A primer on weak recovery conditions

Context

"Sparse Solution of Underdetermined Linear Equations by Stagewise Orthogonal Matching Pursuit" David L. Donoho, Yaakov Tsaig, Iddo Drori, Jean-Luc Starck

"Sparse Nonnegative Solution of underdetermined Linear Equations by Linear Programming" David L. Donoho and Jared Tanner

"High-Dimensional Centrally-Symmetric Polytopes With Neighborliness Proportional to Dimension" David L. Donoho

Challenges

Seismic data recovery is extremely large **scale**

Strong recovery conditions are

- prohibitively expensive to calculate
- overly pessimistic

Need a framework for **large** data volumes

- feasible conditions for recovery
- qualitative tests for recovery
- quantitative test for recovery

Only hope for an approximate solution to the recovery problem ...

Sparse recovery

[Candes, Donoho etc.]

Uniform uncertainty principles (UUP):

- **strong** conditions for inverting underdetermined systems
- strong equivalence between l1 and l0
- valid for any sparsity vector $\mathbf{x}_0$
- unique solution in noise-free case
- pessimistic bounds
- unfeasible bounds for 'large' systems

Donoho proposes weak conditions under the slogan:

noiseless underdetermined problems behave like noisy well-determined problems ...

The problem

Linear system

$$\mathbf{Ax} = \mathbf{y}$$

with

$$\mathbf{A} := \mathbf{RMS}^H$$

and

- **x** the N-length sparsity vector (N>>n) with
- k non-zero entries
- **y** the n-length measurement vector

SLIM Seismic Laboratory for Imaging and Modeling

Strong recovery conditions
[Candes et al]

Prescribe recovery conditions for the
- measurement matrix
- sparsity matrix and vector

Recovery depends on
- mutual coherence
- compression of the sparsity vector

Guarantees recovery for every possible restriction and sparsity vector

Recovery conditions impossible to compute & pessimistic

SLIM Seismic Laboratory for Imaging and Modeling

UUP examples

Recovery conditions are
- pessimistic
- in practice much lower
- sharp for each experiment but different

Indication of a phase transition
- large systems with randomness
- UUP's worked well with randomness in the restriction and/or sparsity vector

Opens the way to study large systems in probabilistic a framework.

SLIM Seismic Laboratory for Imaging and Modeling

Weak recovery conditions
[Donoho et al]

Approximate probabilistic framework:
- valid for **typical** sparsity vectors
- requires mixing (randomness)
- more accurate for larger systems

Mixing depends on
- randomness restriction matrix
- incoherence measurement-sparsity matrices
- randomness & compression of sparsity vector

Fast recovery possible for typical sparsity vectors when

Gaussian approximation is valid …

SLIM Seismic Laboratory for Imaging and Modeling

Gaussian approximation

Matched filter minus sparsity vector needs to be Gaussian, i.e.,

$$\mathbf{z} = \tilde{\mathbf{x}} - \mathbf{x}_0 \quad \text{with} \quad \tilde{\mathbf{x}} = \mathbf{A}^H \mathbf{y}$$

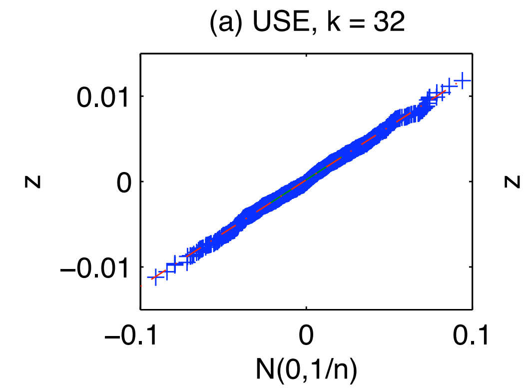
is Gaussian

- depends on the mixing
- can be tested qualitatively with QQ-plot
- related to digital communication theory
- easy to check for large systems

Example:

- Uniform Spherical Ensemble (random matrix with normalized columns)

QQ Plot



linearity corresponds to Gaussianity

Gaussian approximation

Translates in seismic situation to

- source/receiver positions distributed uniformly
- random sampling is known to reduce the adverse effects from under sampling

When valid there exists a fast approximate solution for typical recovery problems.

Exist accurate recovery conditions (phase diagrams) for medium size problems.

Allows design acquisition geometries.

Phase diagrams

Interested in two questions:

- *How many measurements does one need to take to recover typical seismic wavefields to within a prescribed accuracy?*
- *To what accuracy can one reconstruct typical seismic wavefields given a certain acquisition grid and noise level?*

Phase diagrams delineate regions for successful (white) and unsuccessful recovery(dark).

Phase diagrams

Define "measurement density"

$$\delta = n/N$$

as the ratio of the number of measurements over the length of the sparsity vector

and " sparsity density"

$$\rho = k/n.$$

the ratio of the number of non-zero entries in the sparsity vector over the number of measurements

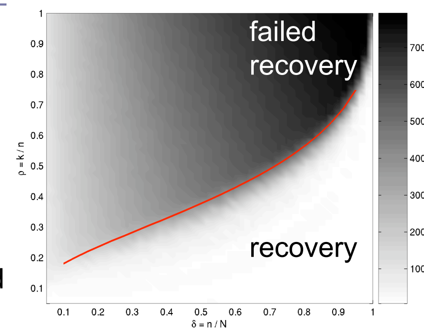
Phase diagrams tests the recovery for combinations $(\delta, \rho) \in [0,1] \times [0,1]$

Phase diagrams

l1 solver [from Donoho et al]

fully sampled rich signal

fully sampled rich signal



under sampled sparse signal

fully sampled sparse signal

In the white region

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \mathbf{A}\mathbf{x} = \mathbf{y}$$

recovers exactly.

Phase diagrams

[Donoho et al]

White area corresponds to **successful** recovery.

Dark area corresponds to **unsuccessful** recovery.

Can be used to answer two main questions

- expensive to compute
- for large systems sharp transition
- theory applies to ideal matrices

How can these results be extended to large to very large systems?

