ELEG 867 - Compressive Sensing and Sparse Signal Representations

Gonzalo R. Arce
Depart. of Electrical and Computer Engineering
University of Delaware

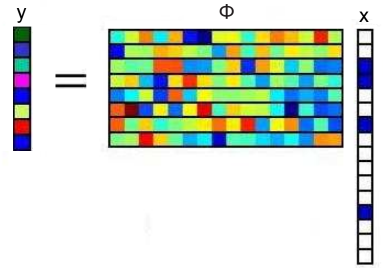
Fall 2011

Outline

- Compressive sensing signal reconstruction
- Geometry of CS
- Reconstruction Algorithms
    - Matching Pursuit
    - Orthogonal Matching Pursuit
    - Iterative Hard Thresholding
    - Weighted Median Regression

Compressive Sensing Signal Reconstruction

- Goal: Recover signal x from measurements y
- Problem: Random projection Φ not full rank (ill-posed inverse problem)
- Solution: Exploit the sparse/compressible geometry of acquired signal x



Suppose there is a S-sparse solution to $y = \Phi x$

- Combinatorial optimization problem

$$(P0) \quad \min_x \|x\|_0 \quad \text{s.t.} \quad \Phi x = y$$

- Convex optimization problem

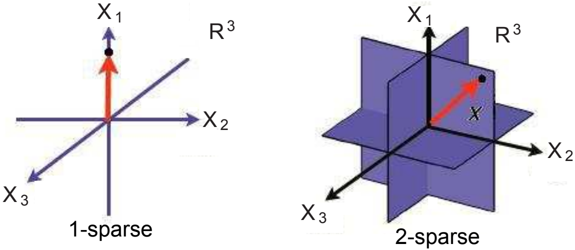
$$(P1) \quad \min_x \|x\|_1 \quad \text{s.t.} \quad \Phi x = y$$

- If $2\delta_{3s} + \delta_{2s} \leq 1$, the solutions $(P0)$ and $(P1)$ are the same[†].

[†] E. Candès."Compressive Sampling". Proc. Intern. Congress of Math., China, April 2006.

The Geometry of CS

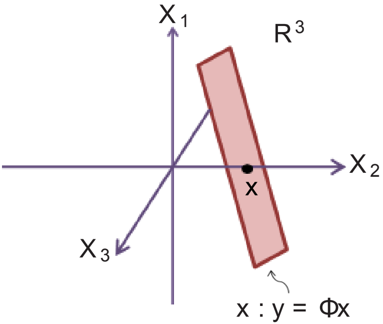
- Sparse signals have few non-zero coefficients



ℓ_0 Recovery

Reconstruction should be

- Consistent with the model: x as sparse as possible min $\|x\|_0$
- Consistent with the measurements: $y = \Phi x$



ℓ_0 Recovery

$$\min_x \|x\|_0 \text{ s.t. } \Phi x = y$$

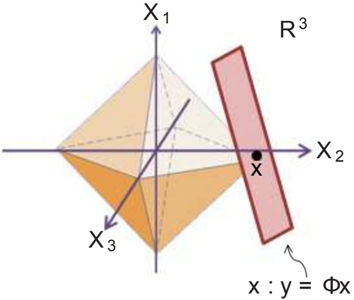
- $\|x\|_0$ number of nonzero elements
- Sparsest signal consistent with the measurements y
- Requires only $M << N$ measurements
- Combinatorial NP-hard problem: for $x \in \mathbb{R}^N$ with sparsity S, the complexity is $O(N^S)$
- Too slow for implementation

ℓ_1 Recovery

A more realistic model

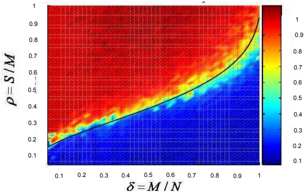
$$\min_x \|x\|_1 \quad \text{s.t.} \quad \Phi x = y$$

- The ℓ_1 norm also induces sparsity.
- The constraint is given by $y = \Phi x$.



