Grid-Independent Compressed Sensing

Albert Fannjiang & Wenjing Liao
Math Dept, UC Davis

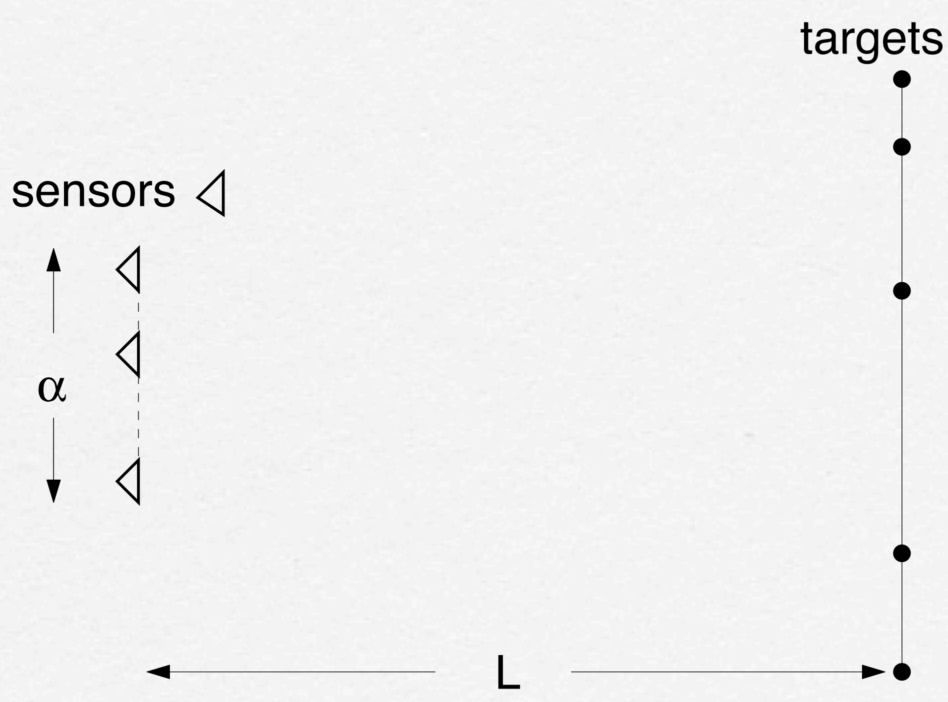
UC Irvine, Feb 15, 2013

Outline

- Imaging set-up
- Compressed sensing formulation
- Theory for grid-independent CS
- Numerics for GICS
- Comparisons
- Super-resolution
- conclusion

Fourier measurement

- Source localization



Paraxial green function

$$
\begin{aligned}
G(r,\xi) &= \frac{e^{i\omega L}}{4\pi L} \times \exp\left(\frac{i\omega|r-\xi|^2}{2L}\right) \\
&= \frac{e^{i\omega L}}{4\pi L} \exp\left(\frac{i\omega r^2}{2L}\right) \exp\left(\frac{-i\omega r\xi}{L}\right) \exp\left(\frac{i\omega \xi^2}{2L}\right)
\end{aligned}
$$

Grid model

Source locations: $x_l : l = 1, ..., s$

Source strengths: $c_l, l = 1, ..., s$.

Signal model: at the sensor located at $\xi_l, l = 1, ..., N$

$$y_l = \sum_{j=1}^{s} c_j G(\xi_l, x_j) + n_l.$$

Approximate x_j by the closest subset of cardinality s of a regular grid $\mathcal{G} = \{p_1, \ldots, p_M\}, M \gg s,$.

Write $\mathbf{x} = (x_j) \in \mathbb{C}^M$ where $x_j = c_j$ whenever the grid points are the nearest grid points to the targets and zero otherwise.

Linear inversion

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

Measurement matrix $\mathrm{A} = \mathrm{D}_1 \Phi \mathrm{D}_2$

$$\Phi_{jl} = \frac{1}{\sqrt{N}} \exp\left(\frac{-i\omega x_l \xi_j}{L}\right)$$

$$\mathrm{D}_1 = \mathrm{diag}\left(\exp\left(\frac{i\omega \xi_j^2}{2L}\right)\right)$$

$$\mathrm{D}_2 = \mathrm{diag}\left(\exp\left(\frac{i\omega x_l^2}{2L}\right)\right)$$

Error = external noise + gridding error

Resolution limit

W/O additional prior information, we can only hope to recover targets separated by at least one Rayleigh length

$$\ell = \frac{\lambda L}{a} = 1$$

a = aperture, L = distance, λ =wavelength

Compressed sensing (CS)

Candes, Donoho, Romberg, Tao, Tibshirani, Tropp......

- Restricted isometry property (RIP)

$$(1-\delta_k)\|\mathbf{x}\|_2^2 \le \|\mathbf{A}\mathbf{x}\|_2^2(1+\delta_k)\|\mathbf{x}\|_2^2, \quad \text{where } \mathbf{x} \text{ is } k\text{-sparse}$$

$$k = 2s, \quad \delta_{2s} < \sqrt{2} - 1$$ (Candes 08)

- Incoherence property (IP)

$$\mu(\mathbf{A}) = \max_{j \neq l} \mu(k,l), \quad \mu(k,l) = \frac{|\langle \mathbf{a}_k, \mathbf{a}_l \rangle|}{|\mathbf{a}_k||\mathbf{a}_l|}$$

$$\mu \sim 1/\sqrt{N}$$ (well separated columns)

CS methods

- ℓ_1-minimization/regularization

Basis Pursuit: $\min \|\mathbf{z}\|_1, \quad \|\mathbf{A}\mathbf{z} - \mathbf{y}\|_2 \leq \varepsilon$

Lasso: $\min \frac{1}{2}\|\mathbf{A}\mathbf{z} - \mathbf{y}\|_2^2 + \lambda\|\mathbf{z}\|_1$

$\lambda = \sigma\sqrt{2 \log M}, \quad \lambda = 0.5\sigma\sqrt{\log M}$

- Greedy algorithms:

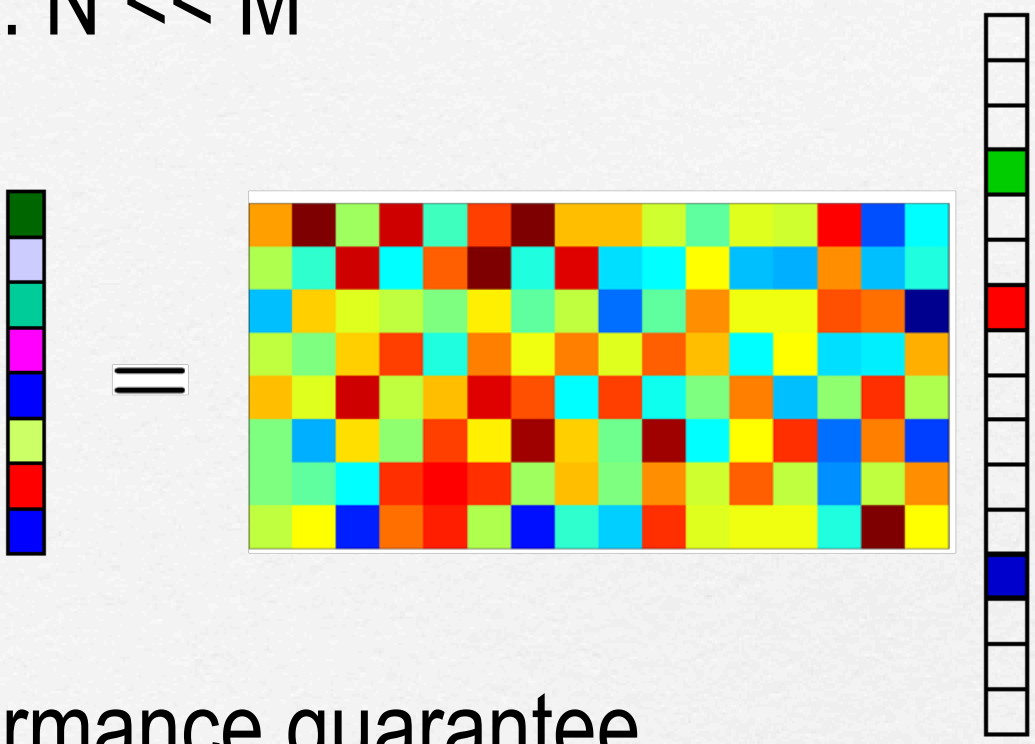
Orthogonal matching pursuit (OMP)
Subspace pursuit (SP)
Compressed sampling matching pursuit (Co-SaMP)
Iterative hard thresholding (IHT)
……. etc

