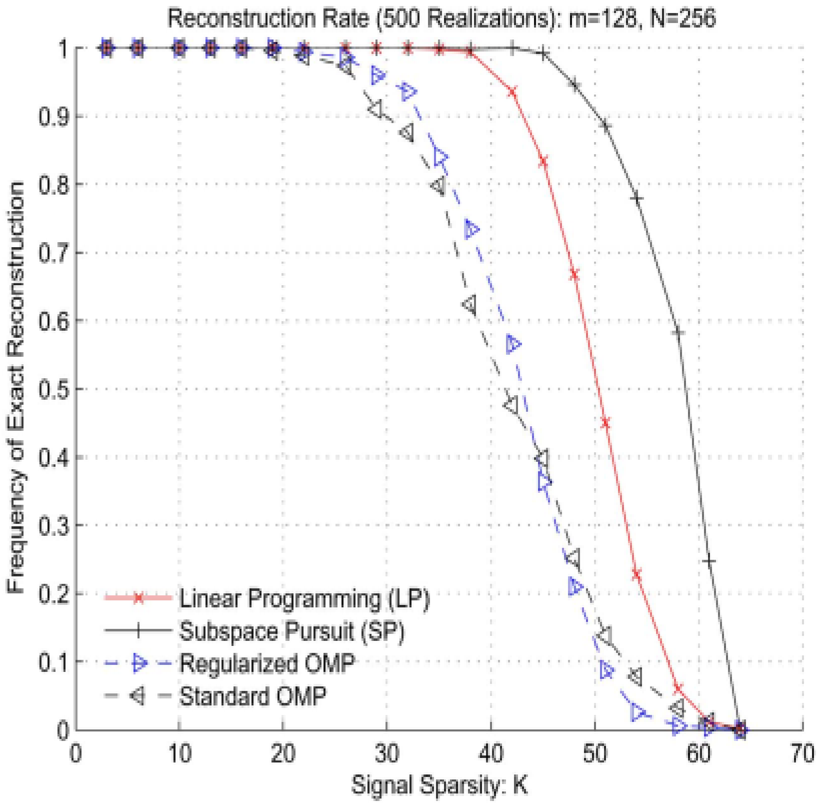
$$\rho_{\min} = \min_{1\le i\le n} |x_i| / \|\mathbf{x}\|, \text{ and } c_K := \frac{2\delta_{3K}(1+\delta_{3K})}{(1-\delta_{3K})^3}$$

- Satpathi and Chakraborty [15] later found the number of iterations for convergence to $\lceil ck \rceil$, where $c = \dfrac{\log(4/\rho_{3K}^2)}{\log(1/\rho_{4K}^2)}$,

where $\rho_{mK} := \sqrt{\dfrac{2\delta_{mK}^2(1+\delta_{mK}^2)}{1-\delta_{mK}^2}}$, $m \in \mathbb{Z}^+$.



(b) Simulations for zero-one sparse signals: both OMP and ROMP starts to fail when $K \geq 10$, ℓ_1-LP begins to fail when $K \geq 35$, and the SP algorithm fails when $K \geq 29$.



(a) Simulations for Gaussian sparse signals: OMP and ROMP start to fail when $K \geq 19$ and when $K \geq 22$ respectively, ℓ_1-LP begins to fail when $K \geq 35$, and the SP algorithm fails only when $K \geq 45$.

Courtesy Dai and Milenkovic [16]

Iterated Hard Thresholding (IHT)

Input : Measurement vector $\mathbf{y} \in \mathbb{R}^m$, sensing matrix $\boldsymbol{\Phi} \in \mathbb{R}^{m \times n}$,

sparsity level K, initial estimate $\mathbf{x}^0$, ϵ

Initialize : counter $k = 0$

While $(\|\mathbf{y} - \boldsymbol{\Phi}\mathbf{x}^k\|_2 > \epsilon)$

$$\mathbf{x}^{k+1} = H_K\left(\mathbf{x}^k + \boldsymbol{\Phi}^T(\mathbf{y} - \boldsymbol{\Phi}\mathbf{x}^k)\right)$$

$$k = k + 1$$

End While

Output : $\hat{\mathbf{x}} = \mathbf{x}^{k-1}$

- This algorithm is motivated by the constrained gradient descent approach:
  - The IHT algorithm solves the following problem:

$$\min_{\mathbf{x}:\|\mathbf{x}\|_0 \leq K} \|\mathbf{y} - \mathbf{\Phi}\mathbf{x}\|_2^2$$

