Recovery from unstructured data

C. S. Sastry, G. Hennenfent and F.J. Herrmann

Organization

- The NFFT Problem
- Recent signal reconstruction methods
  - Basis Pursuit (BP)
  - Uniform Uncertainty Principle (UUP)
  - Orthogonal Matching Pursuit (OMP)
- Some problems with BP/UUP/OMP
- Our attempt
  - StOMP for NFFT

The NFFT problem [Kunis et. al. [5]]

- Discrete Fourier Transform of a 1D periodic signal f is

$$f(x) = \sum_{k=0}^{N-1} \hat{f}_k e^{2\pi ikx}.$$

- Given f values at samples $\{x_j\}_{j=0}^{M-1} \subset [0, 1)$, one can write DFT of f as

$$f(x_j) = \sum_{k=0}^{N-1} \hat{f}_k e^{2\pi ikx_j} \quad OR \quad f = A\hat{f}.$$

- When $M = N$ and $x_j = \frac{j}{M}$, the matrix A is invertible. Hence $\hat{f}$ values at uniform samples can be computed easily.
- Various practical problems, such as those in Seismic imaging, pose limitation on data acquisition at uniform samples. That is, x_j are not uniformly spaced and may at times be insufficient in numbers.
  Hence, inversion of A is not a simple process.

Signal recovery from (incomplete) measurements

<u>Basis Pursuit:</u> [Donoho et. al SIAM Review, 2001]

- To recover $x \in \mathbb{R}^n$ from given measurements $y \in R^m$ satisfying

$$y = Ax \quad for \quad A \in \mathbb{R}^{m\times n} \quad with \quad m << n$$

BP obtains solution via

$$\min_{\alpha} \|\alpha\|_1 \quad subject \quad to \quad A\alpha = y$$

provided x is sufficiently sparse and the **mutual coherence** parameter defined by

$$\mu = \max_{j\neq k} |< a_j, a_k >|.$$

satisfies

$$Sparsity(x) \leq \frac{1+\mu}{2\mu},$$

- it holds for all sparsity vectors and restrictions
- it's a very pessimistic bound, though l_1 problem can be solved using simple techniques.

Uniform Uncertainty Principle (UUP), [Candes et. al [2]]

- UUP provides a stable recovery algorithm for all sparse vectors
- **UUP:** Let $1 \leq S \leq n$. For all submatrices A_T of columns of A not exceeding S in numbers, there is $\delta_S > 0$ such that the columns in A_T are approximately orthonormal, as follows

$$(1 - \delta_S)\|c\|_2^2 \leq \|A_T c\|_2^2 \leq (1 + \delta_S)\|c\|_2^2, \quad \forall \quad c.$$

When $\delta_S + \delta_{2S} + \delta_{3S} < 1$, and $Sparsity(x) \leq S$, one can recover x satisfying

$$y = Ax$$

uniquely *with probability one* from

$$\min_{\alpha} \|\alpha\|_1 \quad subject \quad to \quad A\alpha = y,$$

where A is a general matrix possessing unit normed columns.

SLIM Seismic Laboratory for Imaging and Modeling

Orthogonal Matching Pursuit (OMP), [J.A. Tropp et. al [3]]

- OMP is an iterative procedure. At each iteration, OMP finds the column of A that is most strongly correlated with remaining part of y.
- it then subtracts off its contribution to y and then iterates on the residual
- it is an M - step process.
- it is too expensive

SLIM Seismic Laboratory for Imaging and Modeling

Problems with BP/UUP/OMP

- In BP, conditions on no. of measurements, sparsity of vectors are too pessimistic
- In UUP, conditions involving restricted isometry are difficult to verify in practice
- Though effective numerically, OMP has no theoretical guarantee on the qualitative behaviour of the solution.

Desirable features

- should ideally assume relatively weaker and easily verifiable conditions
- should admit easier implementation and be widely applicable
- be fast for very large scale problems

SLIM Seismic Laboratory for Imaging and Modeling

StOMP fits the bill very well

Comparison of execution times of different algorithms:

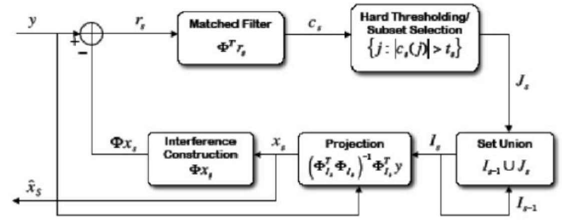
Problem Suite (k,n,N)	ℓ_1	OMP	CFAR	CFDR
(10,100,1000)	0.97	0.37	0.02	0.03
(100,500,1000)	22.79	0.79	0.42	0.32
(100,1000,10000)	482.22	7.98	5.24	3.28
(500,2500,10000)	7767.16	151.81	126.36	88.48

SLIM Seismic Laboratory for Imaging and Modeling

Stagewise Orthogonal Matching Pursuit (StOMP)

- In StOMP at each iteration, a set of columns (as opposed to one column in OMP) having higher correlations is identified
- Like BP, StOMP enjoys strong theoretical guarantees
- StOMP involves a fixed number S ($= 10$) of steps.
- The choice of S and threshold parameters are made such that
  - ➢ all nonzeros in x_0 are selected at the end of S^{th} iteration
  - ➢ Sparsity of x_0 (k) $\leq$ No. of measurements (n), which ensures perfect recovery