CS benefits

- Sparse measurement: N << M

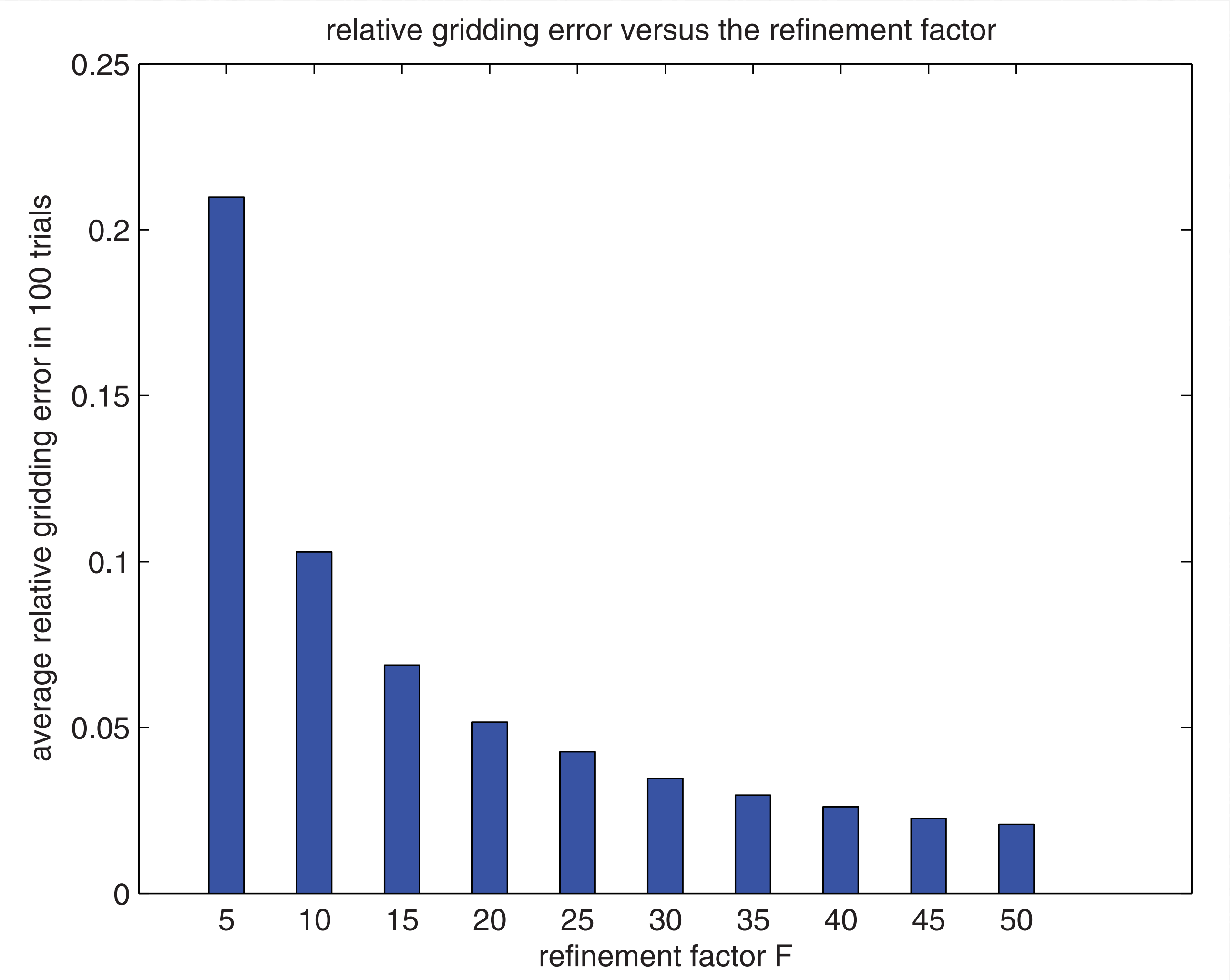
RIP: $N \sim s$

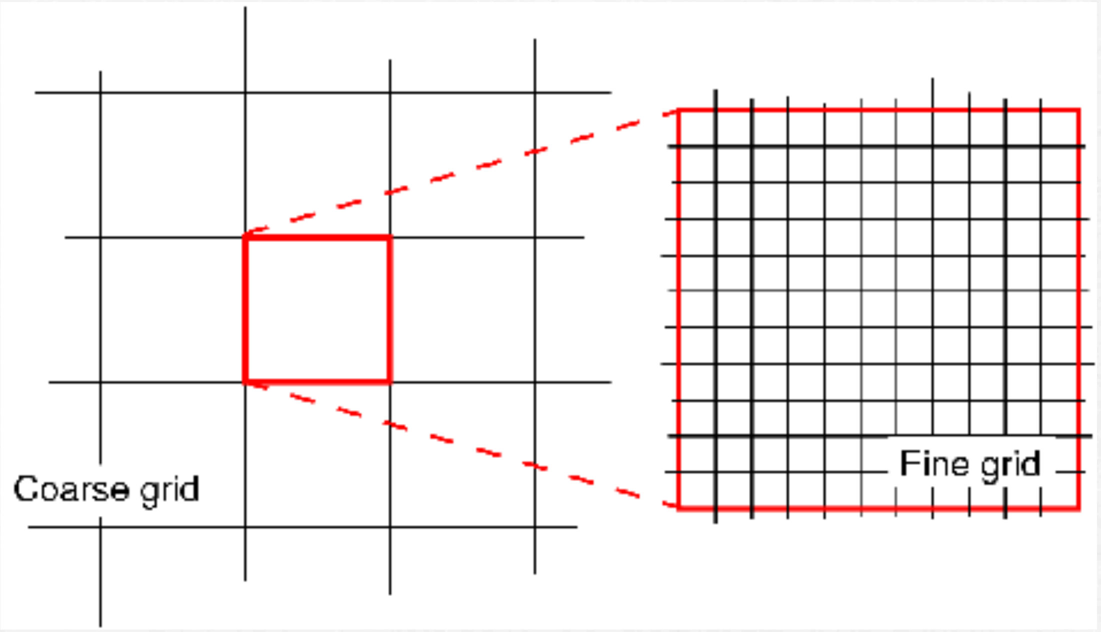
IP: $N \sim s^2$



- Non-asymptotic performance guarantee
- Effective algorithms
- **Caveat ?**

Gridding error

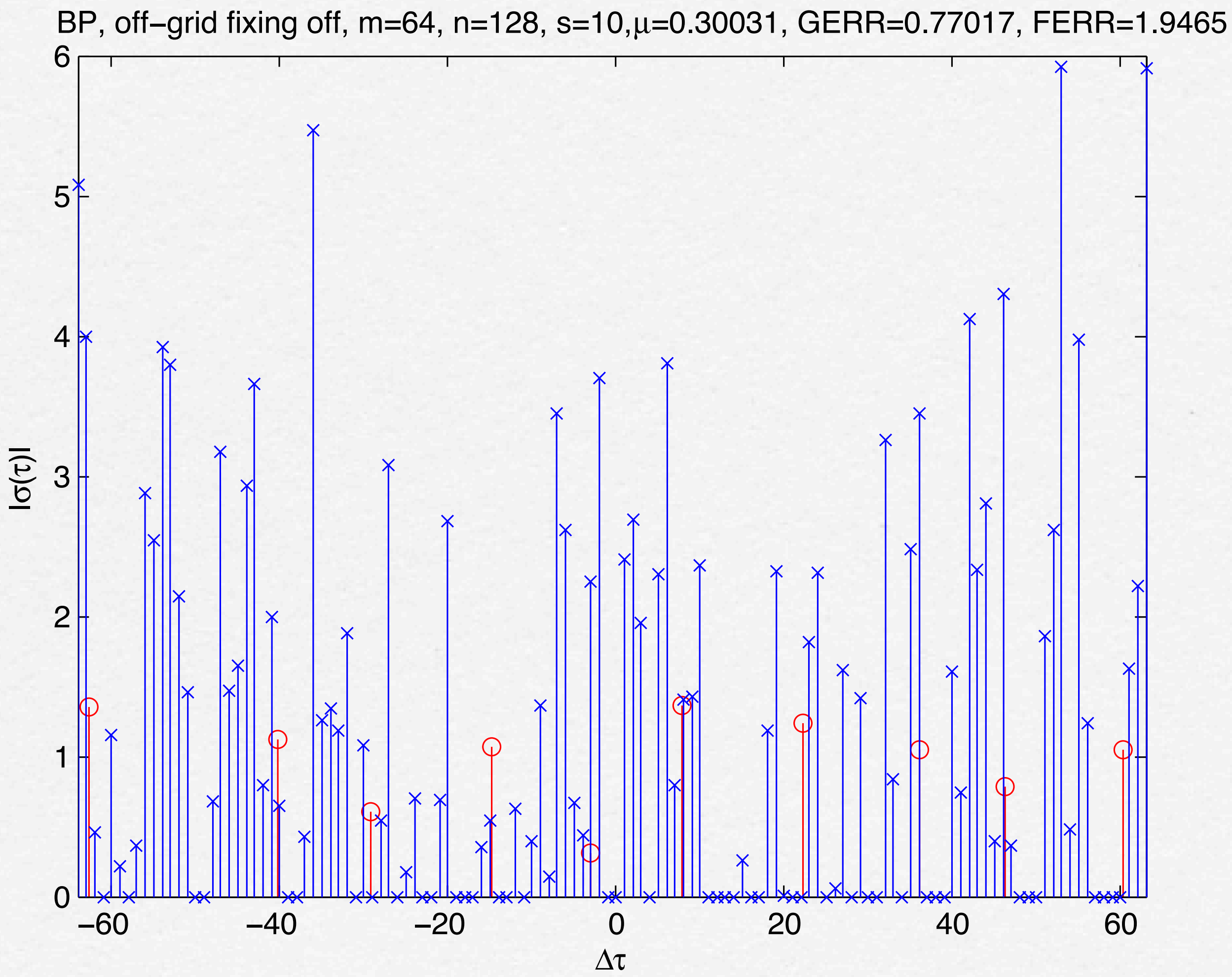




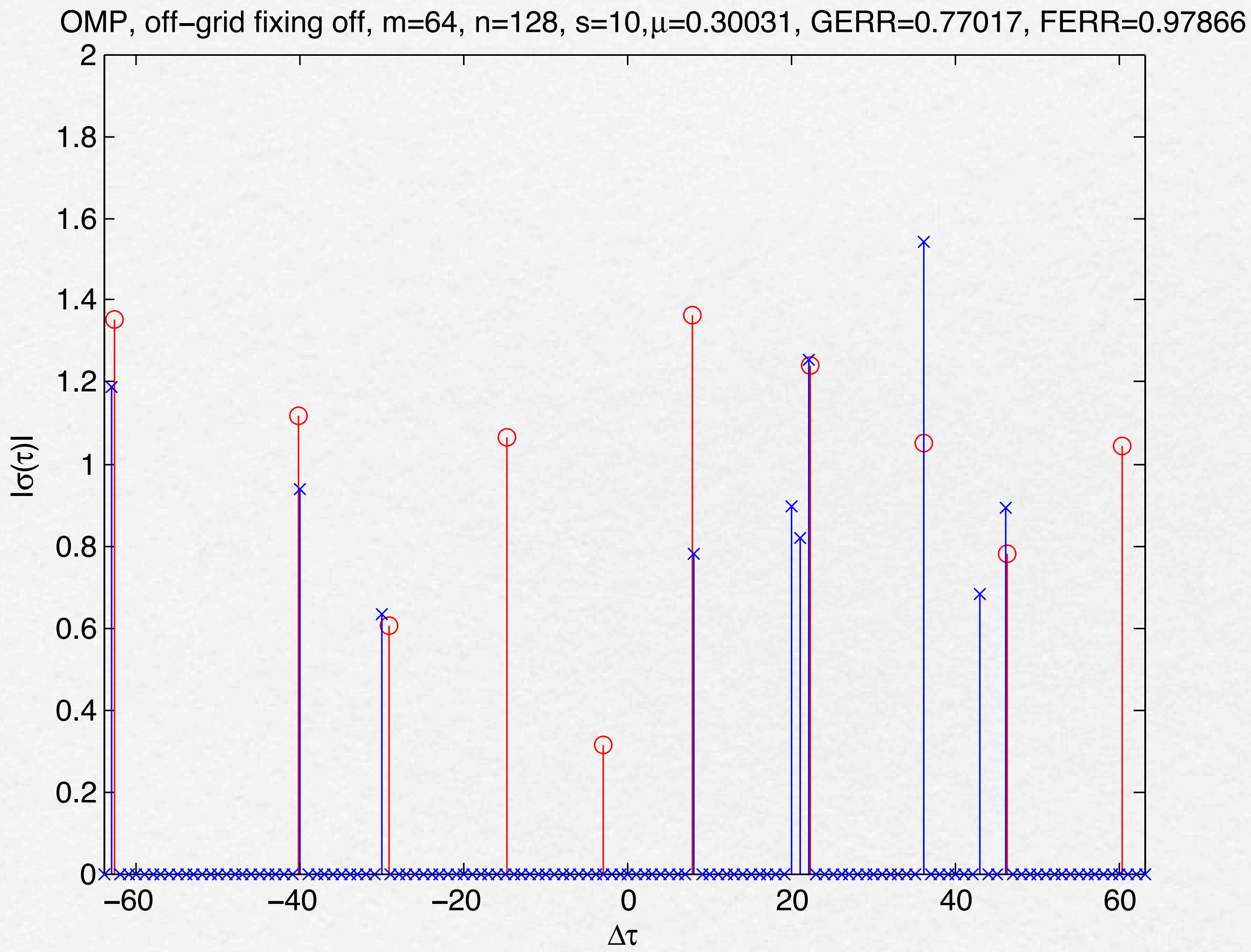
Gridding error inversely proportional to the refinement factor