StOMP

[from Donoho et al]

Stage-wise Orthonormal Matching Pursuit:

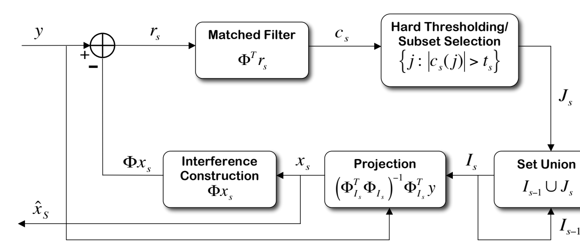
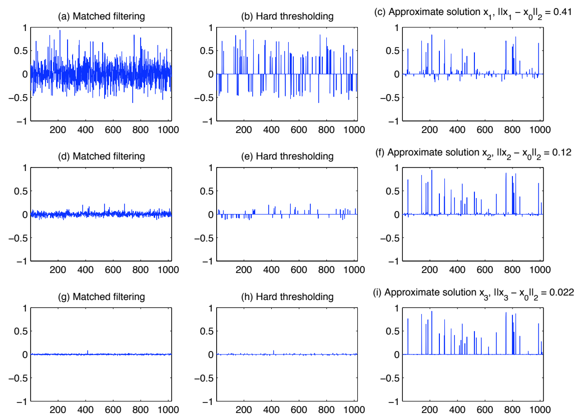


Figure 1: Schematic Representation of the StOMP algorithm.

Solves problem approximately for typical problems ...

StOMP
[from Donoho et al]



StOMP
[from Donoho et al]

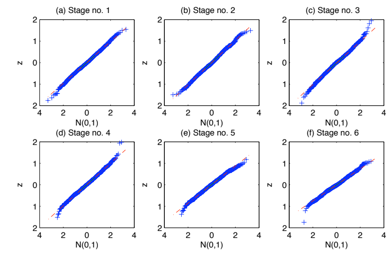


Figure 4: QQ plots comparing residual MAI with Gaussian distribution. Quantiles of residual MAI at different stages of StOMP are plotted against Gaussian quantiles. Near-linearity indicates approximate Gaussianity.

StOMP

Norm-one solvers are prohibitively expensive:

Algorithm	Dense Matrices	Fast Operators
StOMP	$S(\nu+2)nN + O(N)$	$S(\nu+2)\cdot V + O(N)$
OMP	$4n^3/3 + n^2N + O(N)$	$4n^3/3 + 2n\cdot V + O(N)$
ℓ_1 min. with PDCO	$S(2N)^3/3 + O(nN)$	$2N \cdot V + O(nN)$

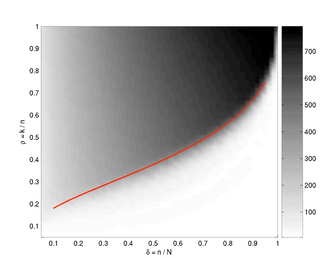
with

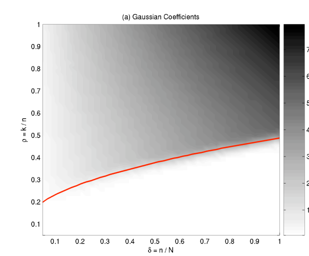
- **N length sparsity vector**
- **V=cost 1 matrix-vector multiplication**
- **S=10 the number of loops of StOMP**
- ν **number of lsqr iterations**

Fast operators and StOMP save!

StOMP versus BP

Stage-wise Orthonormal Matching Pursuit:





- **approximate**
- **good for sparse problems**
- **Order of magnitude faster**

StOMP

[from Donoho et al]

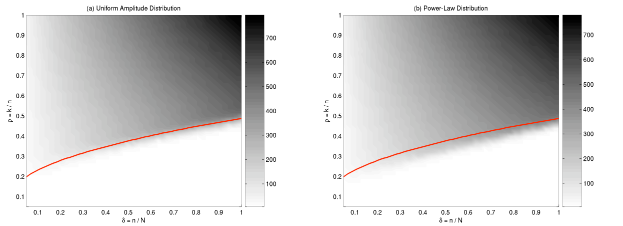


Figure 10: Phase Diagrams for CFAR thresholding when the nonzeros have nonGaussian distributions. Panel (a) uniform amplitude distribution; Panel (b) Power law distribution $x_0(j) = c/j$, $j = 1, \ldots, k$. Compare to Figure 6. Shaded attribute: number of entries where reconstruction misses by 10^{-4}.

StOMP

[from Donoho et al]

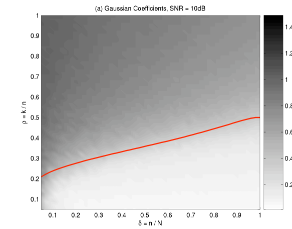
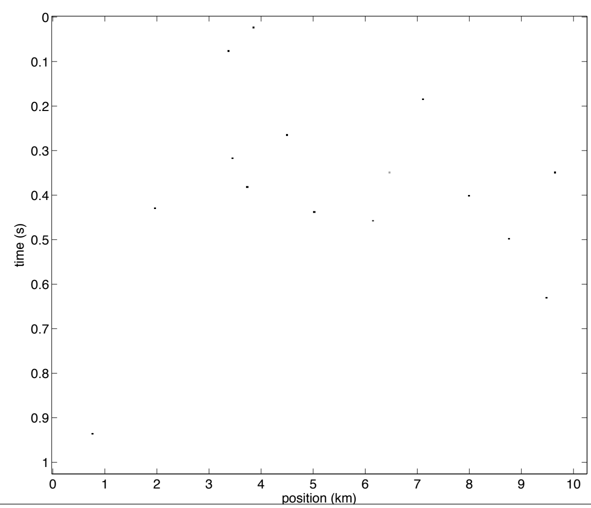
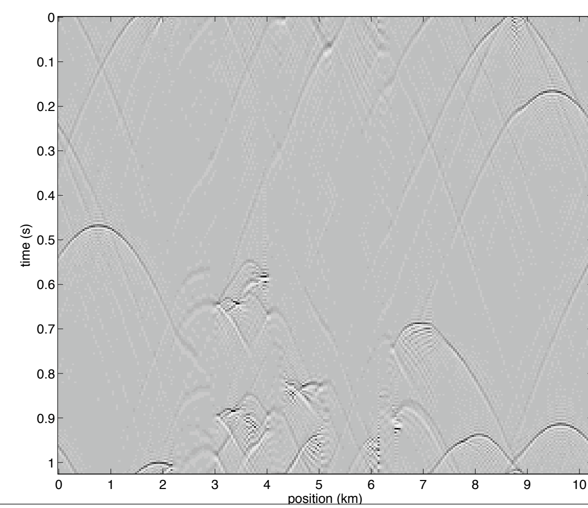


Figure 11: Performance of CFDR thresholding, noisy case. Relative ℓ_2 error as a function of indeterminacy and sparsity. Performance at signal-to-noise ratio 10. Shaded attribute gives relative ℓ_2 error of reconstruction. Signal-to-Noise ratio is $\|x_0\|_2/\|\Phi^T\xi\|_2$.