  - The problem is non-convex in nature as the constraint set is non-convex.
  - However, an heuristic approach is to use gradient descent to first solve the unconstrained convex problem $\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{\Phi}\mathbf{x}\|_2^2$ by the following update at each step

$$\mathbf{x}^k + \mu\mathbf{\Phi}^T(\mathbf{y} - \mathbf{\Phi}\mathbf{x}^k)$$

    then restrict each update of the gradient descent to a K-sparse vector.

- This amounts to projecting the gradient descent update on the $union\ of\ the \binom{n}{K}\ subspaces\ containing\ K-sparse\ vectors$
- The resulting update becomes

$$\mathbf{x}^{k+1} = H_K(\mathbf{x}^k + \mu\mathbf{\Phi}^T(\mathbf{y} - \mathbf{\Phi}\mathbf{x}^k))$$

- The “heuristic” approach of deriving IHT has been formalized by Blumensath and Gilbert [17].
    - Instead of directly solving the actual constrained optimization problem, they attempt to solve another constrained optimization problem where the objective function is a majorization of the actual objective function, with the constraint set unchanged.
    - Specifically, they define the following functional:

$$C(\mathbf{x},\mathbf{z}) = \|\mathbf{y}-\mathbf{\Phi}\mathbf{x}\|_2^2 - \|\mathbf{\Phi}\mathbf{x}-\mathbf{\Phi}\mathbf{z}\|_2^2 + \|\mathbf{x}-\mathbf{z}\|_2^2$$

    - Note that $C(\mathbf{x},\mathbf{x})$ is the actual objective function, and under the condition $\|\mathbf{\Phi}\|_{2\to 2} < 1$, $C(\mathbf{x},\mathbf{x}) \le C(\mathbf{x},\mathbf{z}),\ \forall \mathbf{z}$.

- Thus the prescription for the minimization follows the so called "Maximization – Minimization" $\left(\text{MM}\right)$ approach, formally, $\mathbf{x}^{k+1} = \arg\min_{\mathbf{u}:\|\mathbf{u}\|_0 \le K} C(\mathbf{u}, \mathbf{x}^k), \forall k \ge 0.$
- The updates turn out to be the updates of IHT.
- Also note that, $C(\mathbf{x}^{k+1}, \mathbf{x}^{k+1}) \le C(\mathbf{x}^{k+1}, \mathbf{x}^k) \le C(\mathbf{x}^k, \mathbf{x}^k),$ where the first inequality follows from the majorization property, and the second inequality follows from the minimization property.

- A very simple convergence proof has been given by Foucart $\left[14\right]$
  - Assume the measurement model $\mathbf{y} = \mathbf{\Phi x},$ with the unknown $\mathbf{x}$ having known sparsity K
  - Then, from the k^{th} update of IHT, it follows that

$$\| \mathbf{x}^{k+1} - (\mathbf{x}^k + \mathbf{\Phi}^T(\mathbf{y} - \mathbf{\Phi x}^k)) \|_2^2 \;\le\; \| \mathbf{x} - (\mathbf{x}^k + \mathbf{\Phi}^T(\mathbf{y} - \mathbf{\Phi x}^k)) \|_2^2$$

$$\Rightarrow \| (\mathbf{x}^{k+1} - \mathbf{x}) \|_2^2 \;\le\; 2\langle \mathbf{x}^{k+1} - \mathbf{x}, \mathbf{x}^k + \mathbf{\Phi}^T(\mathbf{y} - \mathbf{\Phi x}^k) - \mathbf{x} \rangle$$

$$\Rightarrow \| (\mathbf{x}^{k+1} - \mathbf{x}) \|_2^2 \;\le\; 2\langle \mathbf{x}^{k+1} - \mathbf{x}, (\mathbf{I} - \mathbf{\Phi}^T\mathbf{\Phi})(\mathbf{x}^k - \mathbf{x}) \rangle$$