F= coarse grid spacing / fine grid spacing

Peril of gridding error

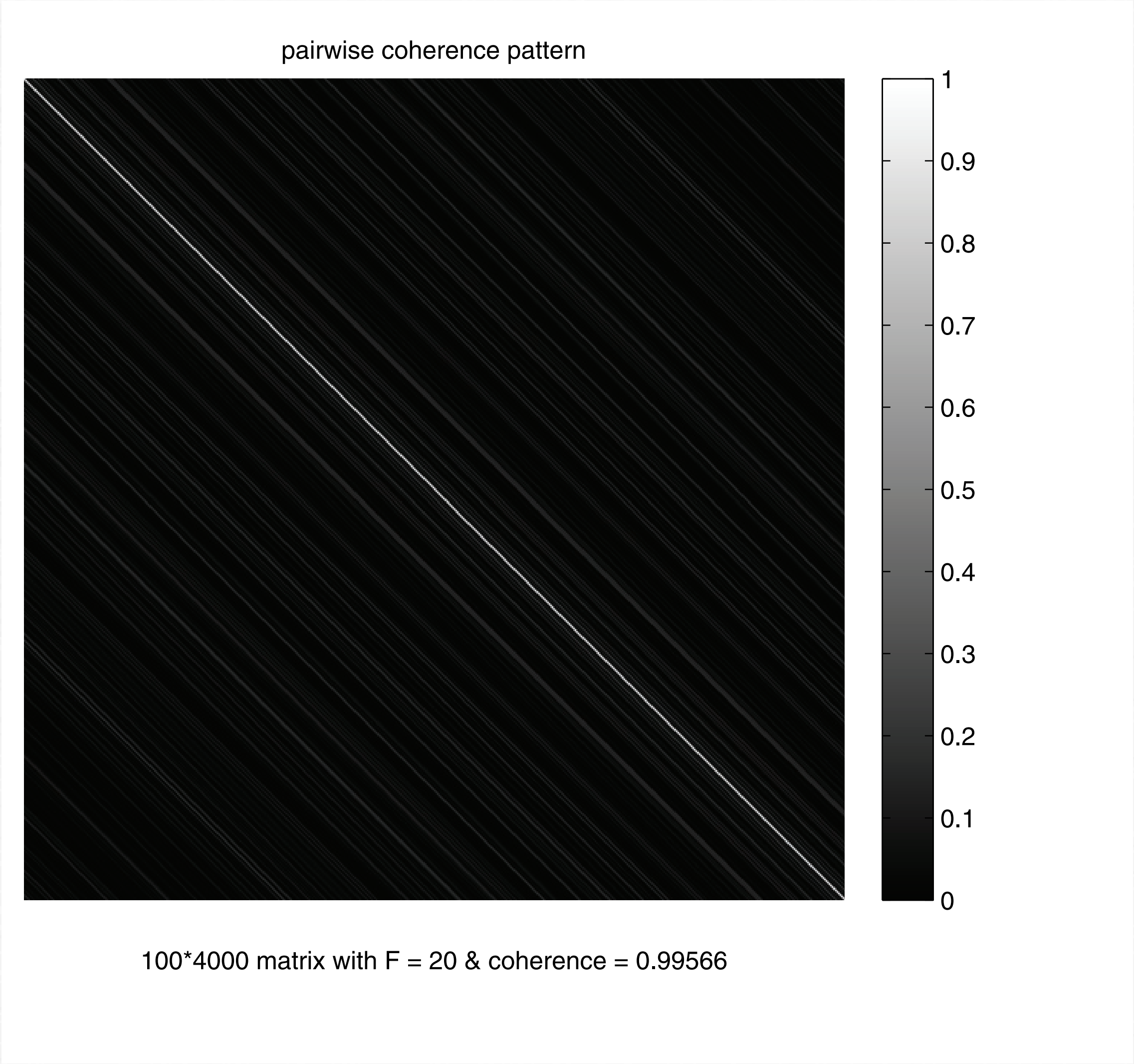


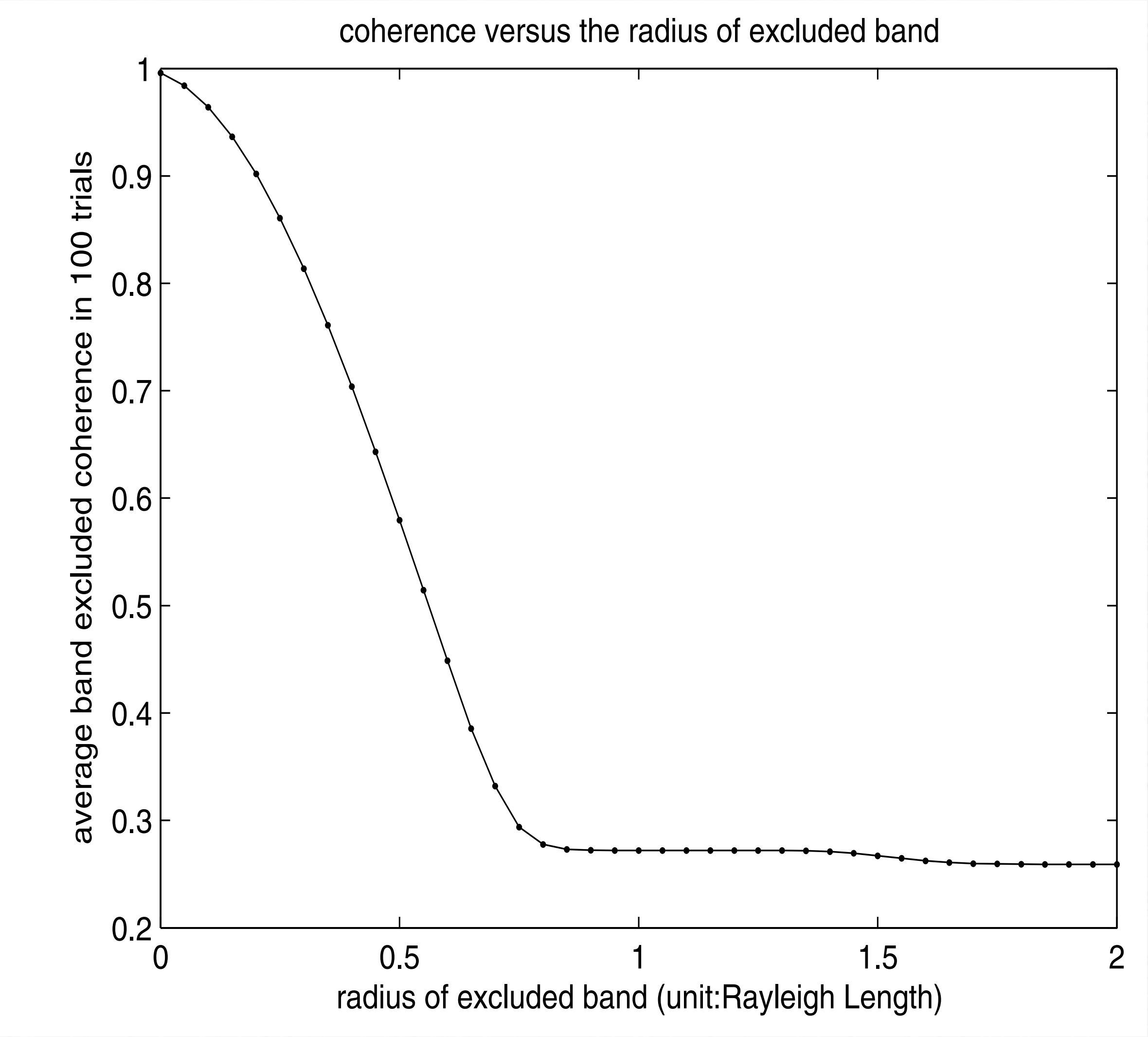
(a) BP



(b) OMP

Coherence pattern





Coherence Band

Let $\eta > 0$. Define the η-coherence band of the index k to be the set

$$B_\eta(k) = \{i \mid \mu(i,k) > \eta\},$$

and the η-coherence band of the index set S to be the set

$$B_\eta(S) = \cup_{k \in S} B_\eta(k).$$

Due to the symmetry $\mu(i,k) = \mu(k,i)$, $i \in B_\eta(k)$ if and only if $k \in B_\eta(i)$.

Denote

$$\begin{aligned} B_\eta^{(2)}(k) &\equiv B_\eta(B_\eta(k)) = \cup_{j \in B_\eta(k)} B_\eta(j) \\ B_\eta^{(2)}(S) &\equiv B_\eta(B_\eta(S)) = \cup_{k \in S} B_\eta^{(2)}(k). \end{aligned}$$

Greedy pursuit

Algorithm Band-excluding Matched Thresholding (BMT)

Input: $\mathbf{A}, \mathbf{y}, \eta > 0$.
Initialization: $S^0 = \emptyset$.
Iteration: For $k = 1, ..., s$,
  1) $i_k = \arg\max_j |\langle \mathbf{y}, \mathbf{a}_j \rangle|, j \notin B_\eta^{(2)}(S^{k-1})$.
  2) $S^k = S^{k-1} \cup \{i_k\}$
Output $\hat{\mathbf{x}} = \arg\min_{\mathbf{z}} \|\mathbf{Az} - \mathbf{y}\|_2$ s.t. $\text{supp}(\mathbf{z}) \subseteq S^s$

BMT performance guarantee

Theorem (F&Liao) Suppose that

$$B_\eta(i) \cap B_\eta(j) = \emptyset, \quad \forall i, j \in \text{supp}(\mathbf{x})$$

and that

$$\eta(2s-1)\frac{x_{\max}}{x_{\min}} + \frac{2\|\mathbf{e}\|_2}{x_{\min}} < 1$$

where

$$x_{\max} = \max_k |x_k|, \quad x_{\min} = \min_k |x_k|.$$

Let $\hat{\mathbf{x}}$ be the BMT reconstruction. Then $\text{supp}(\hat{\mathbf{x}}) \subseteq B_\eta(\text{supp}(\mathbf{x}))$ and moreover every nonzero component of $\hat{\mathbf{x}}$ is in the η-coherence band of a unique nonzero component of $\mathbf{x}$.

Theoretical resolution 2ℓ. Independent of grid refinement!

Compression: for moderate SNR

$$\eta \sim \frac{1}{\sqrt{N}}, \quad N \sim s^2 x_{\max}^2 / x_{\min}^2$$

Sketch of proof

Let supp($\mathbf{x}$) $= \{J_1, \ldots, J_s\}$. Let $J_{\max} \in$ supp($\mathbf{x}$) be the index of the largest component of $\mathbf{x}$ in absolute value.

On the one hand, for $k = 1, ..., s$,

$$\begin{aligned} |\mathbf{y}^\star \mathbf{a}_k| &= |\mathbf{x}_1 \mathbf{a}_1^\star \mathbf{a}_k + ... + \mathbf{x}_{k-1} \mathbf{a}_{k-1}^\star \mathbf{a}_k + x_k + \mathbf{x}_{k+1} \mathbf{a}_{k+1}^\star \mathbf{a}_k + \\ &\quad ... + \mathbf{x}_s \mathbf{a}_s^\star \mathbf{a}_k + \mathbf{e}^\star \mathbf{a}_k| \\ &\geq x_{\min} - (s-1)\eta x_{\max} - \|\mathbf{e}\|_2. \end{aligned}$$

On the other hand, $\forall l \notin B_\eta$(supp($\mathbf{x}$)),

$$\begin{aligned} |\mathbf{y}^\star \mathbf{a}_l| &= |\mathbf{x}_1 \mathbf{a}_1^\star \mathbf{a}_l + \mathbf{x}_2 \mathbf{a}_2^\star \mathbf{a}_l + ... + \mathbf{x}_s \mathbf{a}_s^\star \mathbf{a}_l + \mathbf{e}^\star \mathbf{a}_l| \\ &\leq x_{\max} s \eta + \|\mathbf{e}\|_2. \end{aligned}$$

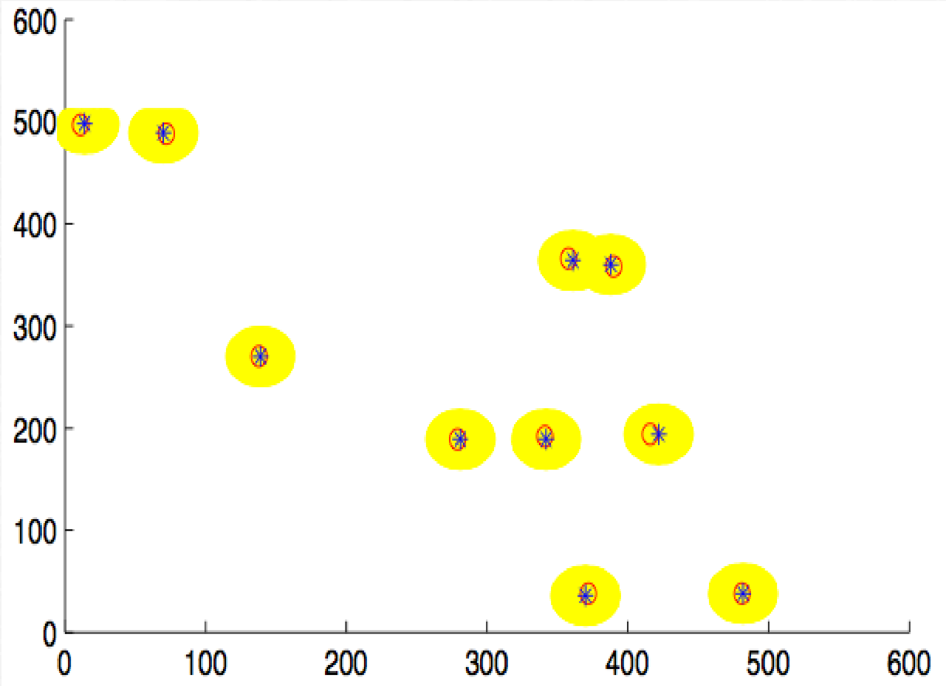
Band-excluding OMP (BOMP)

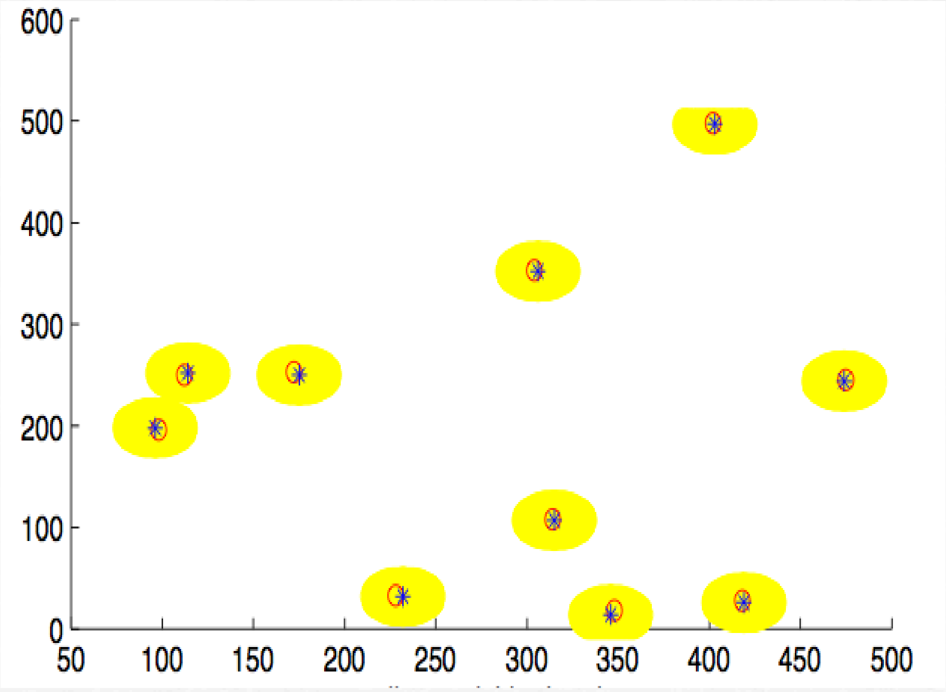
Algorithm 1. BOMP

Input: $\mathbf{A}, \mathbf{y}, \eta > 0$
Initialization: $\mathbf{x}^0 = 0, \mathbf{r}^0 = \mathbf{y}$ and $S^0 = \emptyset$
Iteration: For $n = 1, ..., s$

1) $i_{\max} = \arg\max_i |\langle \mathbf{r}^{n-1}, \mathbf{a}_i \rangle|, i \notin B_\eta^{(2)}(S^{n-1})$
2) $S^n = S^{n-1} \cup \{i_{\max}\}$
3) $\mathbf{x}^n = \arg\min_{\mathbf{z}} \|\mathbf{Az} - \mathbf{y}\|_2$ s.t. $\text{supp}(\mathbf{z}) \in S^n$
4) $\mathbf{r}^n = \mathbf{y} - \mathbf{Ax}^n$

Output: $\mathbf{x}^s$.