Phase Transition Diagram

In the ℓ_1 minimization, there is a defined region on $(M/N, S/M)$ which ensures successful recovery

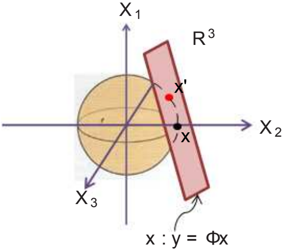


- $\delta = M/N$ is a normalized measure of the problem indeterminacy.
- $\rho = S/M$ is a normalized measure of the sparsity.
- Red region - unsuccessful recovery or exact reconstruction typically fails.
- Blue region - successful recovery or exact reconstruction typically occurs.

ℓ_2 Recovery

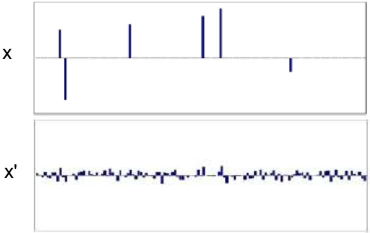
$$\min_x \|x\|_2 \quad \text{s.t.} \quad \Phi x = y$$

- Least square solution $x = (\Phi^T \Phi)^{-1} \Phi^T y$
- Solved by using quadratic programming:
  - ✓ Least squares solution
  - ✓ Interior-point methods
- Solution is almost never sparse



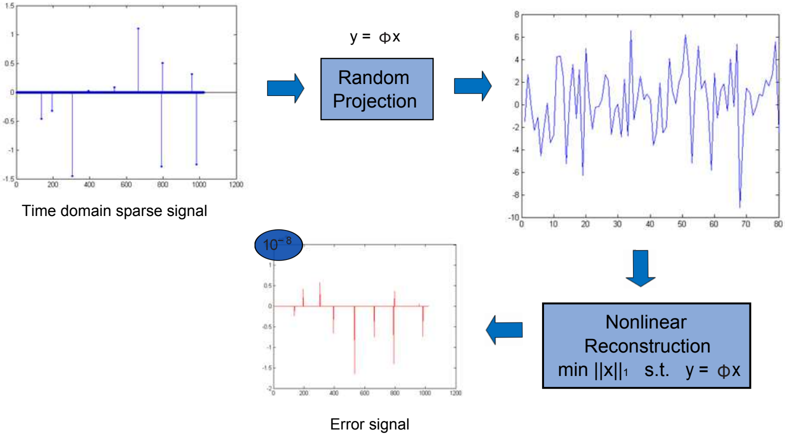
ℓ_2 Recovery

- Problem: small ℓ_2 does not imply sparsity



- x' has small ℓ_2 norm but it is not sparse

CS Example in the Time Domain



CS Example in the Wavelet Domain

- Reconstruction of an image ($N = 1$ megapixel) sparse ($S = 25,000$) in the wavelet domain from $M = 96,000$ incoherent measurements.[†]

Original (25K wavelets)

Recovered Image

[†] E. J. Candès and J. Romberg ”Sparsity and Incoherence in Compressive Sampling.” Inverse Problems. vol.23, pp.969-985. 2006.

Reconstruction Algorithms

- Different formulations and implementations have been proposed to find the sparsest x subject to $y = \Phi x$
- Difficult to compare results obtained by different methods
- Those are broadly classified in:
  - ✓ Regularization formulations (Replace combinatorial problem with convex optimization)
  - ✓ Greedy algorithms (Iterative refinement of a sparse solution)
  - ✓ Bayesian framework (Assume prior distribution of sparse coefficients)

Greedy Algorithms

Iterative algorithms that select an optimal subset of the desired signal x at each iteration. Some examples:

Matching Pursuit (MP)	$\min_x \|\|x\|\|_0 \quad \text{s.t.} \quad \Phi x = y$
Orthogonal Matching Pursuit(OMP)	$\min_x \|\|x\|\|_0 \quad \text{s.t.} \quad \Phi x = y$
Compressive Sampling Matching Pursuit (CoSaMP)	$\min_x \|\|y - \Phi x\|\|_2^2 \quad \text{s.t.} \quad \|\|x\|\|_0 \leq S$
Iterative Hard Thresholding (ITH)	$\min_x \|\|y - \Phi x\|\|_2^2 \quad \text{s.t.} \quad \|\|x\|\|_0 \leq S$
Weighted Median Regression	$\min_x \|\|y - \Phi x\|\|_1 + \lambda \|\|x\|\|_0$

Those methods build an approximation of the sparse signal at each iteration.