  - Let the supports of $\mathbf{x}, \mathbf{x}^k, \mathbf{x}^{k+1}$ be $\Lambda, \Lambda_k, \Lambda_{k+1}$ respectively, and let $V_{k+1} = \Lambda \cup \Lambda_k \cup \Lambda_{k+1},$ so that $|V_{k+1}| \le 3K.$
  - The, it follows from Cauchy – Scwartz inequality,

$$\| (\mathbf{x}^{k+1} - \mathbf{x}) \|_2^2 \le 2 \| \mathbf{I} - \mathbf{\Phi}_{V^{k+1}}^T \mathbf{\Phi}_{V^{k+1}} \|_{2\to 2} \| \mathbf{x}^k - \mathbf{x} \|_2 \| \mathbf{x}^{k+1} - \mathbf{x} \|_2$$

$$\Rightarrow \| (\mathbf{x}^{k+1} - \mathbf{x}) \|_2 \le 2\delta_{3K} \| \mathbf{x}^k - \mathbf{x} \|_2$$

which implies that $\mathbf{x}^k \to \mathbf{x},$ as $k \to \infty,$ if $\delta_{3K} < \dfrac{1}{2}$

Hard Thresholding Pursuit (HTP)

Input : Measurement vector $\mathbf{y} \in \mathbb{R}^m$, sensing matrix $\mathbf{\Phi} \in \mathbb{R}^{m\times n}$,
sparsity level K, initial estimate $\mathbf{x}^0$, ϵ

Initialize : counter $k = 0$

While $(\|\mathbf{y} - \mathbf{\Phi}\mathbf{x}^k\|_2 > \epsilon)$

$$L^{k+1} = supp\left(H_K\left(\mathbf{x}^k + \mathbf{\Phi}^T(\mathbf{y} - \mathbf{\Phi}\mathbf{x}^k)\right)\right)$$

$$x^{k+1} = \arg\min_{\mathbf{u}:\mathbf{u}\in\mathbb{R}^n,\, supp(\mathbf{u})\subset L^{k+1}} \|\mathbf{y} - \mathbf{\Phi}\mathbf{u}\|_2$$

$$k = k + 1$$

End While

Output : $\hat{\mathbf{x}} = \mathbf{x}^{k-1}$

- Foucart [18] proposed HTP motivated by the observation that the number of iterations taken by IHT to converge can be reduced by taking orthogonal projections of the updates on the set of K indices found at an iteration
- Foucart found the recovery condition for HTP for perfect measurements to be $\delta_{3K} < \dfrac{1}{\sqrt{3}}$
- Bouchot et.al [19] has found an upper bound on the number of iteration that HTP take to converge, as $n \le cK$, where c is a constant such that $c \le 3$, whenever, $\delta_{3K} \le 1/\sqrt{5}$

Hard Thresholding Pursuit (HTP)