Two-dimensional case

BOMP performance guarantee

Theorem (F&Liao) Suppose that

$$B_\eta(i) \cap B_\eta^{(2)}(j) = \emptyset, \quad \forall i, j \in \mathsf{supp}(\mathbf{x})$$

and that

$$(5s-4) \cdot \eta \cdot \frac{x_{\max}}{x_{\min}} + \frac{5}{2} \cdot \frac{\|\mathbf{e}\|_2}{x_{\min}} < 1$$

where

$$x_{\max} = \max_k |x_k|, \quad x_{\min} = \min_k |x_k|.$$

Let $\hat{\mathbf{x}}$ be the BOMP reconstruction. Then $\mathsf{supp}(\hat{\mathbf{x}}) \subseteq B_\eta(\mathsf{supp}(\mathbf{x}))$ and moreover every nonzero component of $\hat{\mathbf{x}}$ is in the η-coherence band of a unique nonzero component of $\mathbf{x}$.

`Theoretical resolution` 3ℓ. `Numerical resolution` $\sim 1\ell$.

Local optimization (LO)

Algorithm 2. Local Optimization (LO)

Input: $\mathbf{A}, \mathbf{y}, \eta > 0, S^0 = \{i_1, \ldots, i_k\}$.
Iteration: For $n = 1, 2, ..., k$.
1) $\mathbf{x}^n = \arg\ \min_{\mathbf{z}} \|\mathbf{A}\mathbf{z} - \mathbf{y}\|_2, \quad \text{supp}(\mathbf{z}) = S^{n-1} \cup B_\eta(i_n)$.
2) $S^n = \text{supp}(\mathbf{x}^n)$.
Output: S^k.

Algorithm 3. BLOOMP

Input: $\mathbf{A}, \mathbf{y}, \eta > 0$
Initialization: $\mathbf{x}^0 = 0, \mathbf{r}^0 = \mathbf{y}$ and $S^0 = \emptyset$
Iteration: For $n = 1, ..., s$
1) $i_{\text{max}} = \arg\min_i |\langle \mathbf{r}^{n-1}, \mathbf{a}_i \rangle|, i \notin B_\eta^{(2)}(S^{n-1})$
2) $S^n = \text{LO}(S^{n-1} \cup \{i_{\text{max}}\})$.
3) $\mathbf{x}^n = \arg\min_{\mathbf{z}} \|\mathbf{A}\mathbf{z} - \mathbf{y}\|_2$ s.t. $\text{supp}(\mathbf{z}) \in S^n$
4) $\mathbf{r}^n = \mathbf{y} - \mathbf{A}\mathbf{x}^n$
Output: $\mathbf{x}^s$.

LO performance guarantee

Theorem (F& Liao) Let S^0 and S^k be the input and output, respectively, of the LO algorithm.

If

$$x_{\min} > (\varepsilon + 2(s-1)\eta)\left(\frac{1}{1-\eta} + \sqrt{\frac{1}{(1-\eta)^2} + \frac{1}{1-\eta^2}}\right)$$

and each element of S^0 is in the η-coherence band of a unique nonzero component of $\mathbf{x}$, then each element of S^k remains in the η-coherence band of a unique nonzero component of $\mathbf{x}$.

BLO-based CS-algorithms

BLO Subspace Pursuit (BLOSP)

BLO Co-SaMP (BLO-CoSaMP)

BLO Iterative Hard Thresholding (BLOIHT)

➡ BP-BLOT — Constrained L1-minimization

➡ Lasso-BLOT — Unconstrained L1-minimization

. . . .

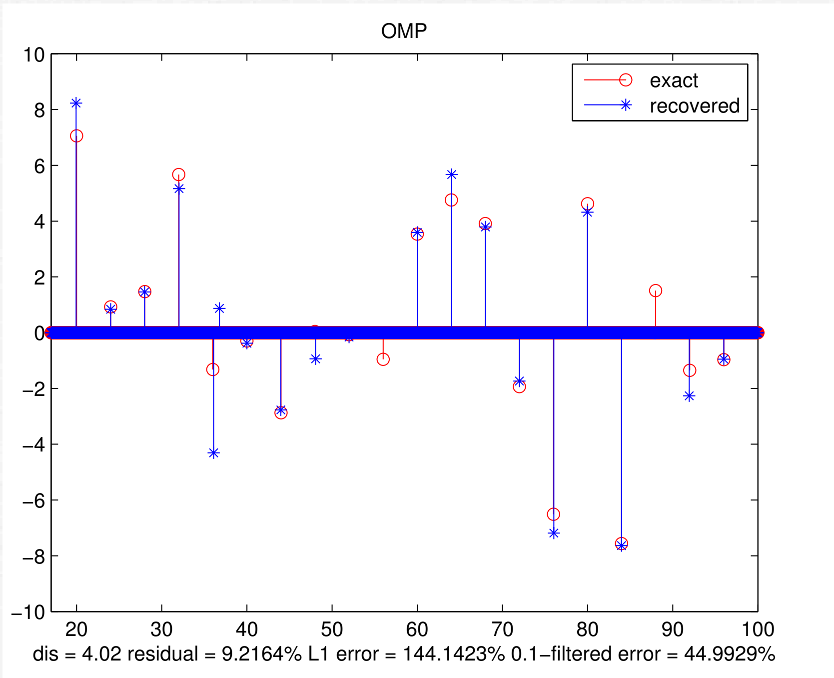
etc.

Comparison with BP

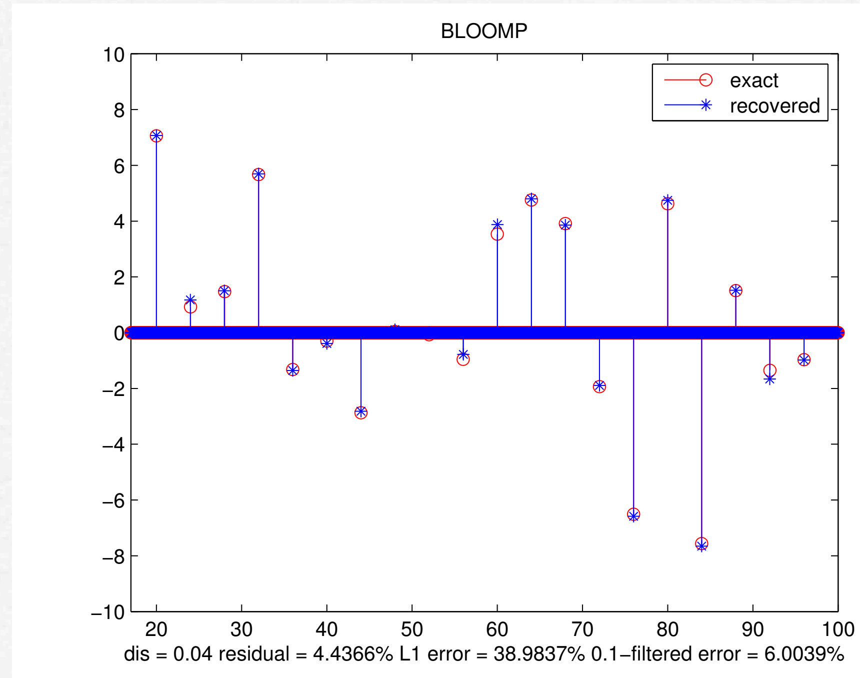
Candes & Fernandez-Granda 2012

$$\text{Error} \leq \text{Constant} \cdot F^2 \cdot \text{Noise}$$

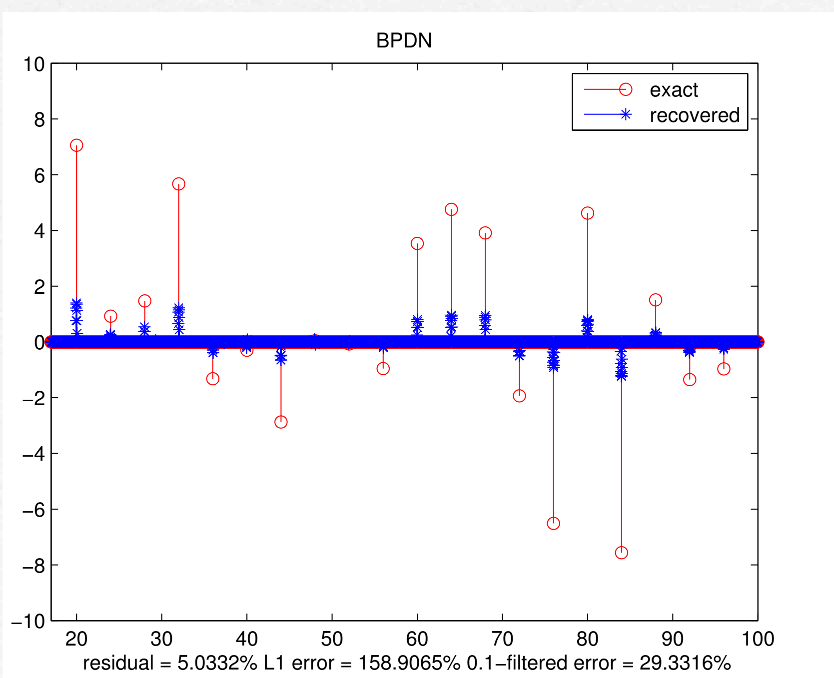
- Requires target separation of at least 4 Rayleigh lengths (RLs)
- Fourier measurement
- Error bound meaningful only with SNR >> 1
- Error > 80% at F = 20 (gridding error ~ 5%) independent of SNR