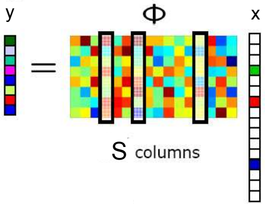
Matching Pursuit (MP)

Sparse approximation algorithm used to find the solution to

$$\min_{x} \|x\|_0 \ \text{ s.t. } \ \Phi x = y$$

Key ideas:

- Measurements $y = \Phi x$ are composed of sum of S columns of Φ.
- The algorithm needs to identify which S columns contribute to y.

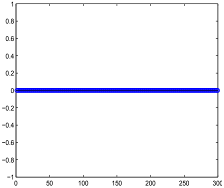


Matching Pursuit (MP) Algorithm

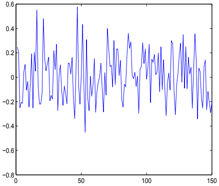
Input and Initialize Parameters	Observation vector $\mathbf{y}$ Measurement Matrix Φ Initial Solution $\widehat{x}^{(0)} = 0$ Initial residual $r^{(0)} = y$
Iteration k	Compute the current error: $c_j^{(k)} = \langle \phi_j, r^{(k)} \rangle$ Identify the index $\widehat{j}$ such that: $\widehat{j} = \max_j \lvert c_j^{(k)} \rvert$ Update $\widehat{x}_{\widehat{j}}^{(k)} = \widehat{x}_{\widehat{j}}^{(k-1)} + c_{\widehat{j}}^{(k)}$ Update $r^{(k)} = r^{(k-1)} - c_{\widehat{j}}^{(k)} \phi_{\widehat{j}}$

D. Donoho et. al, SparseLab Version 2.0. 2007.

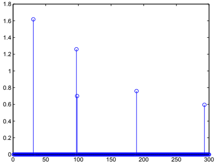
MP Reconstruction Example



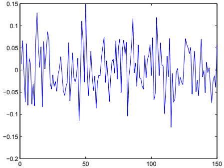
reconstruc. signal
Iteration 0



residue - Iteration 0

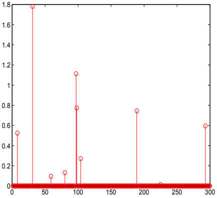


reconstruc. signal
Iteration 5

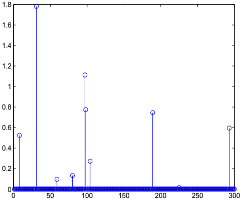


residue - Iteration 5

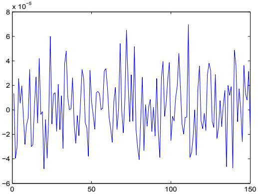
MP Reconstruction Example



Original Signal



reconstruc. signal
Iteration 32



Residue Iteration 32

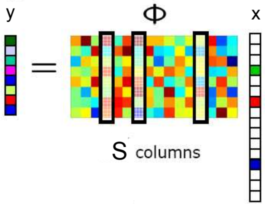
Orthogonal Matching Pursuit (MP)

Sparse approximation algorithm used to find the solution to

$$\min_x \|x\|_0 \;\; \text{s.t.} \;\; \Phi x = y$$

Key ideas:

- Measurements $y = \Phi x$ are composed of sum of S columns of Φ.
- The algorithm needs to identify which S columns contribute to y.



Orthogonal Matching Pursuit (OMP) Algorithm

Input and Initialize Parameters	Observation vector $\mathbf{y}$ Measurement Matrix Φ Initial Solution $\widehat{x}^{(0)} = 0$ Initial Solution Support $S^{(0)}$=support$\{\widehat{x}^{(0)}\} = 0$ Initial residual $r^{(0)} = y$
Iteration k	Compute the current error: $c_j^{(k)} = \langle \phi_j, r^{(k)} \rangle$ Identify the index $\widehat{j}$ such that: $\widehat{j} = \max_j \lvert c_j^{(k)} \rvert$ Update the support $S^{(k)}$=$S^{(k-1)} \cup \widehat{j}$ Update the matrix $\Phi_{S^{(k)}} = [\underline{0}, \ldots, \phi_i, \ldots, \underline{0}, \ldots, \phi_j, \ldots, \underline{0}]$ Update the solution $\widehat{x}^{(k)} = (\Phi_{S^{(k)}}^T \Phi_{S^{(k)}})^{-1} \Phi_{S^{(k)}}^T y$ Update $r^{(k)} = y - \Phi_{S^{(k)}} \widehat{x}^{(k)}$