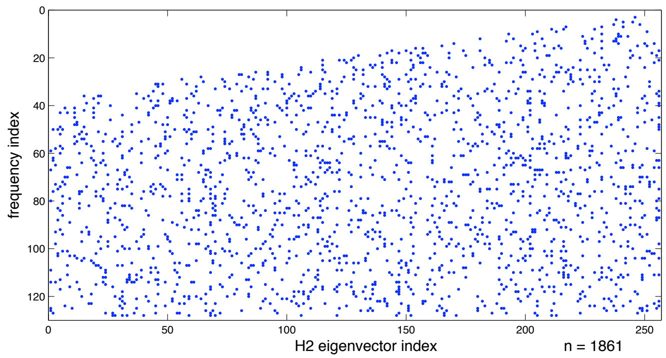
Compressed imaging



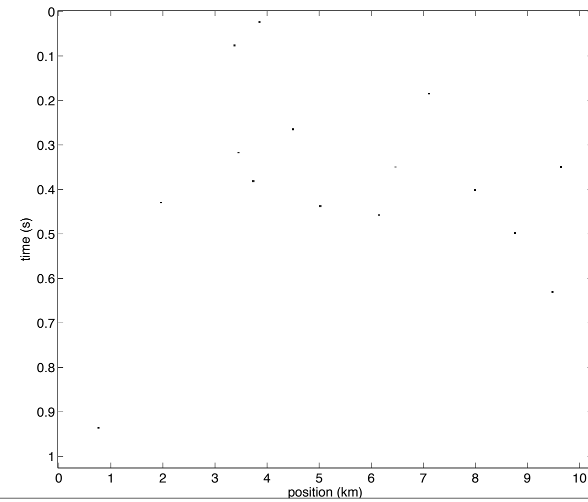
Compressed imaging



Compressed imaging



Compressed imaging



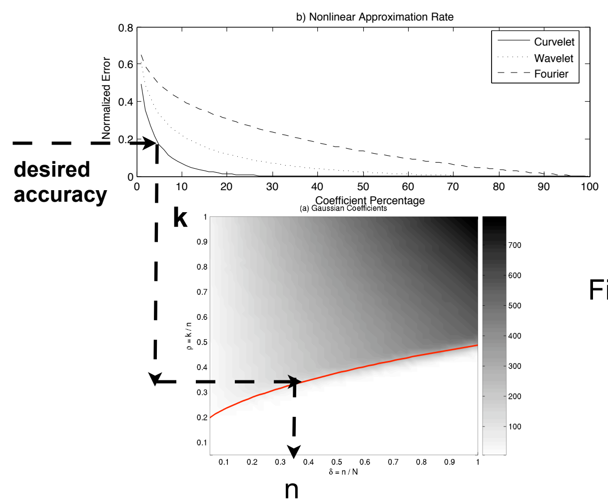
Use of phase diagram

How many measurements does one need to to take to recover typical seismic wavefields to within a prescribed accuracy?

- determine grid size of the problem N
- plots for nonlinear approximation rate => k
- phase diagram => n for which recovery is feasible

Assumes knowledge of the decay of the approximation error for a typical seismic data set.

Use of phase diagram



Fix N

Use of phase diagram

To what accuracy can one reconstruct typical seismic wavefields given a certain acquisition size, geometry and noise level?

- use N (grid size), n (# of measurements) to determine recoverable k
- use k in the NLA plot to compute the recovery error
- for noisy data use adapted phase diagram with the relative recovery error

Assumes knowledge on the phase diagram. Unattainable in practice ...

SLIM Seismic Laboratory for Imaging and Modeling

Use of phase diagram

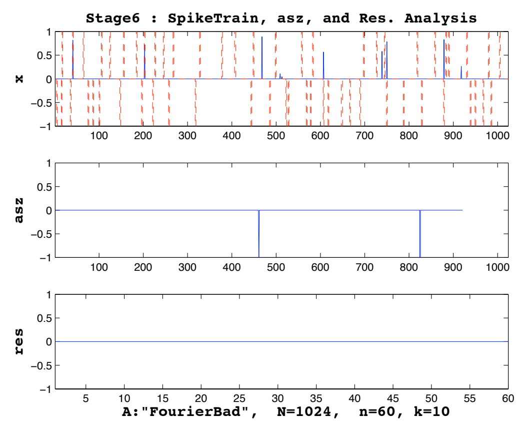
Analytic results for the phase transition are known for idealized matrices.

We will show empirically how far we can extend.

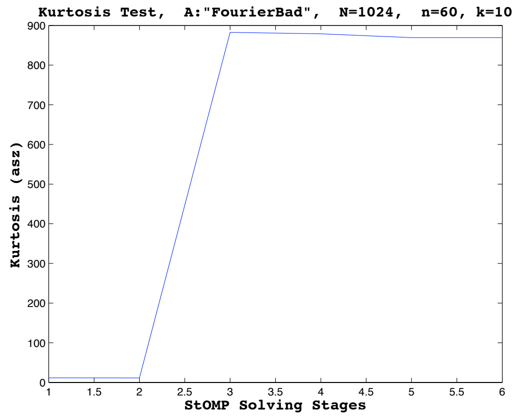
Remaining questions

- can analytic results be extended to other matrices?
- can we come up with qualitative and quantitative recovery conditions for large problems?
- MAI may provide perspective ...

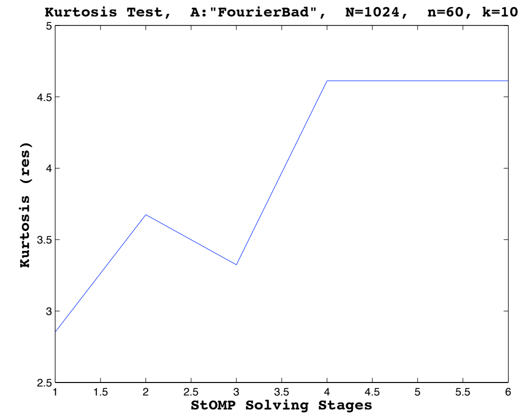
SLIM Seismic Laboratory for Imaging and Modeling



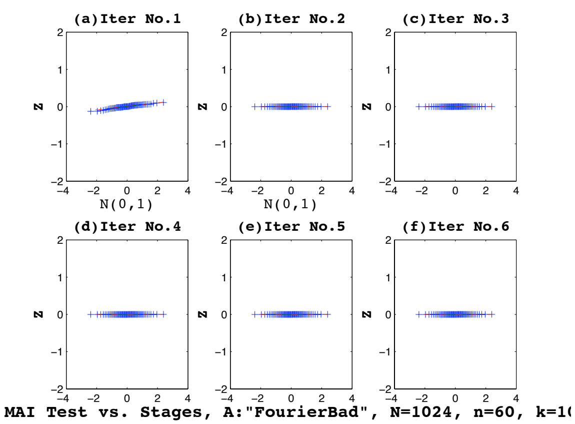
SLIM Seismic Laboratory for Imaging and Modeling