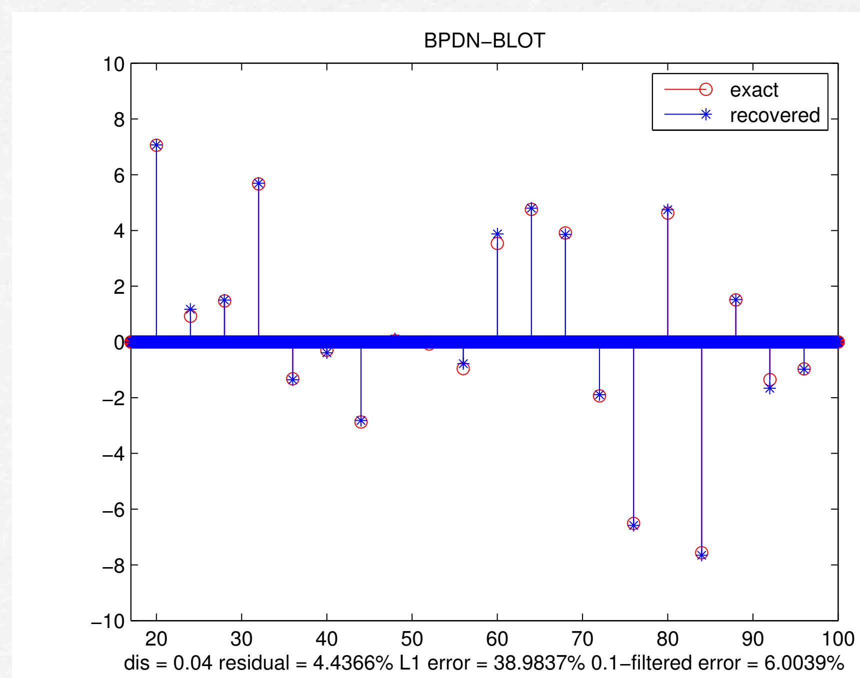
(a) OMP



(b) BLOOMP



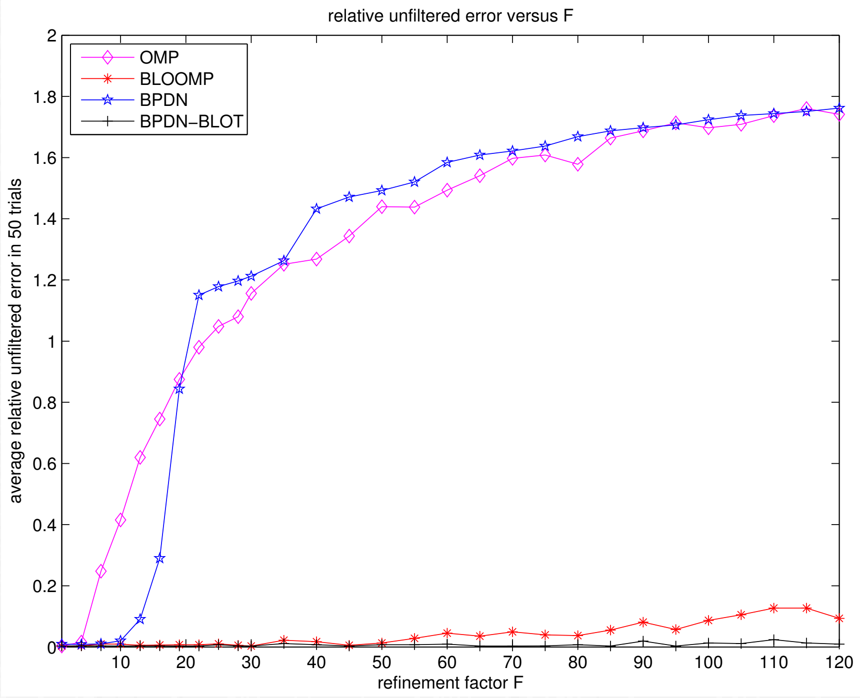
(c) BP



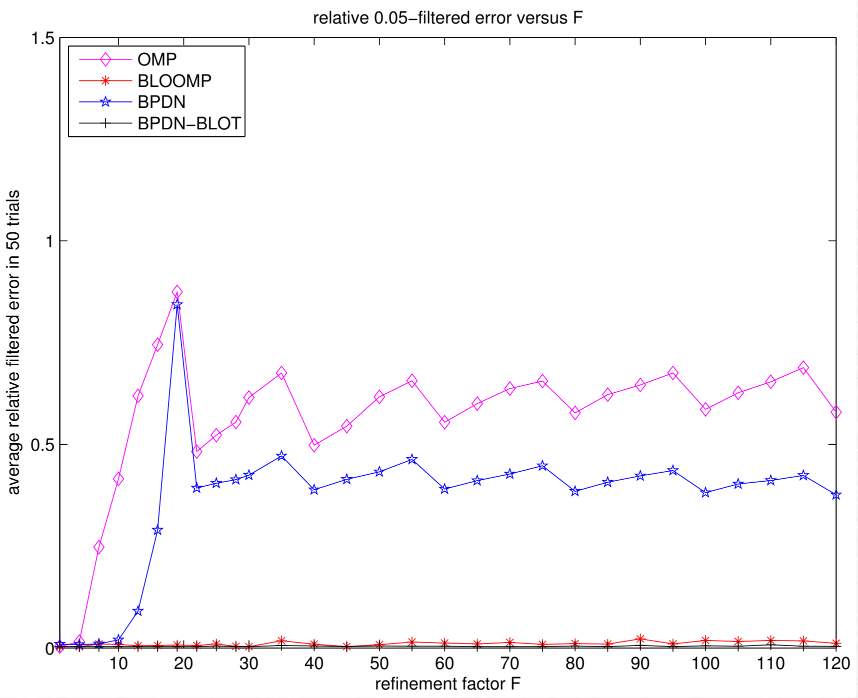
(d) BP-BLOT

- 20 random-phased,well-separated targets
- Real-parts shown
- F=50
- SNR=20
- Accuracy of BLOOMP & BP-BLOT is a few % RL

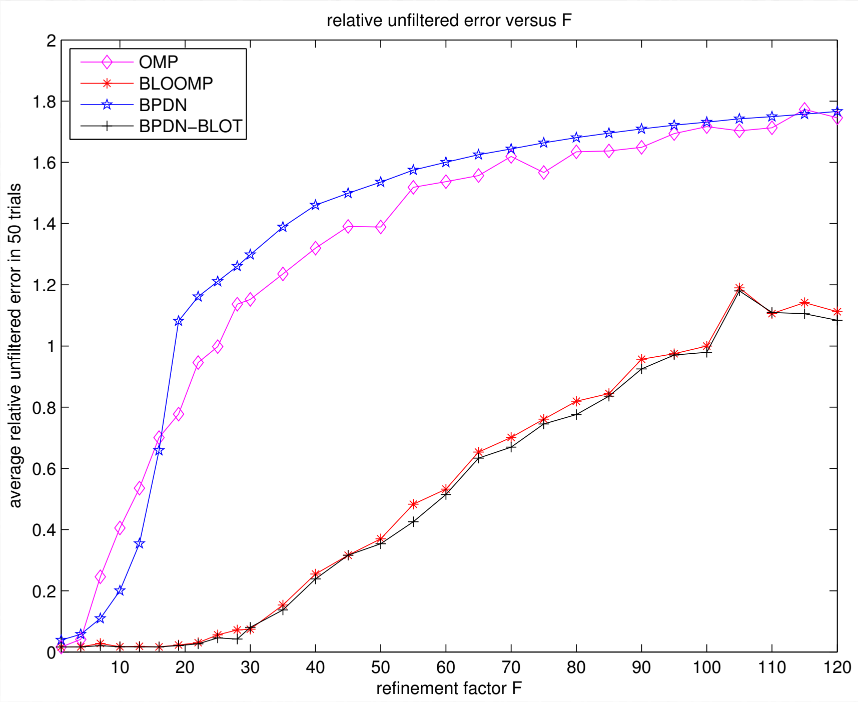
Relative error vs. Refinement factor



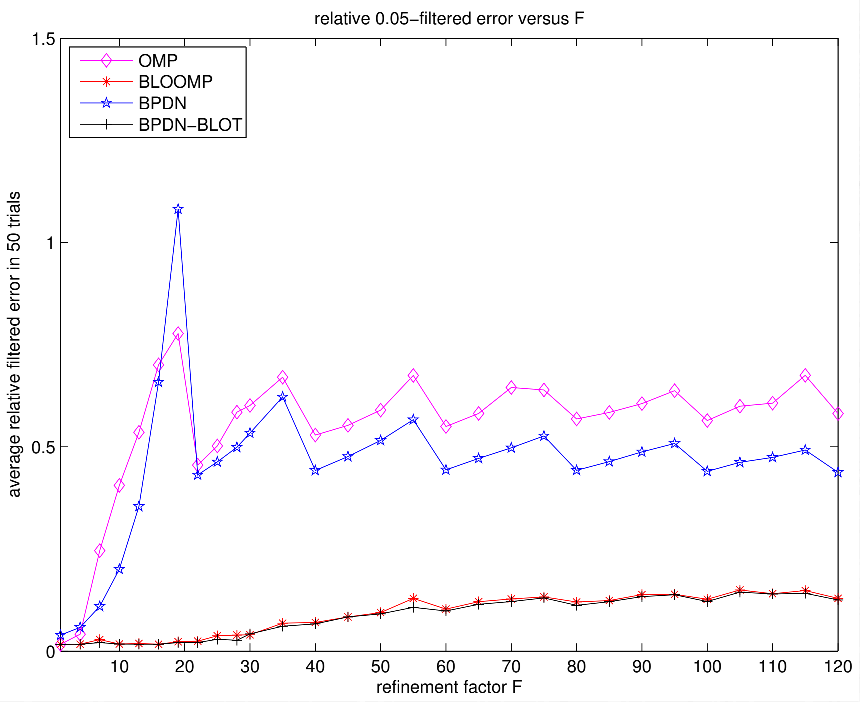
SNR=100, filter width = 0



SNR=100, filter width = 0.1



SNR=20, filter width = 0

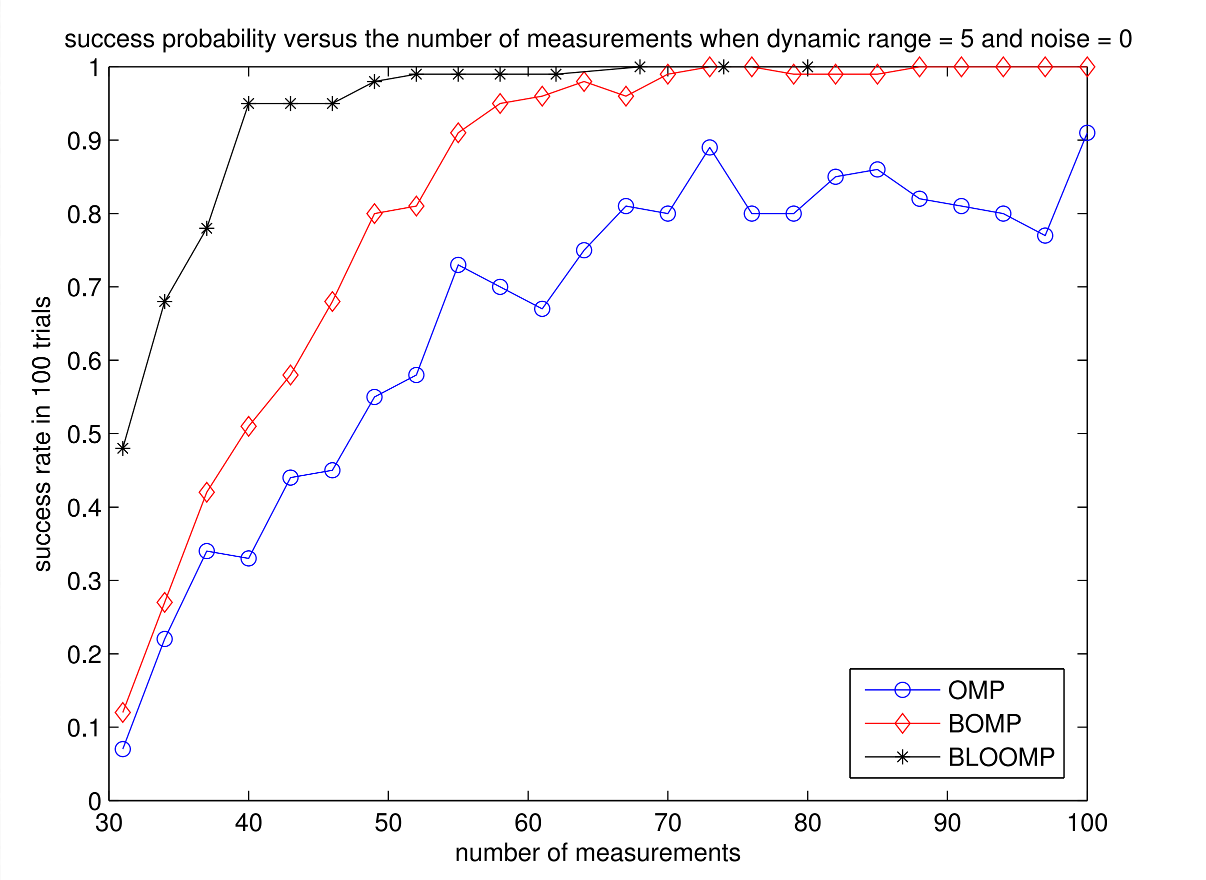


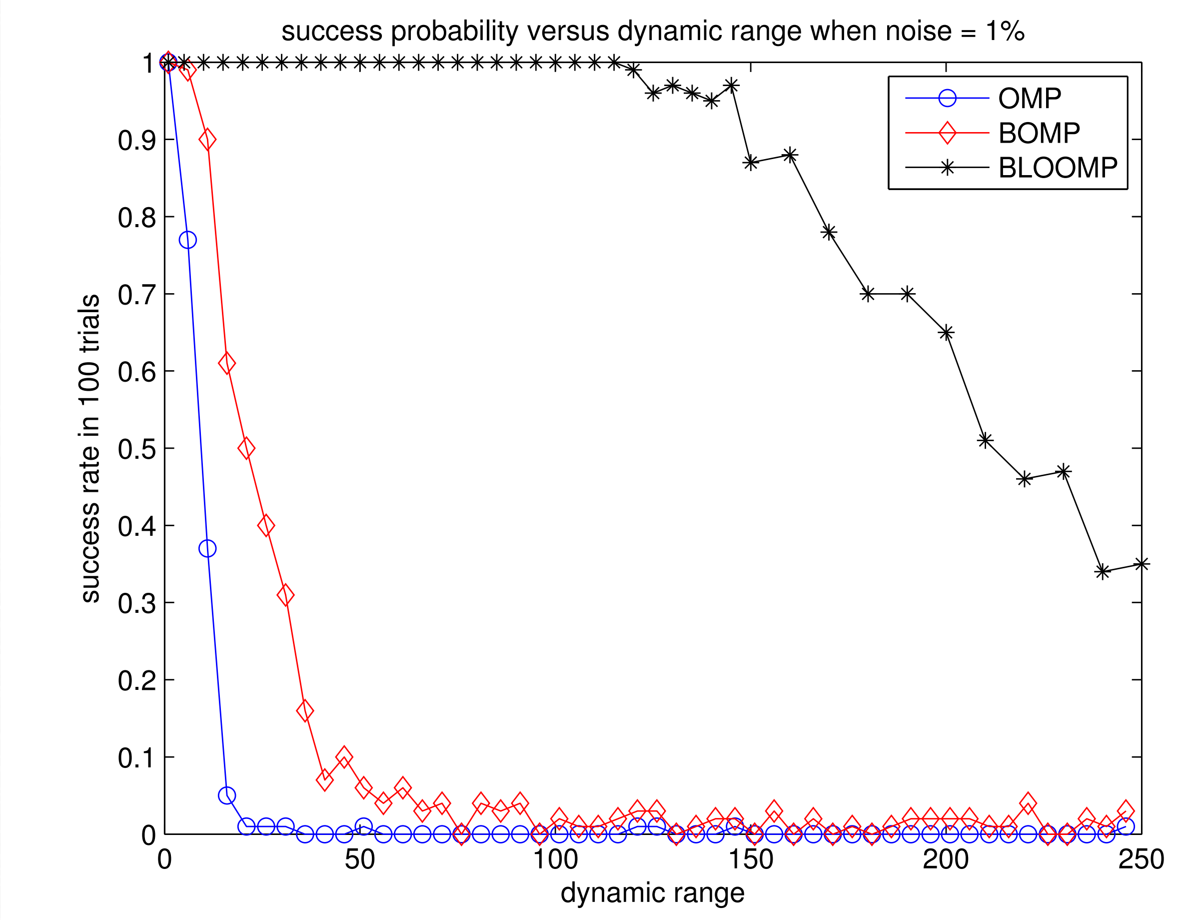
SNR=20, filter width = 0.1

For two subsets A and B in $\mathbb{R}^d$ of the same cardinality, the Bottleneck distance $d_B(A, B)$ is defined as

$$d_B(A, B) = \min_{f \in \mathcal{M}} \max_{a \in A} |a - f(a)|$$

where $\mathcal{M}$ is the collection of all one-to-one mappings from A to B.