D. Donoho et. al, SparseLab Version 2.0. 2007.

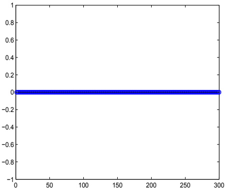
Remarks on the OMP Algorithm

- $S^{(k)}$ is the support of $\widehat{x}$ at the iteration k and $\Phi_{S^{(k)}}$ is the matrix that contains the columns from Φ that belongs to this support.
- The updated solution gives the $\widehat{x}^{(k)}$ that solves the minimization problem $\|\Phi_{S^{(k)}}x - y\|_2^2$. The solution is given by

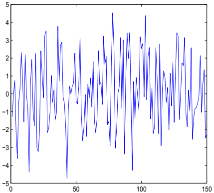
$$\begin{aligned} \Phi_{S^{(k)}}^T\left(\Phi_{S^{(k)}}x - y\right) &= 0 \\ -\Phi_{S^{(k)}}^T r^{(k)} &= 0 \end{aligned}$$

This solution shows that the columns in $\Phi_{S^{(k)}}^T$ are orthogonal to $r^{(k)}$.

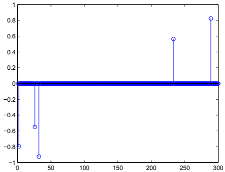
OMP Reconstruction Example



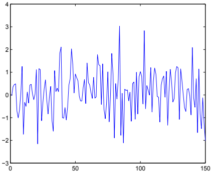
reconstruc. signal
Iteration 0



residue - Iteration 0

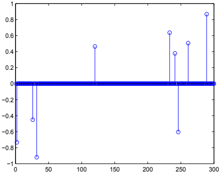


reconstruc. signal
Iteration 5

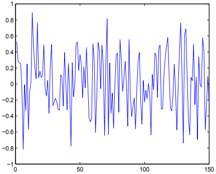


residue - Iteration 5

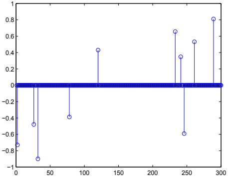
OMP Reconstruction Example



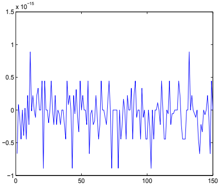
reconstruc. signal
Iteration 9



residue - Iteration 9



reconstruc. signal
Iteration 10



residue - Iteration 10

Iterative Hard Thresholding

- Sparse approximation algorithm that solves the problem

$$\min_x \|y - \Phi x\|_2^2 + \lambda \|x\|_0.$$

- Iterative algorithm that computes a descent direction given by the gradient of the ℓ_2-norm.
- The solution at each iteration is given by finding:

$$x^{(k+1)} = H_\lambda(x^{(k)} + \Phi^T(y - \Phi x^{(k)}))$$

- H_λ is a non-linear operator that only retains the coefficients with the largest magnitude.

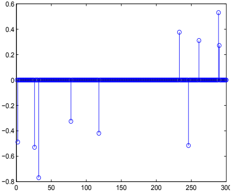
† T. Blumensath and M. Davies. "Iterative Hard Thresholding for Compressed Sensing." Journal of Appl. Comp. Harm. Anal. vol. 27, no. 3, pp. 265-274, 2009.

Iterative Hard Thresholding

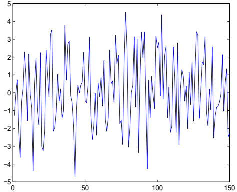
Input and Initialize Parameters	Observation vector **y** Measurement Matrix Φ $\widehat{x}^{(0)} = 0$
Iteration k	Compute an update: $a^{(k+1)} = x^{(k)} + \Phi^T(y - \Phi x^{(k)}))$ Select the largest elements: $x^{(k+1)} = H_\lambda(a^{(k+1)})$. Pull other elements toward zero.