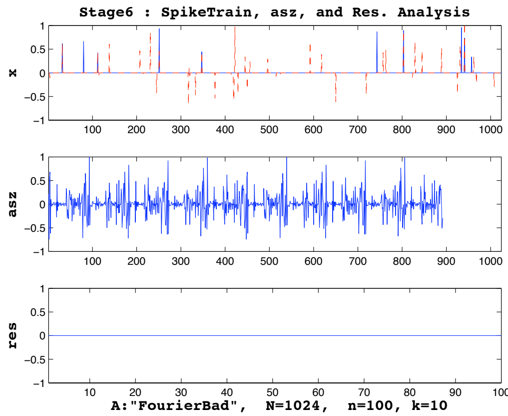
SLIM Seismic Laboratory for Imaging and Modeling



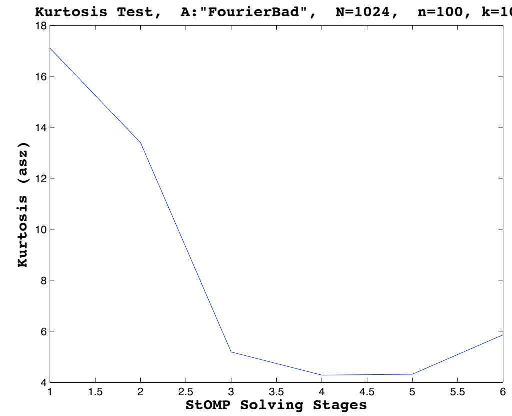
SLIM Seismic Laboratory for Imaging and Modeling



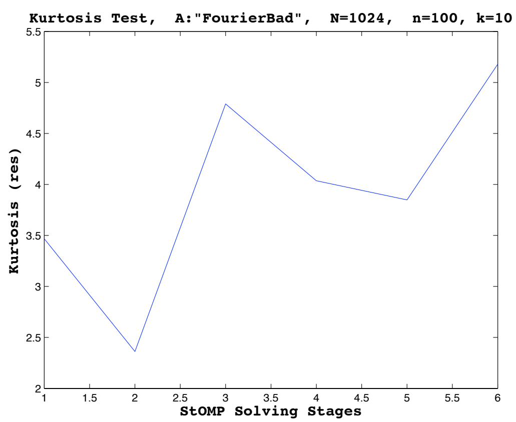
SLIM Seismic Laboratory for Imaging and Modeling



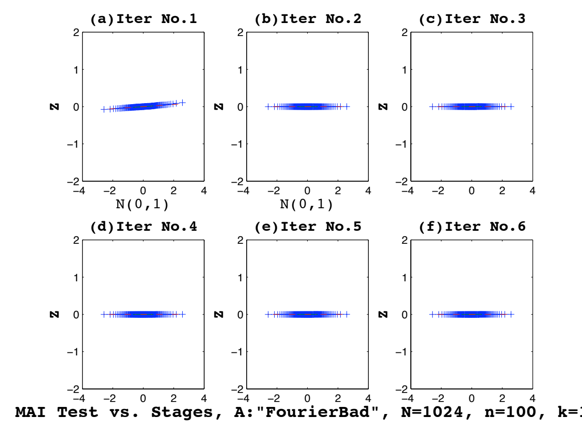
SLIM Seismic Laboratory for Imaging and Modeling



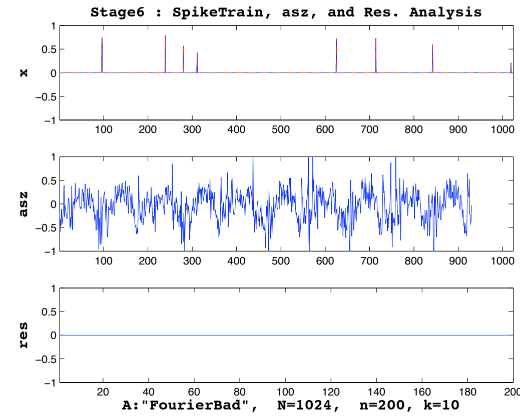
SLIM Seismic Laboratory for Imaging and Modeling



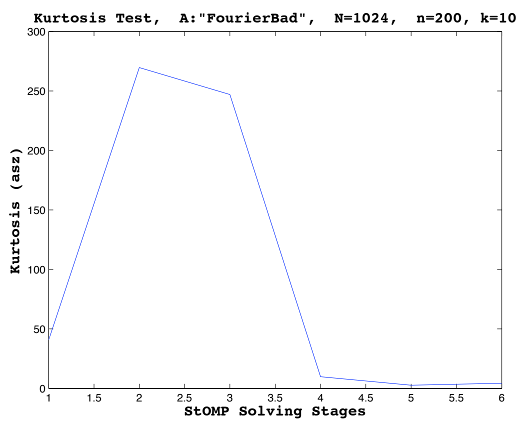
SLIM Seismic Laboratory for Imaging and Modeling



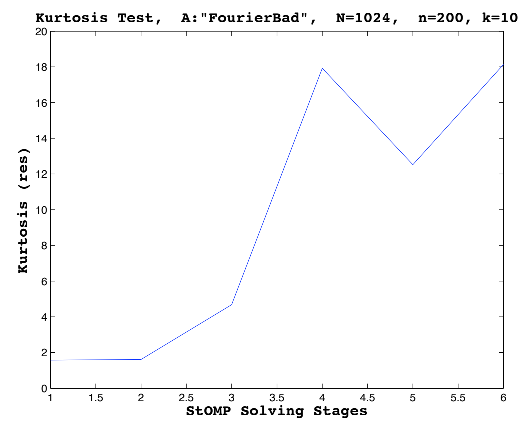
SLIM Seismic Laboratory for Imaging and Modeling



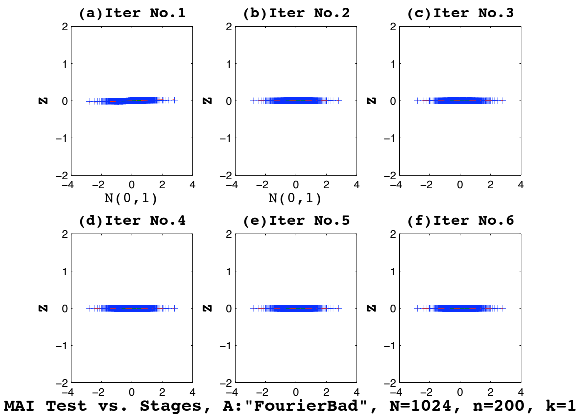
SLIM
Seismic Laboratory for Imaging and Modeling