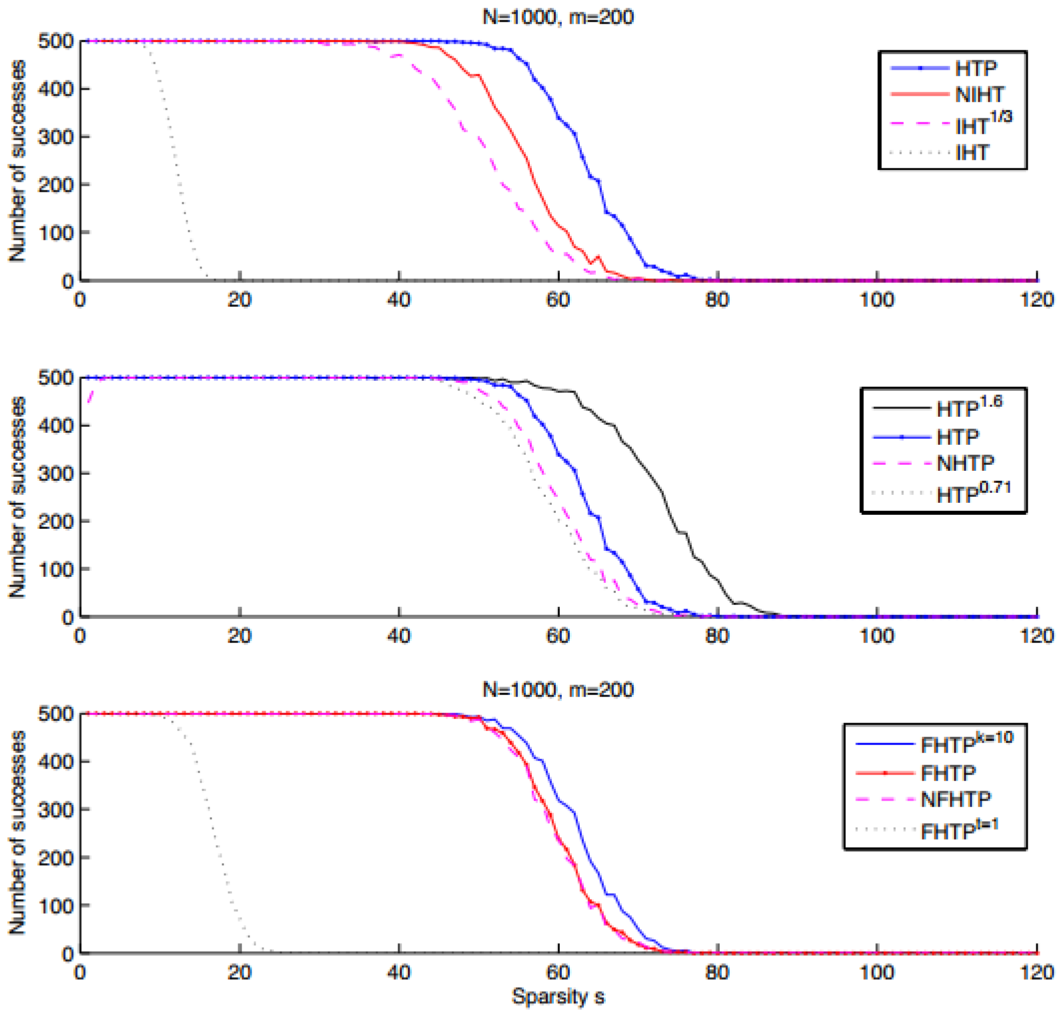


FIG. 4.2. *Number of successes for IHT and HTP algorithms (Gaussian matrices and vectors)*

Courtesy of Foucart [18].