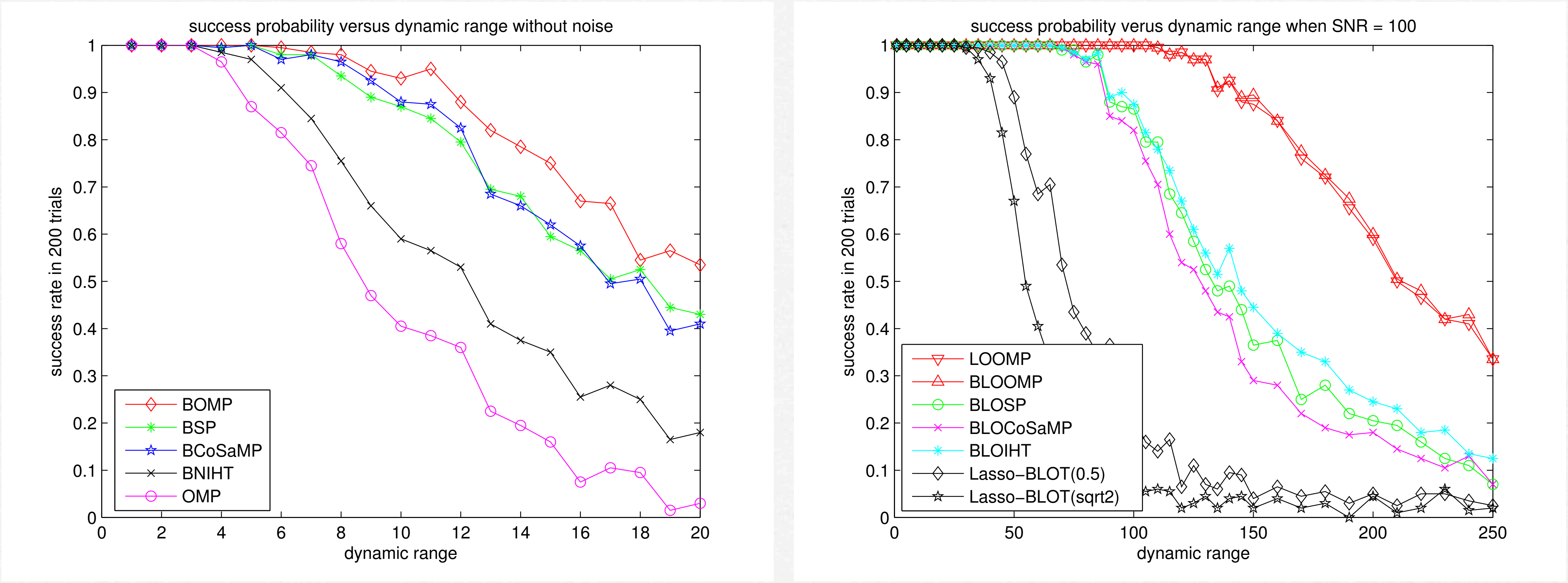
Success rate vs. number of data

Success rate vs. dynamic range

S=10

Performance vs. target range

S=10



LO dramatically improves the performance w.r.t. dynamic range

BLOOMP performs better than Lasso-BLOT

CS with highly redundant dictionary

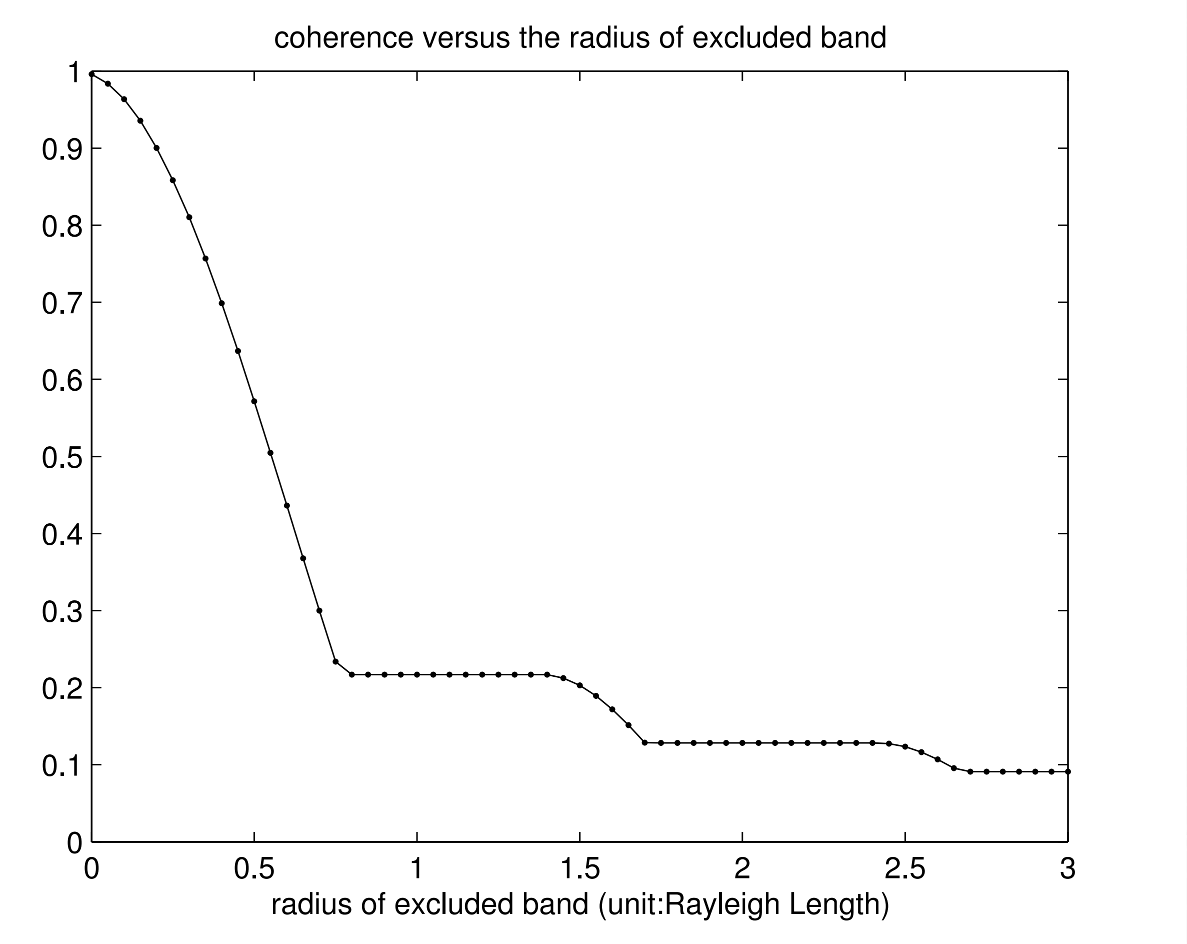
$$\mathbf{y} = \mathbf{\Phi x} + \mathbf{e} = \mathbf{\Phi D}\alpha + \mathbf{e}$$

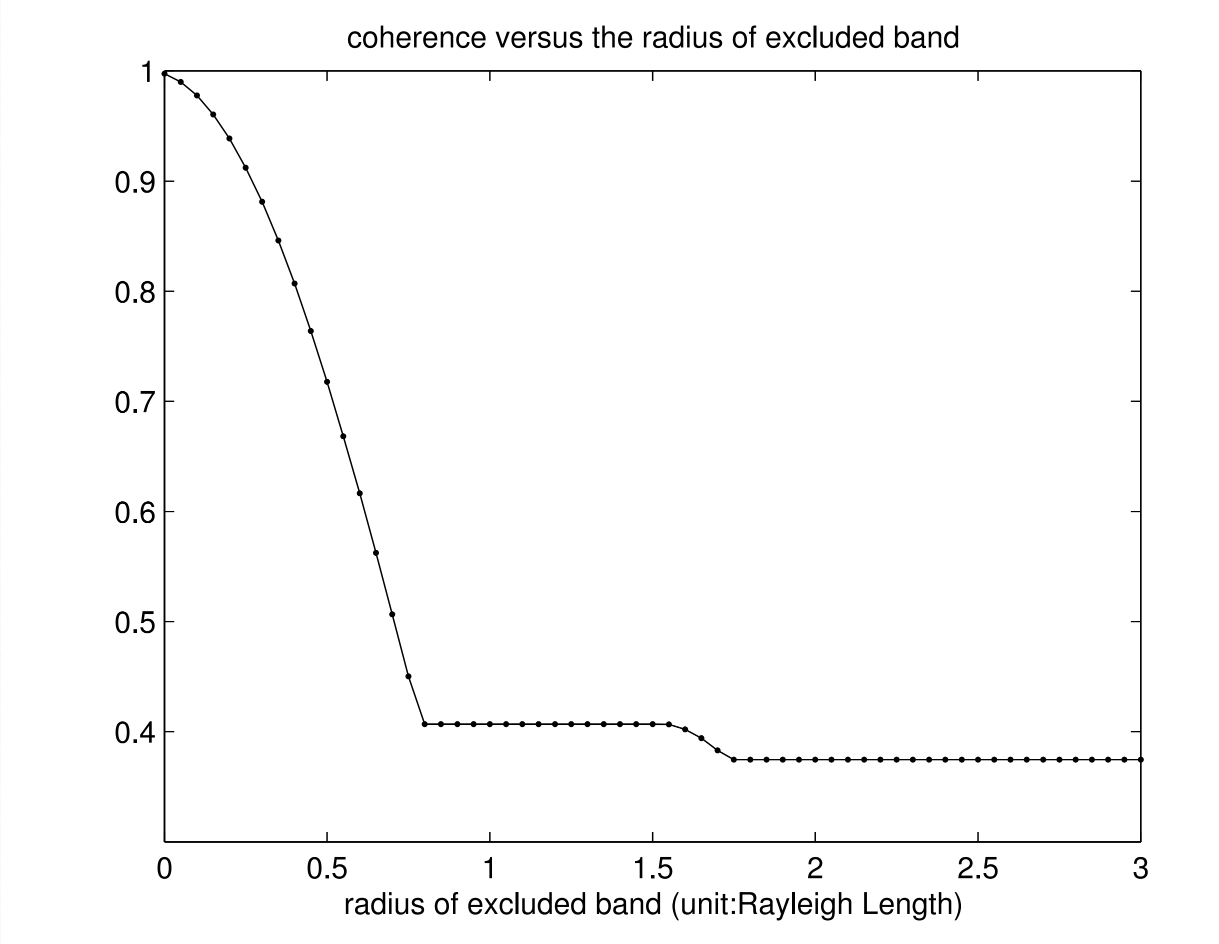
where $\mathbf{\Phi}$ is i.i.d. Gaussian matrix and $\mathbf{D}$ is an oversampled, redundant DFT frame.