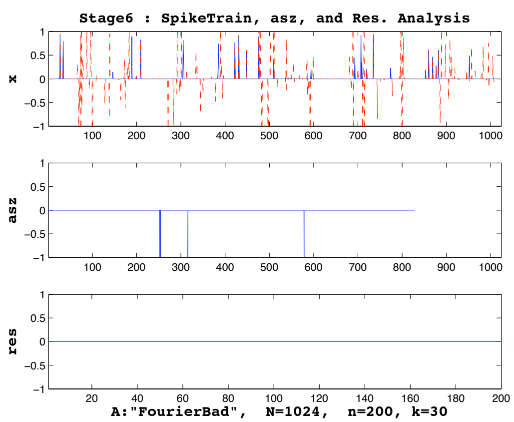
SLIM
Seismic Laboratory for Imaging and Modeling



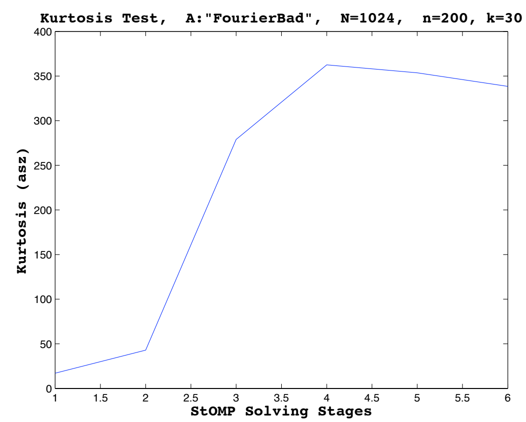
SLIM
Seismic Laboratory for Imaging and Modeling



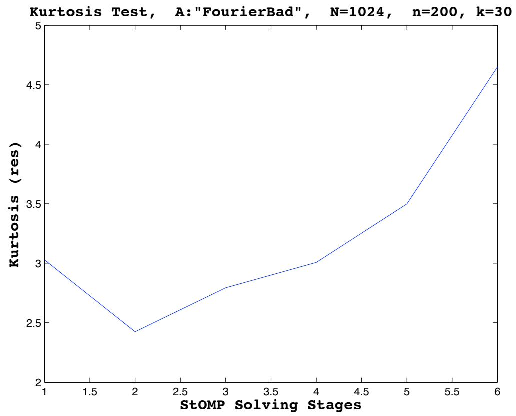
SLIM
Seismic Laboratory for Imaging and Modeling

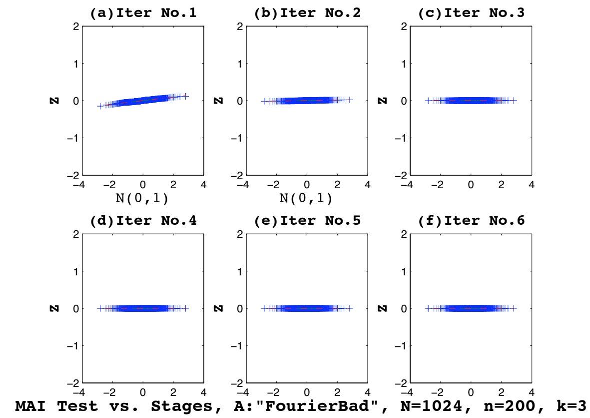


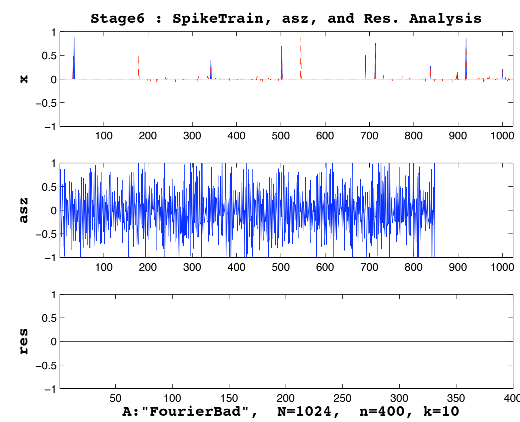
SLIM
Seismic Laboratory for Imaging and Modeling