StOMP block diagram



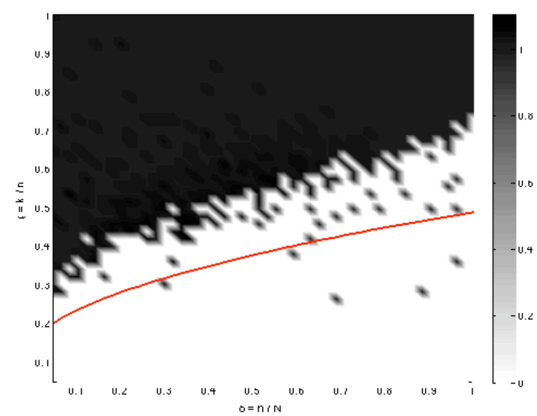
- **Threshold Selection**
  - ➢ made using ideas from statistical decision theory, which assume Gaussian behaviour of residual, that is,

$$\tilde{x} := A^T y \text{ and } z := \tilde{x} - x_0 \text{ is Gaussian}$$

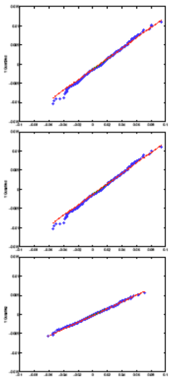
  - ➢ Gausianity is verified by QQ-plots and quantified by KL measure
- **Performance analysis**
  - ➢ shown in terms of phase diagram of points $(n/N, k/n)$
  - ➢ which shows the transition from success to failure for different k and n
- **Complexity analysis**
  - ➢ When A is explicit, matrix - vector multiplication involves MN flops
  - ➢ When A has operator form (like FT, WT etc), implementation can be carried out fast

- **Application to Fourier matrix ensemble**

  I. with incomplete and regular measurements



Gausianity test



n=208,k=144, slope measure=11, Kullback-Leibler distance: 0.058

n=237,k=78, slope measure=9.2, KL-distance: 0.05

n=560,k=105, Slope measure: 4.6 KL distance: 0.0062, distribution is close to being Gaussian

II. with incomplete and irregular measurements

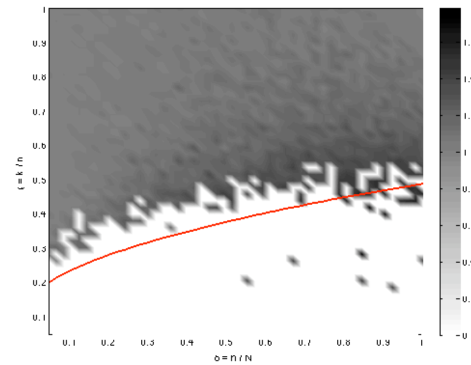


Figure 1: StOMP with NFFT for random set of measurements

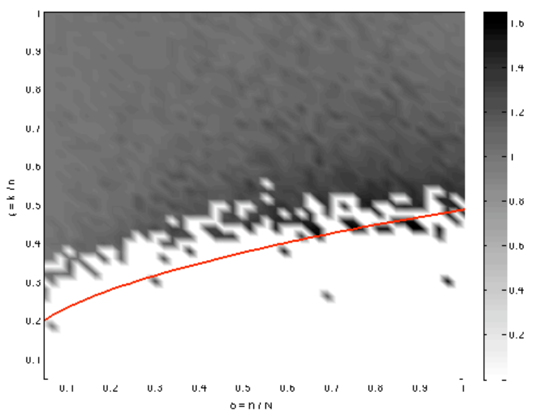


Figure 2: StOMP with NFFT for random set of measurements

Observations:

- For Fourier matrix ensemble with regular measurements, empirical decision boundary can be less pessimistic than the theoretical boundary
- For Fourier matrix ensemble with irregular measurements, actual and theoretical decision boundaries almost coincide
- The recovery conditions are weaker and the empirical recovery is well within the theoretical bounds

Discrepancy of a set $I = \{x_i\}_1^M$

- \- gives quantitative measure of deviation of the sequence from good (uniform) distribution

$$D_M = \sup_{x_1 \leq \alpha < \beta \leq x_M} \left| \frac{\#(I \cap [\alpha, \beta))}{M} - (\beta - \alpha) \right|$$

$$\frac{1}{M} \leq D_M \leq 1$$

- When the measurements are uniformly distributed, every subinterval gets its due share.

SLIM Seismic Laboratory for Imaging and Modeling

Work in progress/objective

- implementations using 'fast operator' computations, which avoids the time-consuming vector matrix multiplications
- studying decision boundary in terms of discrepancy of data points obeying different distributions
- analyzing the applicability of the method in terms of Gaussianity test

Application:

- As seismic data are typically irregularly sampled and curvelet implementations are carried out in frequency domain, the equispaced Fourier frequencies so computed using NFFT can be directly used in the computation of curvelet coefficients (See Gilles & Herrmann [6]), and hence the methodology can be used for various Seismic imaging applications.
- From phase diagram, knowing k, n can be determined, and vice versa.

SLIM Seismic Laboratory for Imaging and Modeling

References

[1] Chen S.S., Donoho D.L., Saunders M. A. (2001), "Atomic decomposition by basis pursuit,", SIAM Review, Vol. 43, No.1 pp: 129-159.

[2] E. J. Candes, J. Romberg and T. Tao, "Stable Signal Recovery from Incomplete and Inaccurate Measurements," Submitted, 2005

[3] J.A. Tropp, A. Gilbert, "Signal recovery from partial information via orthogonal matching pursuit," submitted 2005.

[4] D.L. Donoho, Y. Tsaig, I. Drori, J.L. Starck, "Sparse solutions underdetermined linear equations by stagewise orthogonal matching pursuit.

[5] Kunis and D. Potts, "NFFT2.0," 2005.

[6] G. Hennenfent and F.J. Herrmann, "Seismic denoising with non-uniformly sampled curvelets," Submitted, 2005.

SLIM Seismic Laboratory for Imaging and Modeling