D. Donoho et. al, SparseLab Version 2.0. 2007.

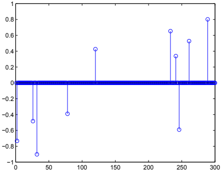
ITH Reconstruction Example



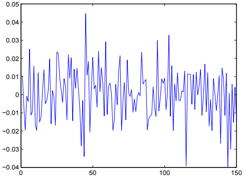
reconstruc. signal
Iteration 1



residue - Iteration 1

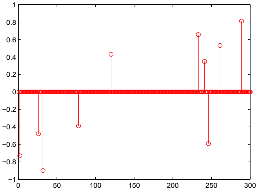


reconstruc. signal
Iteration 5

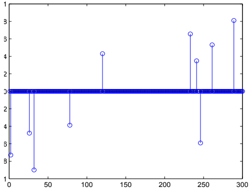


residue - Iteration 5

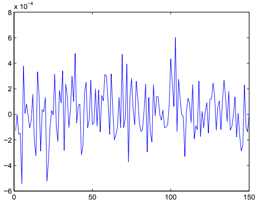
ITH Reconstruction Example



Original Signal



reconstruc. signal
Iteration 10



residue - Iteration 10

Weighted Median Regression Reconstruction Algorithm

When the random measurements are corrupted by noise the most widely used CS reconstruction algorithms solve the problem

$$\min_x \|y - \Phi x\|_2^2 + \tau \|x\|_1.$$

where τ is the regularization parameter that balances the conflicting task of minimizing the ℓ_2 norm while yielding a sparse solution.

- As τ increases the solution is sparser.
- As τ decreases the solution approximates to least square solution.

In the solution to the problem

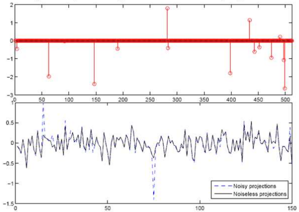
$$\min_{x} \|y - \Phi x\|_2^2 + \tau \|x\|_1 .$$

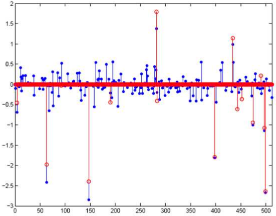
- The ℓ_2 term is the data fitting term, induced by the Gaussian assumption of the noise contamination.
- The ℓ_1 term is the sparsity promoting term.

However, the ℓ_2 term leads to the least square solution which is known to be very sensitive to high-noise level and outliers present in the contamination.

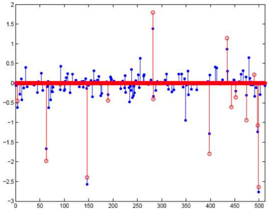
Example of signal reconstruction when outliers are present in the random projections

Original sparse signal and random projections





Matching Pursuit Solution



L1-Ls solution

To mitigate the effect of the impulsive noise in the compressive measurements, a more robust norm for the data-fitting term is

$$\min_{x} \underbrace{\|y - \Phi x\|_1}_{\text{Data fitting}} + \underbrace{\lambda \|x\|_0}_{\text{Sparsity}} \tag{1}$$

- LAD offers robustness to a broad class of noise, in particular to heavy tail noise.
- Optimum under the maximum likelihood principle when the underlying contamination follows a Laplacian-distributed model.

† J. L. Paredes and G. Arce. ”Compressive Sensing Signal Reconstruction by Weighted Median Regression Estimate.” ICASSP 2010.

Drawbacks:

- Solving ℓ_0-LAD optimization is N_p-hard problem.
- Direct solution is unfeasible even for modest-sized signal.

To solve this multivariable minimization problem:

- Solve a sequence of scalar minimization subproblems.
- Each subproblem improves the estimate of the solution by minimizing along a selected coordinate with all other coordinates fixed.
- Closed form solution exists for the scalar minimization problem.