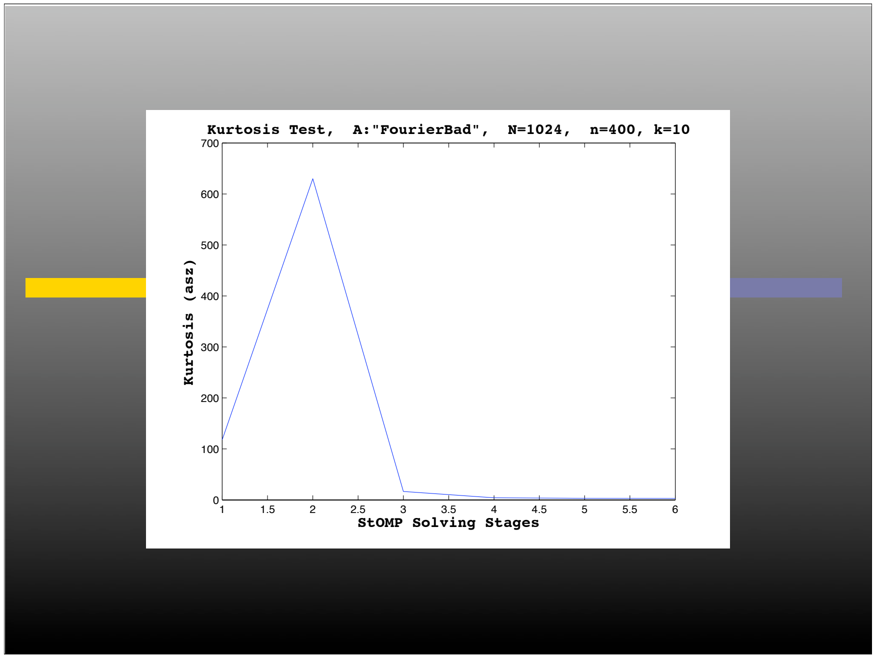


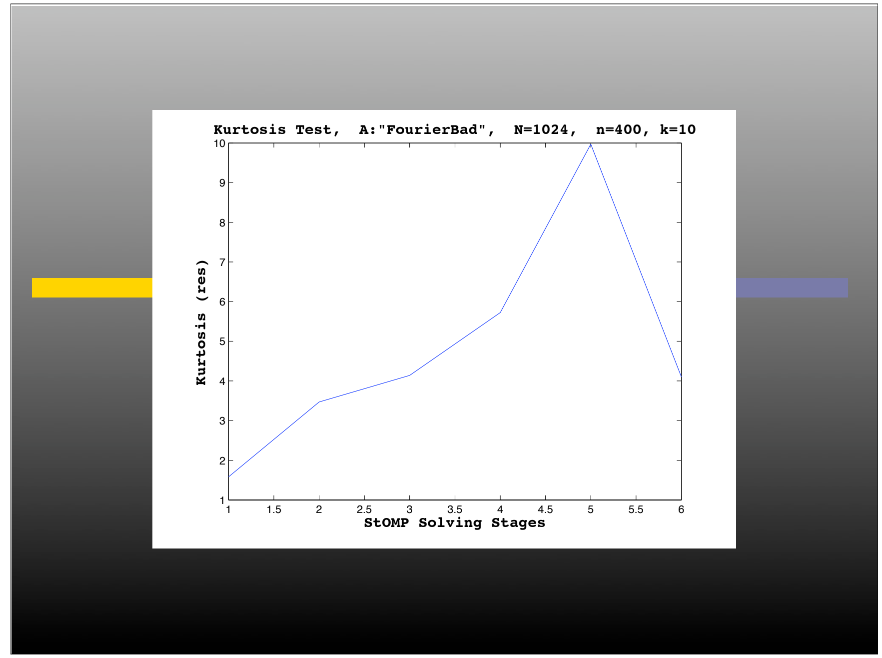
SLIM
Seismic Laboratory for Imaging and Modeling

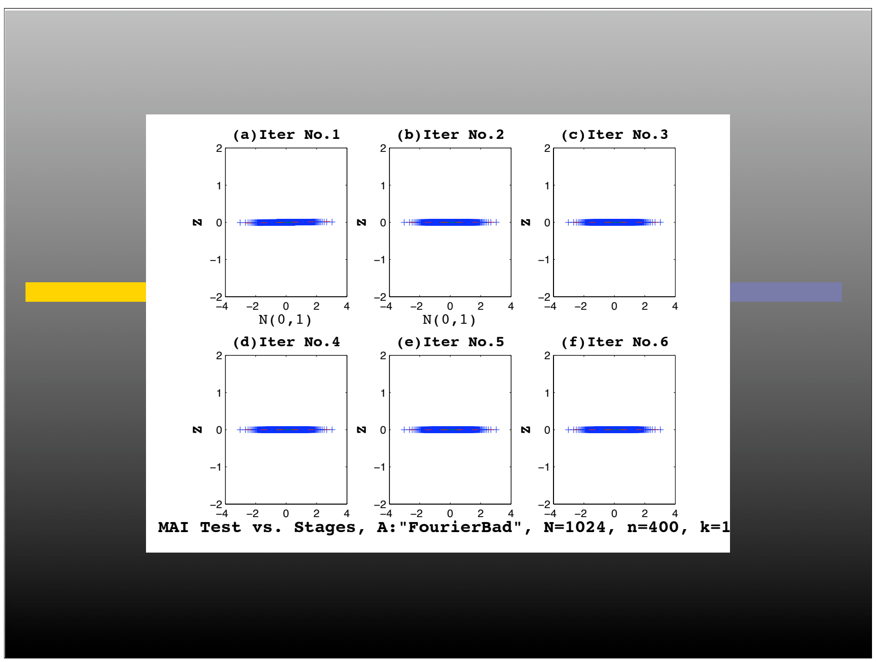

Kurtosis Test, A:"FourierBad", N=1024, n=200, k=30
Kurtosis (res)
StOMP Solving Stages

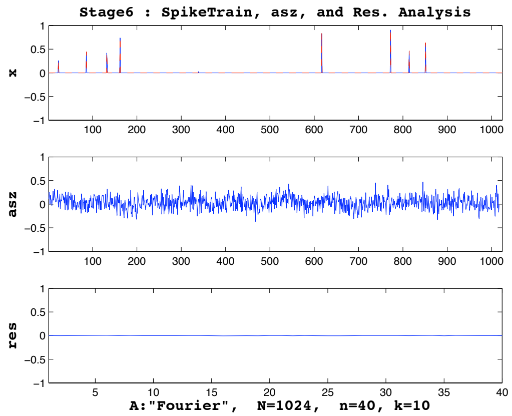

(a)Iter No.1
(b)Iter No.2
(c)Iter No.3
(d)Iter No.4
(e)Iter No.5
(f)Iter No.6
z
N(0,1)
MAI Test vs. Stages, A:"FourierBad", N=1024, n=200, k=3


Stage6 : SpikeTrain, asz, and Res. Analysis
x
asz
res
A:"FourierBad", N=1024, n=400, k=10

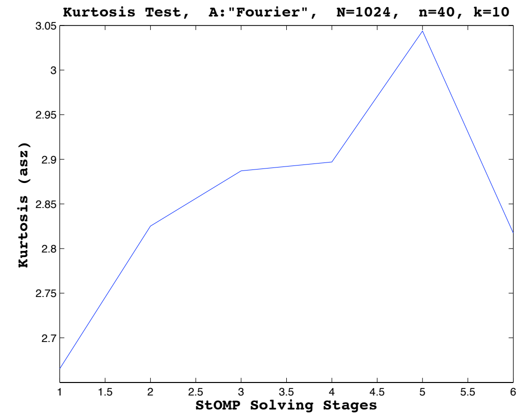

Kurtosis Test, A:"FourierBad", N=1024, n=400, k=10
Kurtosis (asz)
StOMP Solving Stages


Kurtosis Test, A:"FourierBad", N=1024, n=400, k=10
Kurtosis (res)
StOMP Solving Stages

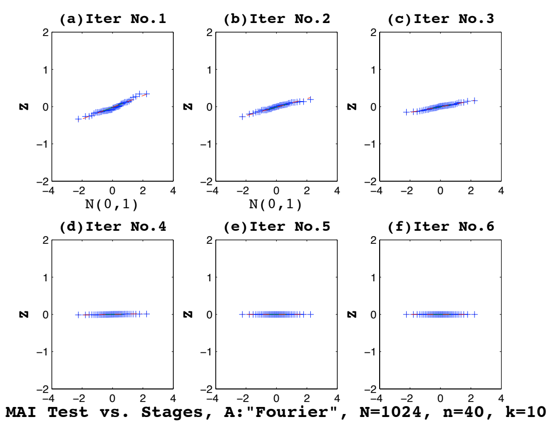

(a)Iter No.1
(b)Iter No.2
(c)Iter No.3
(d)Iter No.4
(e)Iter No.5
(f)Iter No.6
N(0,1)
MAI Test vs. Stages, A:"FourierBad", N=1024, n=400, k=1