Performance metric:

$$\frac{\|\mathbf{D}(\alpha - \hat{\alpha})\|}{\|\mathbf{D}\alpha\|}$$

Coherence pattern





- Coherence band of the dictionary
- Coherence band of the sensing matrix

Analysis approach: Frame-baed BP

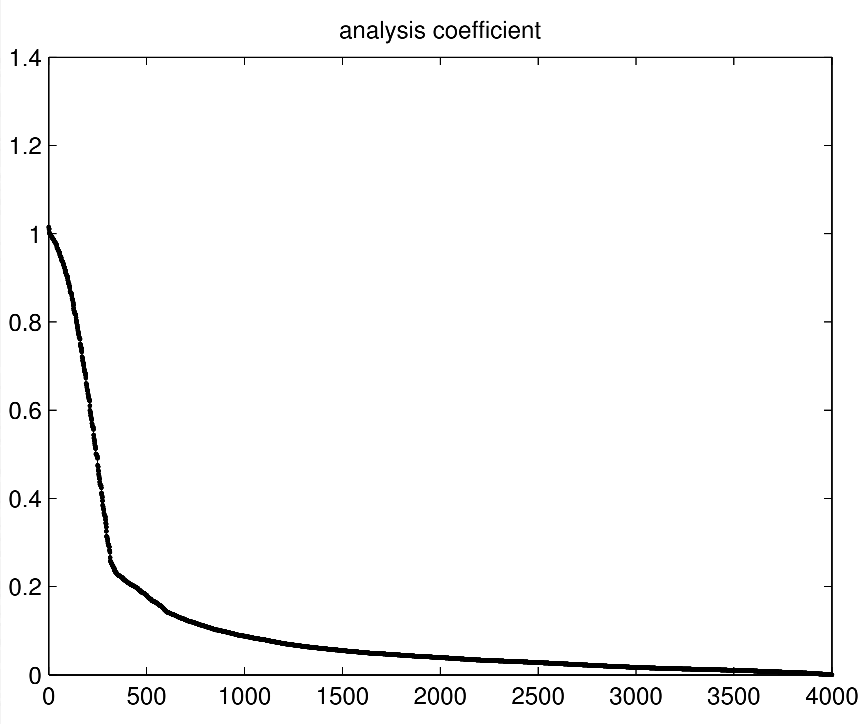
Candes-Eldar-Needel-Randal 2011

$$\min_{\mathbf{z}} \|\mathbf{D}^*\mathbf{z}\|_1, \quad \|\mathbf{A}\mathbf{z}-\mathbf{y}\|_2 \leq \varepsilon, \quad \mathbf{A} = \Phi\mathbf{D}$$

Assumptions: 1) Frame-adapted RIP

$$(1-\delta)\|\mathbf{D}\mathbf{z}\|_2^2 \leq \|\mathbf{A}\mathbf{z}\|_2^2 \leq (1+\delta)\|\mathbf{D}\mathbf{z}\|_2^2, \quad \|\mathbf{z}\|_0 \leq 2s$$

2) sparsity or compressibility of analysis coefficients



But, unless with a tight frame, analysis coefficients have long tail

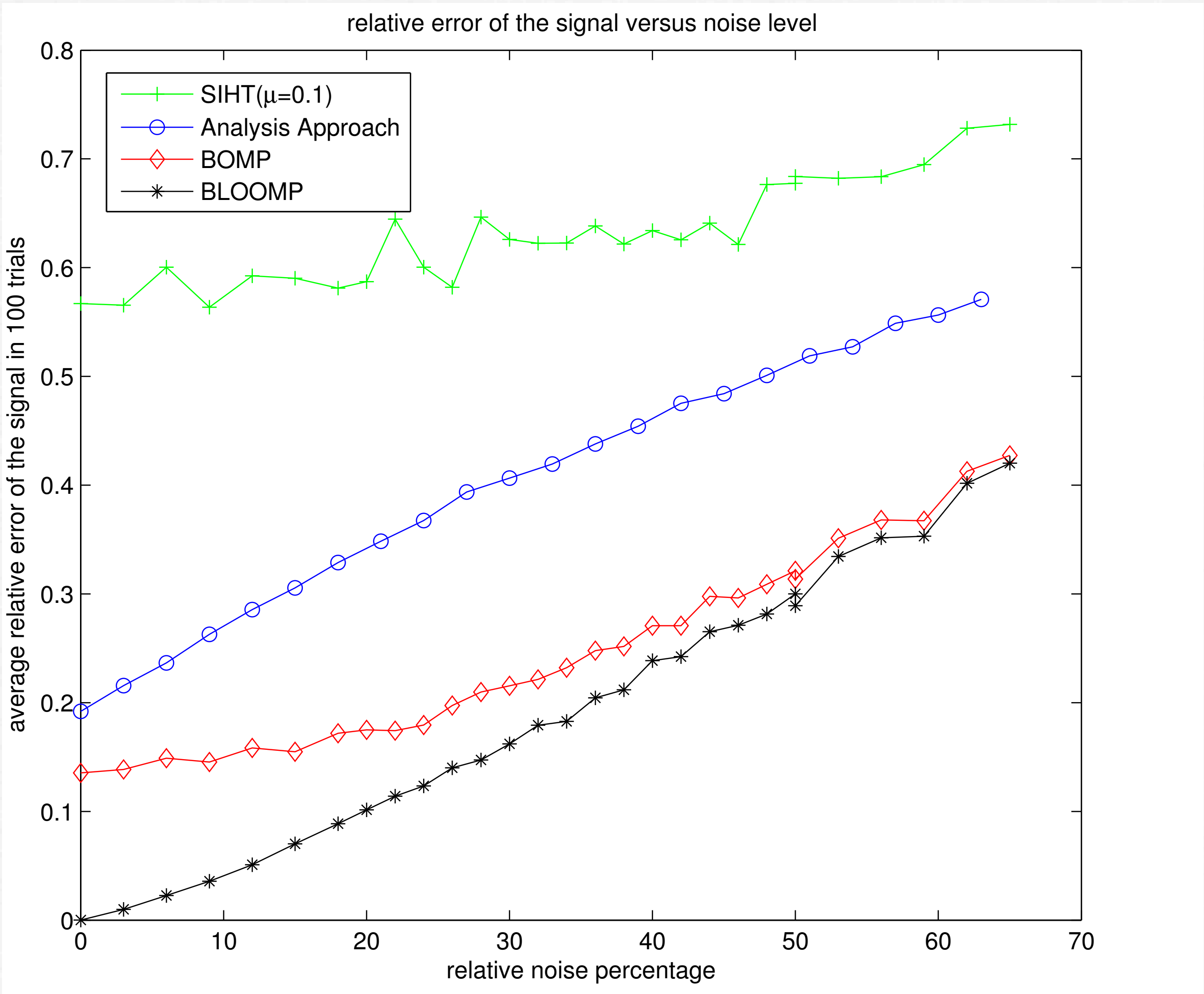
Analysis approach: Spectral CS

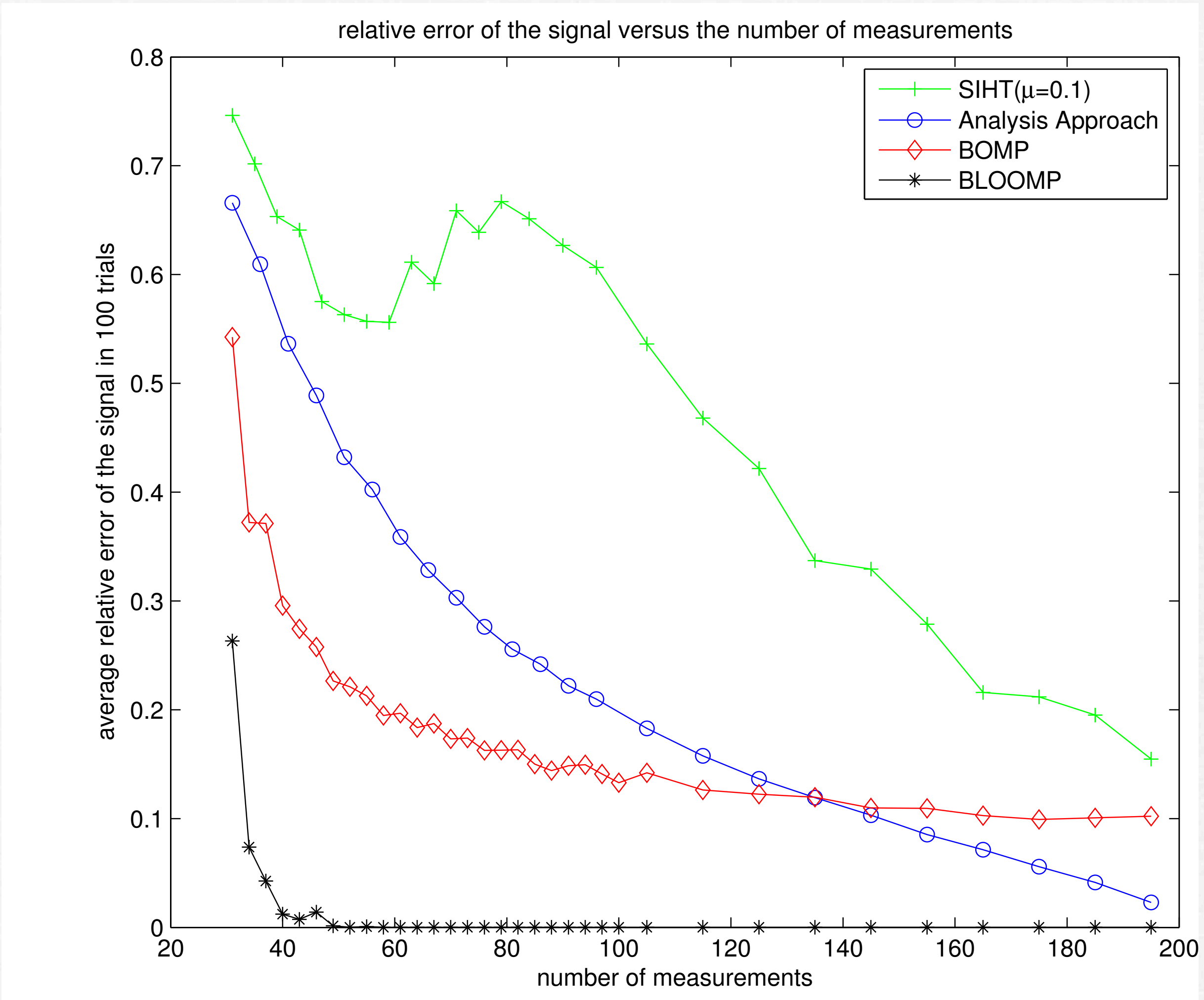
Duarte-Baraniuk 2012: model-based CS (SIHT)

IHT: $\mathbf{x}^{n+1} = T_s(\mathbf{x}^n + \mathbf{\Phi}^*(\mathbf{y} - \mathbf{\Phi}\mathbf{x}^n))$

Coherence-inhibiting structured sparse approximation is implemented by the heuristics of selecting the s largest, well separated **analysis coefficients**.

Comparison: analysis vs. synthesis





Error vs. percentage noise

Error vs. number of data

S=10, Target range = 10

Super-resolution w. Fourier measurement

Donoho 1992: optimal recovery theory, no explicit algorithm

$$\text{Error} \leq \text{Constant} \cdot F^{\alpha} \cdot \text{Noise}, \quad \alpha \leq 2R+1$$

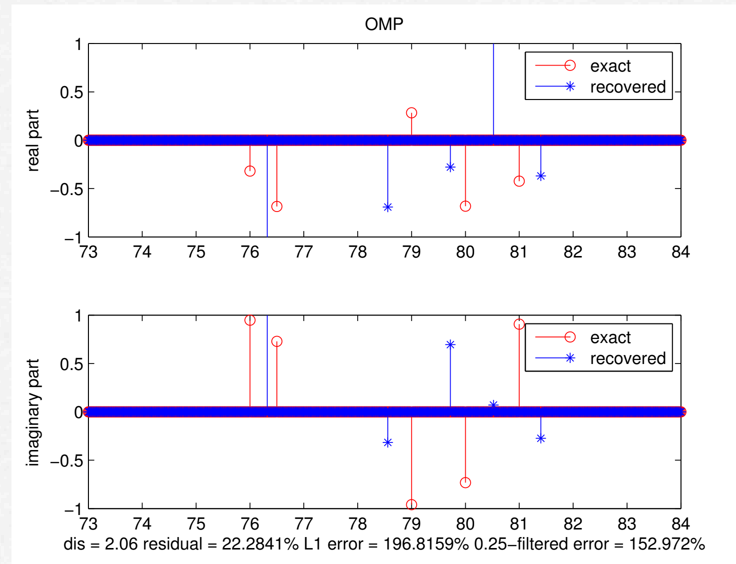
Rayleigh index $R(S) = \min\left\{ r : r \geq \sup_t \#(S \cap [t, t+4\ell r)) \right\}$

Rayleigh index of a set S is at most r if every interval of length $4\ell r$ contains at most r points in S. $R(S)$ is the smallest of such r.