- Signal reconstruction

$$\widehat{x} = \min_{x} \sum_{i=1}^{M} |(y - \Phi x)_i| + \lambda \|x\|_0$$

The solution using Coordinate Descent Method is given by

$$\widehat{x_n} = \min_{x_n} \sum_{i=1}^{M} |y_i - \sum_{j=1, j\neq n}^{N} \phi_{i,j} x_j - \phi_{i,n} x_n| + \lambda |x_n|_0 + \sum_{j=1, j\neq n}^{N} \lambda |x_j|_0$$

$$\widehat{x_n} = \min_{x_n} \underbrace{\sum_{i=1}^{M} |\phi_{in}| \left| \frac{y_i - \sum_{j=1, j\neq n}^{N} \phi_{i,j} x_j}{\phi_{in}} - x_n \right|}_{\text{Weighted least absolute deviation}} + \underbrace{\lambda |x_n|_0}_{\text{Regulariz.}}$$

$Q(x_n)$

- The solution to the minimization problem is given by

$$\widehat{x_n} = \min_{x_n} Q(x_n) + \lambda |x_n|_0$$

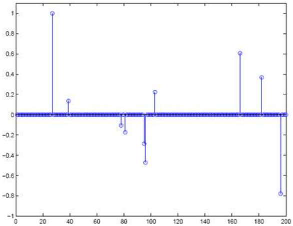
$$\tilde{x}_n = \text{MEDIAN}(|\phi_{in}| \diamond \frac{y_i - \sum_{j=1, j\neq n}^{N} \phi_{i,j} x_j}{\phi_{in}}) |_{i=1}^{M};\ n = 1, 2, ..., N$$

- At k^{th} iteration:

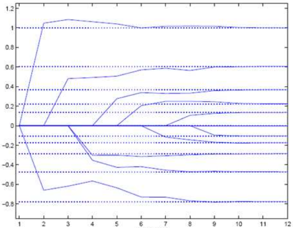
$$\widehat{x}_n^{(k)} = \begin{cases} \tilde{x}_n, & \text{if } \ Q(\tilde{x}_n) + \lambda \leq Q(0) \\ 0, & \text{otherwise} \end{cases}$$

The iterative algorithm detects the non-zero entries of the sparse vector by

- *Robust signal estimation:* computing a rough estimate of the signal that minimizes the LAD by the weighted median operator.
- *Basis selection problem:* applying a hard-threshold operator on the estimated values.
- Removing the entry's influence from the observation vector.
- Repeating these steps again until a performance criterion is reached.



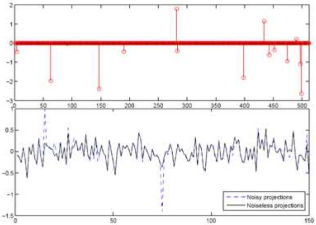
Original Signal



Iteration

Input and Initialize Parameters	Observation vector y Measurement Matrix Φ Number of Iterations K_0 Iteration counter: $k = 1$ Hard-thresholding value: $\lambda = \lambda_i$ Estimation at $k = 1$, $\widehat{x}^{(0)} = 0$
Iteration Step A	For the n-th entry of x, $n = 1, 2, \ldots, N$ compute $\tilde{x}_n = \text{MEDIAN}\left\{ \phi_{in} \diamond \frac{y_i - \sum_{j=1;j\neq n}^{N} \phi_{ij}\hat{x}_j}{\phi_{in}} \right\}\Big\|_{i=1}^{k}$ $\widehat{x}_n^{(k)} = \begin{cases} \tilde{x}_n, & \text{if } \; Q(\tilde{x}_n) + \lambda \leq Q(0) \\ 0, & \text{otherwise,} \end{cases}$
Step B	Update the hard-thresholding parameter and the estimation of x $x^{(k+1)} = x^{(k)}$ $\lambda = \lambda_i \beta^k$
Step C	Check stopping criterium If $k \leq K_0$ then set $k = k + 1$ and go to Step A; otherwise, end.
Output	Recovered sparse signal $\hat{\mathbf{x}}$