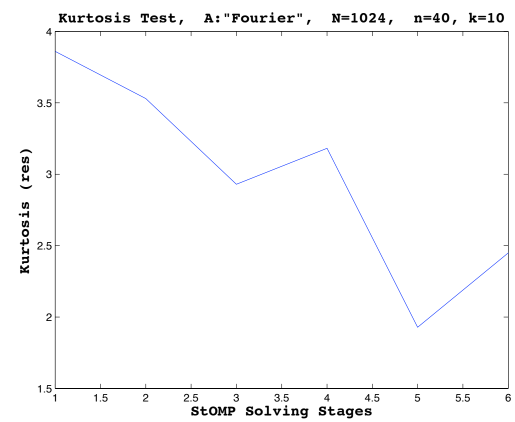
Stage6 : SpikeTrain, asz, and Res. Analysis
x
asz
res
A:"Fourier", N=1024, n=40, k=10



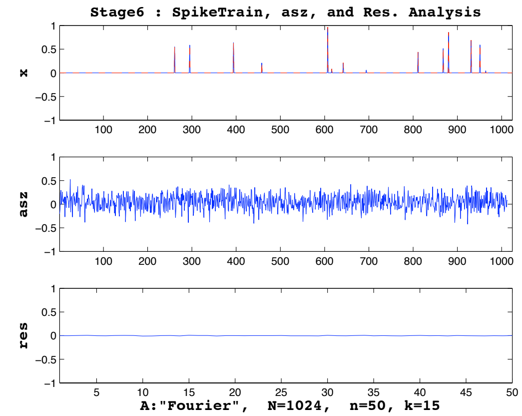
Kurtosis Test, A:"Fourier", N=1024, n=40, k=10
Kurtosis (asz)
StOMP Solving Stages

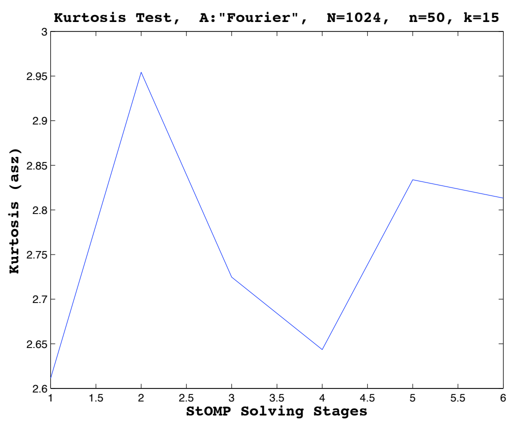


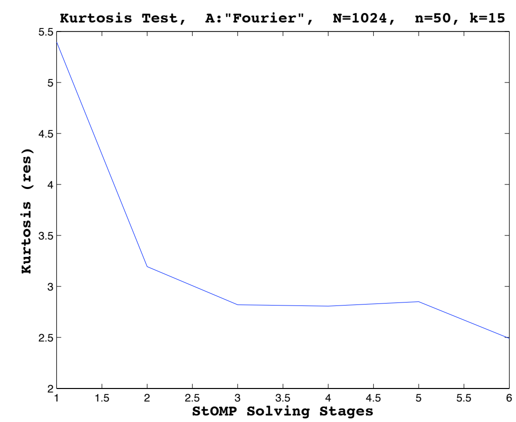
(a)Iter No.1
(b)Iter No.2
(c)Iter No.3
(d)Iter No.4
(e)Iter No.5
(f)Iter No.6
z
N(0,1)
MAI Test vs. Stages, A:"Fourier", N=1024, n=40, k=10

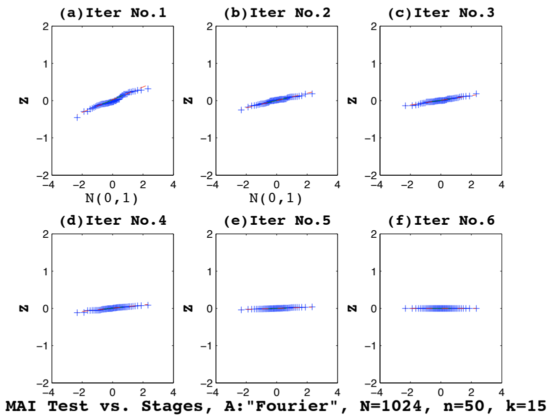


Kurtosis Test, A:"Fourier", N=1024, n=40, k=10
Kurtosis (res)
StOMP Solving Stages

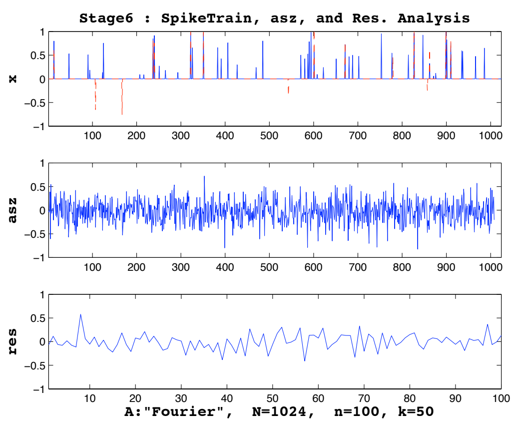

Stage6 : SpikeTrain, asz, and Res. Analysis
x
asz
res
1
0.5
0
−0.5
−1
100
200
300
400
500
600
700
800
900
1000
5
10
15
20
25
30
35
40
45
50
A:"Fourier", N=1024, n=50, k=15


Kurtosis Test, A:"Fourier", N=1024, n=50, k=15
Kurtosis (asz)
3
2.95
2.9
2.85
2.8
2.75
2.7
2.65
2.6
1
1.5
2
2.5
3
3.5
4
4.5
5
5.5
6
StOMP Solving Stages

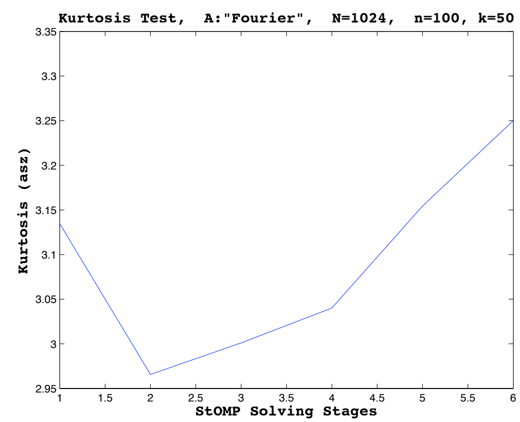

Kurtosis Test, A:"Fourier", N=1024, n=50, k=15
Kurtosis (res)
5.5
5
4.5
4
3.5
3
2.5
2
1
1.5
2
2.5
3
3.5
4
4.5
5
5.5
6
StOMP Solving Stages