Models of sparsity

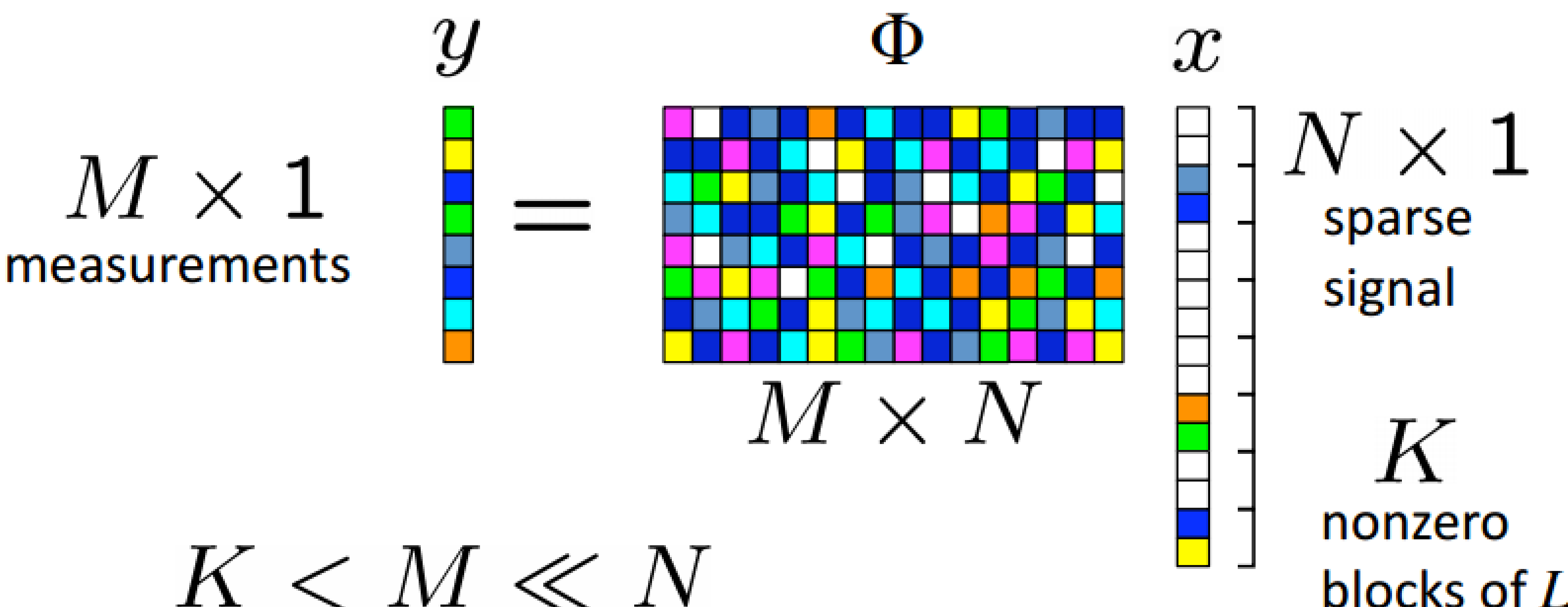
Block sparsity

- These are sparse vectors where the non – zero coefficients occur in clusters.
- Let a vector can be written as

$\mathbf{x} = [x_1 \cdots x_d \; x_{d+1} \cdots x_{2d} \cdots x_{N-d+1} \cdots x_N]^T$. Let $N = Ld$,

and let $\mathbf{x}[l] := [x_{(l-1)d+1} \cdots x_{ld}]^T$, so that each of these $\mathbf{x}[\cdot]$

represents a block of length d.

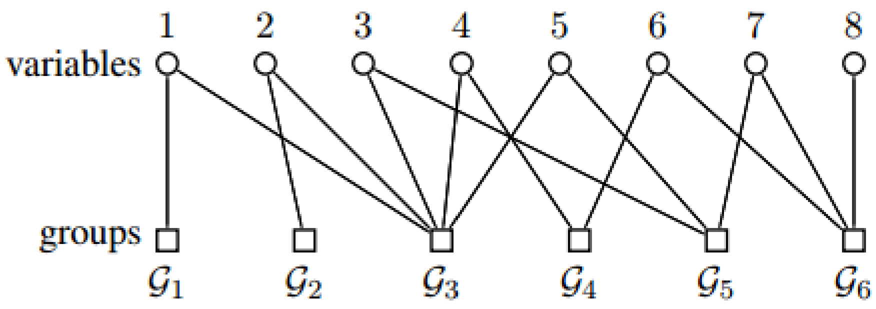


- The block sparsity has been defined [20] as.

$$\|\mathbf{x}\|_{2,0} = \sum_{l=1}^{L} I\left(\|\mathbf{x}[l]\|_2 > 0\right)$$

Group Sparsity

- A generalization of block sparsity, where the blocks may not be overlapping
- Consider a set of indices $\mathcal{N} = \{1, 2, \cdots, n\}$, and consider a class $\mathfrak{G}$, called a group structure, which is a collection of some subsets of $\mathcal{N}$, i.e, $\mathfrak{G} = \{\mathcal{G}_1, \cdots, \mathcal{G}_L\}$, such that $\mathcal{G}_i \subseteq \mathcal{N}$, $1 \le i \le L$, and $\bigcup_{\mathcal{G} \in \mathfrak{G}} \mathcal{G} = \mathcal{N}$
- A vector $\mathbf{x}$ is called a $G-group-sparse$ vector [21] with respect to the group structure $\mathfrak{G}$, if the support of $\mathbf{x}$ is contained in the union of at most G groups form the group structure $\mathfrak{G}$



Union of Subspace (UoS) model and Model sparse signals

- Another generalization of block sparse model that tries to capture the effect of overlapping blocks
- Let $\mathbf{x} \in \mathbb{R}^N$ be a $K-$sparse vector, but with unknown support, i.e. the support of $\mathbf{x}$ can be any of the $\binom{N}{K}$ supports of cardinality K, numbered as Λ_1 through $\Lambda_{\binom{N}{K}}$
- For each i, $1 \le i \le \binom{N}{K}$, define the sets

$$\mathcal{V}_i = \{\mathbf{u} \in \mathbb{R}^N \mid u_i = 0 \ \forall i \in \Lambda_i\}$$