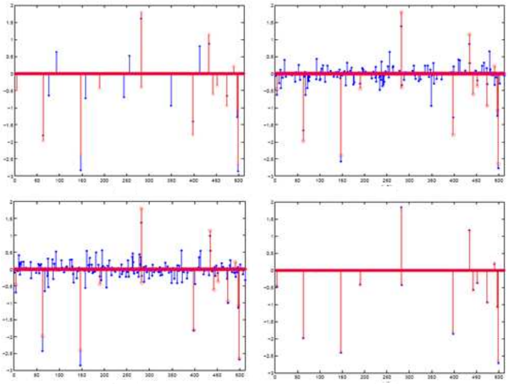
Example Signal Reconstruction from Projections in Non-Gaussian Noise



Original signal and noisy projections

CoSaMP

L1-Ls

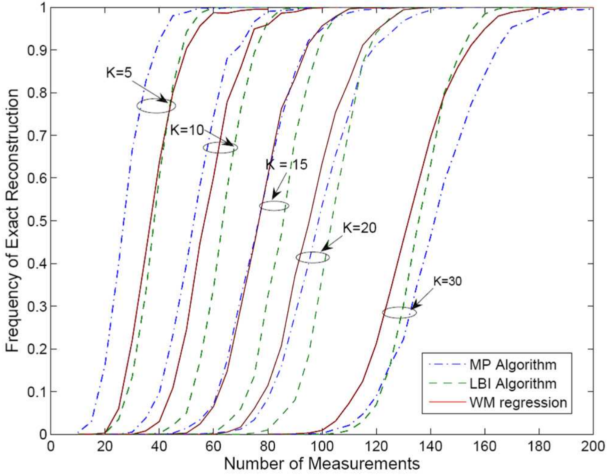


Matching Pursuit

Weighted Median

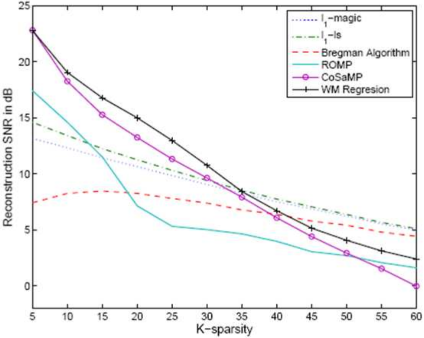
Performance Comparisons

Exact reconstruction for noiseless signals (normalized error is smaller than 0.1% of the signal energy).



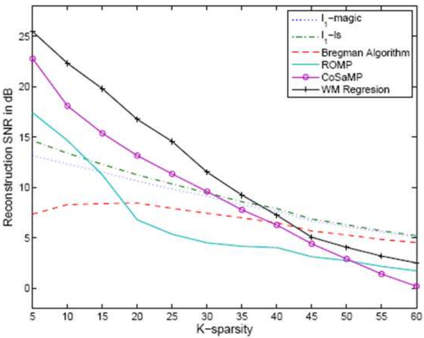
Performance Comparisons

SNR vs. Sparsity for different types of noise



(a)

Gaussian Noise



(b)

Laplacian Noise

Convex Relaxation Approach:

The ℓ_0-LAD problem can be relaxed by solving the following problem

$$\min_{x} \underbrace{\|y - \Phi x\|_1}_{\text{Data fitting}} + \underbrace{\lambda \|x\|_1}_{\text{Sparsity}} . \tag{2}$$

- This problem is convex and can be solved by linear programming.
- ℓ_1-norm may not induce the necessary sparsity in the solution.
- It requires more measurements to achieve a target reconstruction error.

The solution for the problem

$$\widehat{x} = \min_x \sum_{i=1}^{M} |y_i - \Phi_i x| + \lambda \|x\|_1$$

is given by

$$\hat{x} = \text{MEDIAN}(\lambda \diamond 0, |\phi_1| \diamond \frac{y_1}{\phi_1}, \ldots, |\phi_M| \diamond \frac{y_M}{\phi_M}).$$

- The regularization process induced by the ℓ_1-term leads to a weighting operation with the zero-valued sample in the WM operation.
- Large values of the regularization parameter implies large value for the weight corresponding to the zero-valued sample (this favors sparsity).
- The solution is biased.