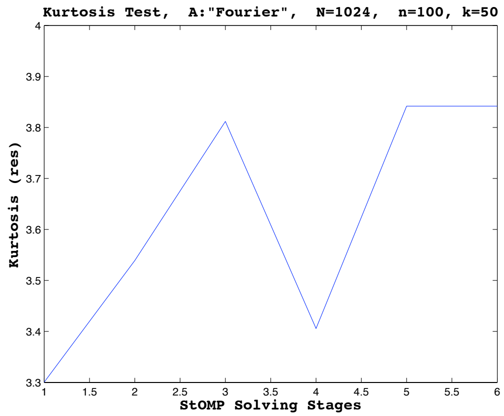
SLIM Seismic Laboratory for Imaging and Modeling

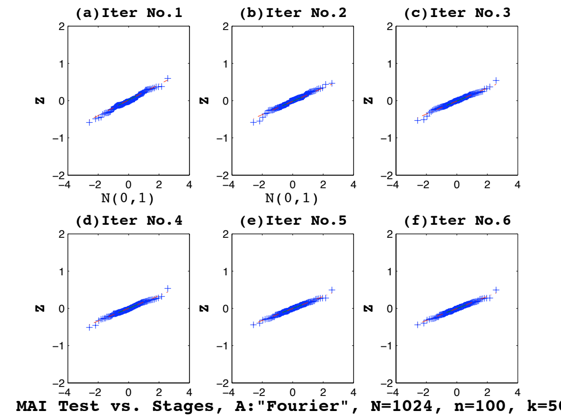


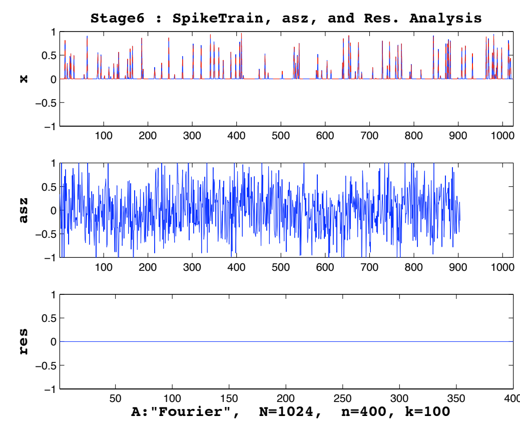
SLIM Seismic Laboratory for Imaging and Modeling

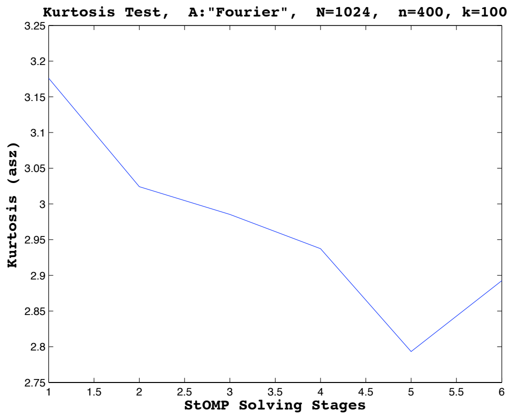


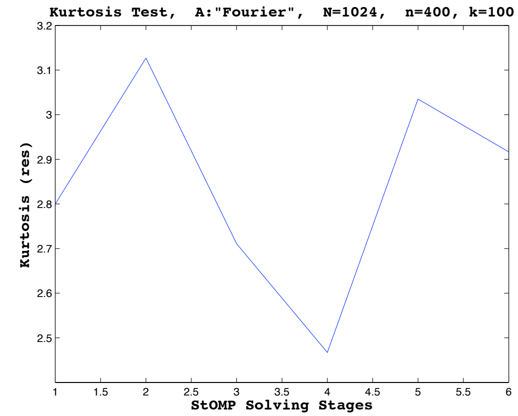
SLIM Seismic Laboratory for Imaging and Modeling

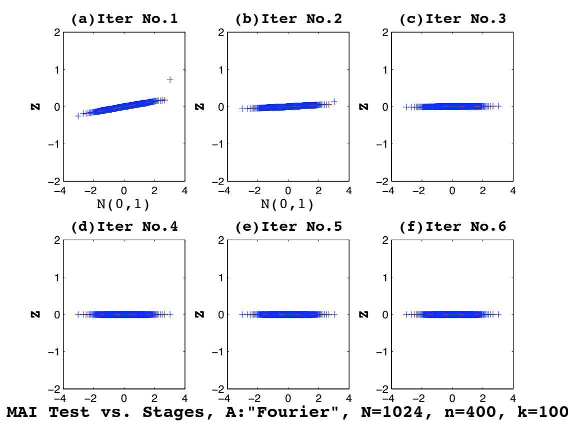

Kurtosis Test, A:"Fourier", N=1024, n=100, k=50
Kurtosis (res)
StOMP Solving Stages


(a)Iter No.1
(b)Iter No.2
(c)Iter No.3
(d)Iter No.4
(e)Iter No.5
(f)Iter No.6
N(0,1)
z
MAI Test vs. Stages, A:"Fourier", N=1024, n=100, k=50


Stage6 : SpikeTrain, asz, and Res. Analysis
x
asz
res
A:"Fourier", N=1024, n=400, k=100







Conclusions

For the good Fourier, i.e. mixing matrix

- if kurtosis of MAI & residue are app. 3 then likely recovery
- when close to non-recoverable region then behavior difficult to predict
- if both kurtosi are far from >> 3 then no recovery likely

For the bad Fourier, i.e. no mixing matrix

- difficult to predict
- recoverable region is much smaller
- kurtosis is not a good indicator

Outlook

StOMP is orders of magnitude faster then BP

Weak recovery conditions better suited for

Test for Gaussianity and detectability may be feasible for given

- typical sparsity vector (e.g. permutation of histogram)
- typical acquisition geometry
- sparsity matrix

Findings are an extension of known matched filter arguments (randomness = good)

May lead to feasible tests for recovery.

SLIM Seismic Laboratory for Imaging and Modeling

Outlook

Applications discussed during the meeting

- event detection (Mohhammad)
- compressed imaging (Tim)
- NFFT's (Sastry)

Extension to large systems

Relation to other solvers

Monte-Carlo sampling for the phase diagram with sparse recovery?

SLIM Seismic Laboratory for Imaging and Modeling