- It is not difficult to see that each V_i is a subspace of dimension K, but are, in general, overlapping, that is, in general, $\mathcal{V}_i \cap \mathcal{V}_j \setminus \{\mathbf{0}\} \neq \varnothing$

- Thus, $\mathbf{x} \in \mathcal{U}$ where $\mathcal{U} = \bigcup_{i=1}^{\binom{N}{K}} \mathcal{V}_i$

- In general, let $\mathcal{M}_K = \cup_{i=1}^{m_K} \mathcal{V}_i$, where $1 \le m_K \le \begin{pmatrix} N \\ K \end{pmatrix}$ then $\mathcal{M}_K$ defines the $K-$model sparse signal model and the elements of $\mathcal{M}_K$ are called the $K-$model sparse signals.

References

[1] M. Davenport, M. B. Wakin *et al.*, "Analysis of orthogonal matching pursuit using the restricted isometry property," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4395–4401, 2010.

[2] J. Wang and B. Shim, "On the recovery limit of sparse signals using orthogonal matching pursuit," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4973–4976, 2012.

[3] L.-H. Chang and J.-Y. Wu, "An improved rip-based performance guarantee for sparse signal recovery via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 60, no. 9, pp. 5702–5715, 2014.

[4] J. Tropp*et al.*, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, 2004.

[5] Y. Chi and R. Calderbank, "Coherence-based performance guarantees of orthogonal matching pursuit," in *Communication, Control, and Computing (Allerton)*, 2012 50*th Annual Allerton Conference on*. IEEE, 2012, pp. 2003–2009.

[6] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[7] S. K. Sahoo and A. Makur, "Signal recovery from random measurements via extended orthogonal matching pursuit," *IEEE Trans. Signal Process.*, vol. 63, no. 10, pp. 2572–2581, 2015

[8] J. Wang, S. Kwon, and B. Shim, "Generalized orthogonal matching pursuit," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6202–6216, 2012.

[9] S. Satpathi, R. L. Das, and M. Chakraborty, "Improving the bound on the rip constant in generalized orthogonal matching pursuit," *IEEE Signal Process. Lett.*, vol. 20, no. 11, pp. 1074–1077, 2013.

[10] C. Soussen, R. Gribonval, J. Idier, and C. Herzet, "Joint k–step analysis of orthogonal matching pursuit and orthogonal least squares," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 3158–3174, 2013.

[11] S. Mukhopadhyay, P. Vashishtha, and M. Chakraborty, "Signal recovery in uncorrelated and correlated dictionaries using orthogonal least squares," *arXiv preprint arXiv*:1607.08712, 2016.

[12] J. Wang and P. Li, "Recovery of Sparse Signals Using Multiple Orthogonal Least Squares," *arXiv preprint arXiv*:1410.2505, Oct. 2014.

[13] D. Needell and J. A. Tropp, "Cosamp: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 301–321, 2009.

[14] S. Foucart, "Sparse recovery algorithms: sufficient conditions in terms of restricted isometry constants," in *Approximation Theory XIII: San Antonio* 2010. Springer, 2012, pp. 65–77.

[15] S. Satpathi and M. Chakraborty, "On the number of iterations for convergence of cosamp and sp algorithm," *arXiv preprint arXiv*:1404.4927, 2014.

[16] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2230–2249, 2009.

[17] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Appl. Comput. Harmon. Anal.*, vol. 27, no. 3, pp. 265–274, 2009.

[18] S. Foucart, "Hard thresholding pursuit: an algorithm for compressive sensing," *SIAM J. Numer. Anal.*, vol. 49, no. 6, pp. 2543–2563, 2011.

[19] J.–L. Bouchot, S. Foucart, and P. Hitczenko, "Hard thresholding pursuit algorithms: number of iterations," *Appl. Comput. Harmon. Anal.*, 2016.

[20] Y. C. Eldar and G. Kutyniok, *Compressed sensing: theory and applications*. Cambridge University Press, 2012.

[21] L. Baldassarre, N. Bhan, V. Cevher, A. Kyrillidis, and S. Satpathi, "Group–sparse model selection: Hardness and relaxations," *arXiv preprint arXiv*:1303.3207, 2013.