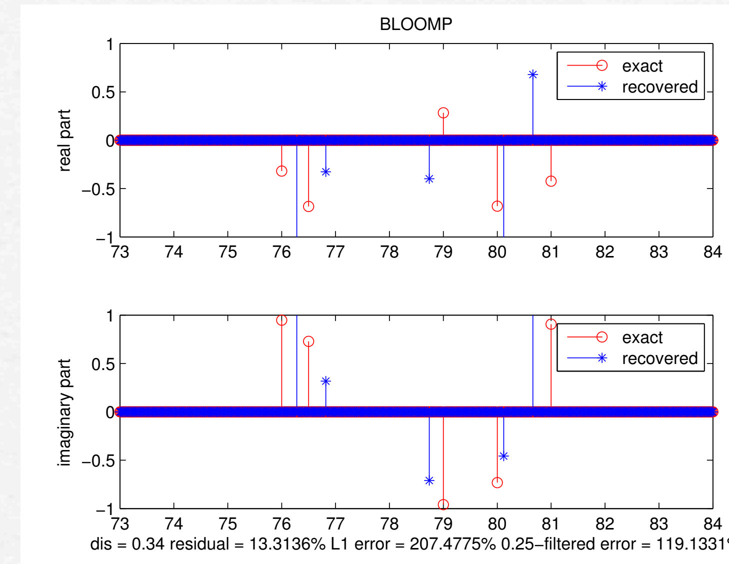
Candes & Fernandez-Granda 2012 (BP) ⟶ R=1

F& Liao: 2 (BMT) or 3 (BOMP, BLOOMP) RLs ⟶ R= any

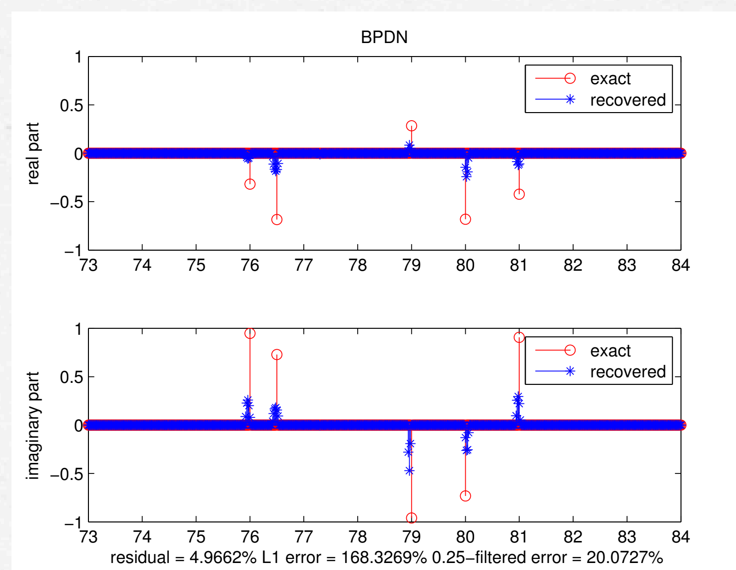
- Use prior info as constraint in search



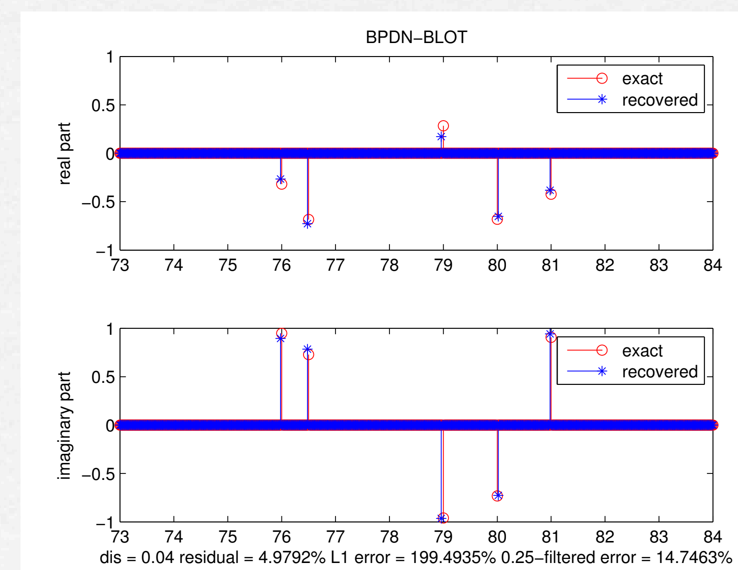
(a) OMP



(b) BLOOMP



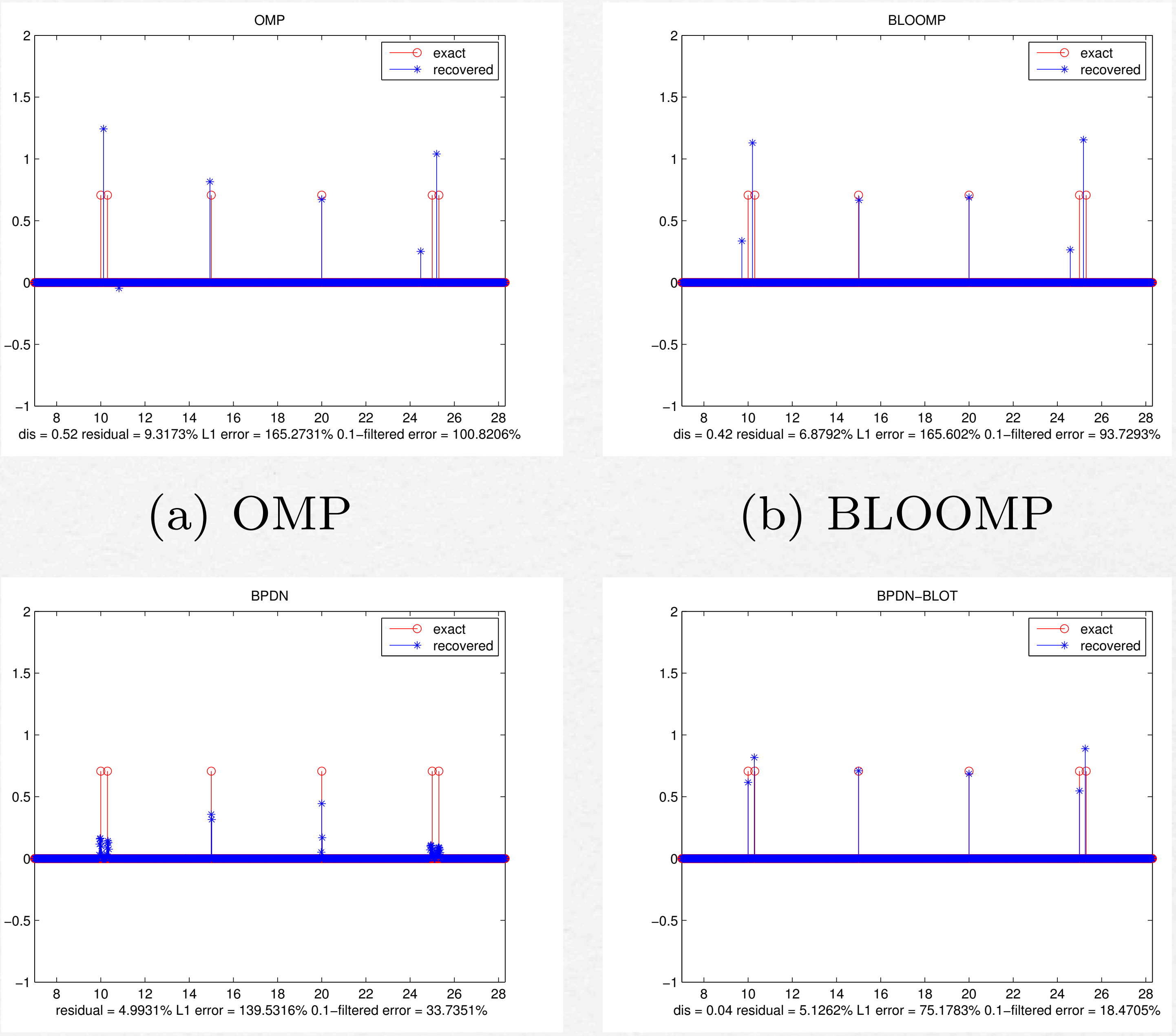
(c) BP



(d) BP-BLOT

5 random-phased spikes at ***76, 76.5***, 79, 80, 81 (R=5) with F=50, SNR=20

- Use prior info as constraint in search



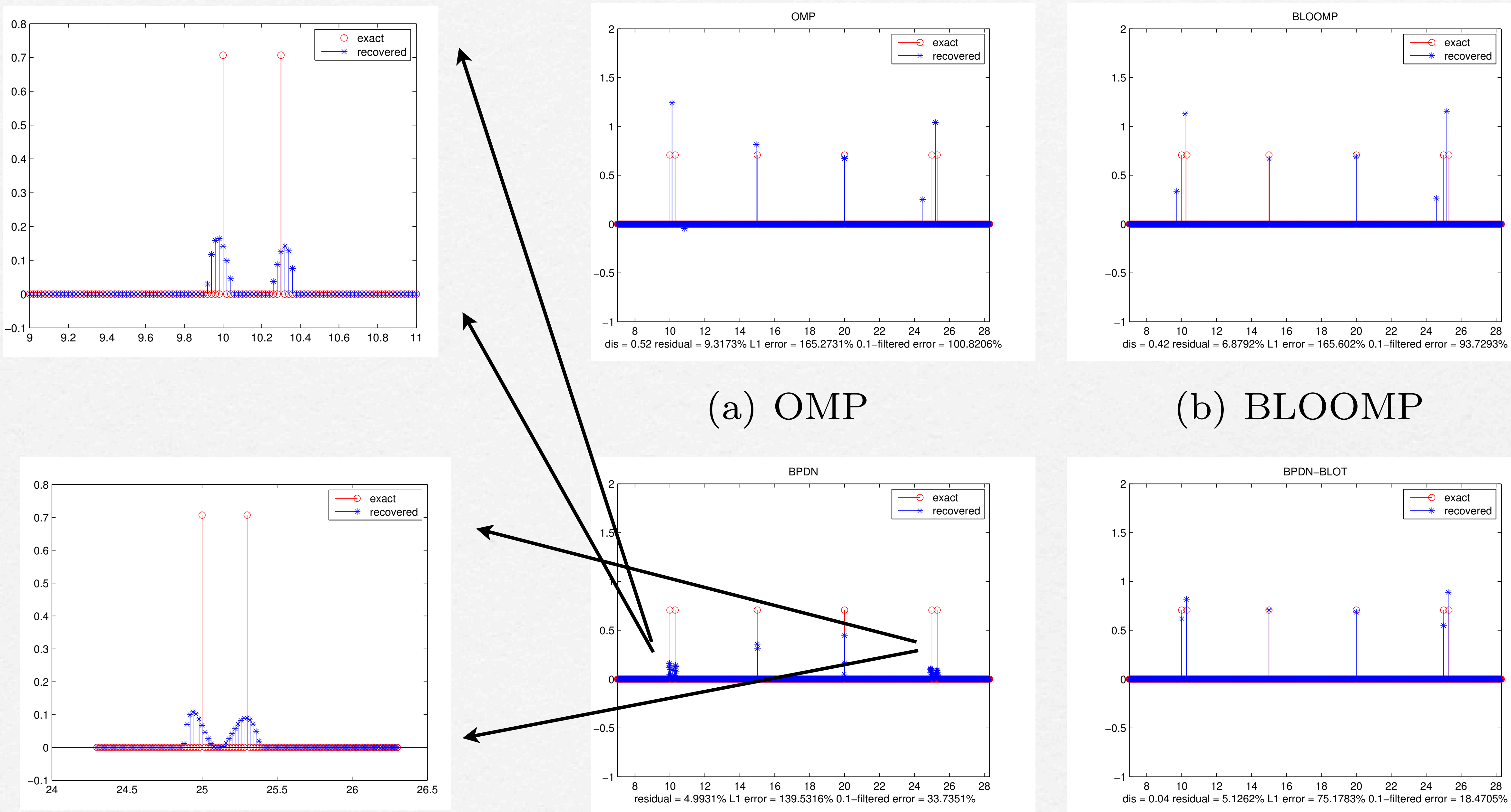
(a) OMP

(b) BLOOMP

(c) BP

(d) BP-BLOT

5 random-phased spikes (real part) at ***10, 10.3***, 15, 20, ***25, 25.3*** (R=6) with F=50, SNR=20



(a) OMP (b) BLOOMP

(c) BP (d) BP-BLOT

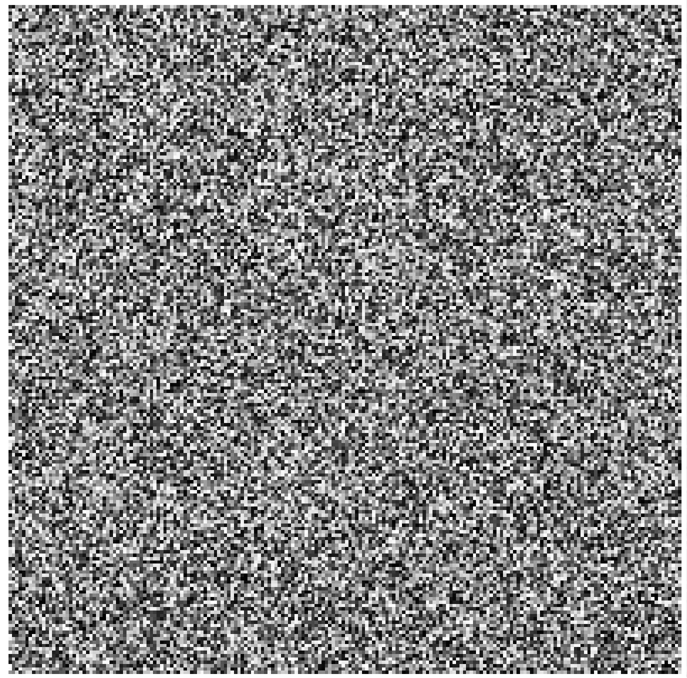
5 random-phased spikes (real part) at ***10, 10.3***, 15, 20, ***25, 25.3*** (R=6) with F=50, SNR=20

High resolution illumination

F=10, OR=2, # random shifts=***50***

data = 2 # pixels

Algorithm: alternating projection



HRI

200 x 200 phantom

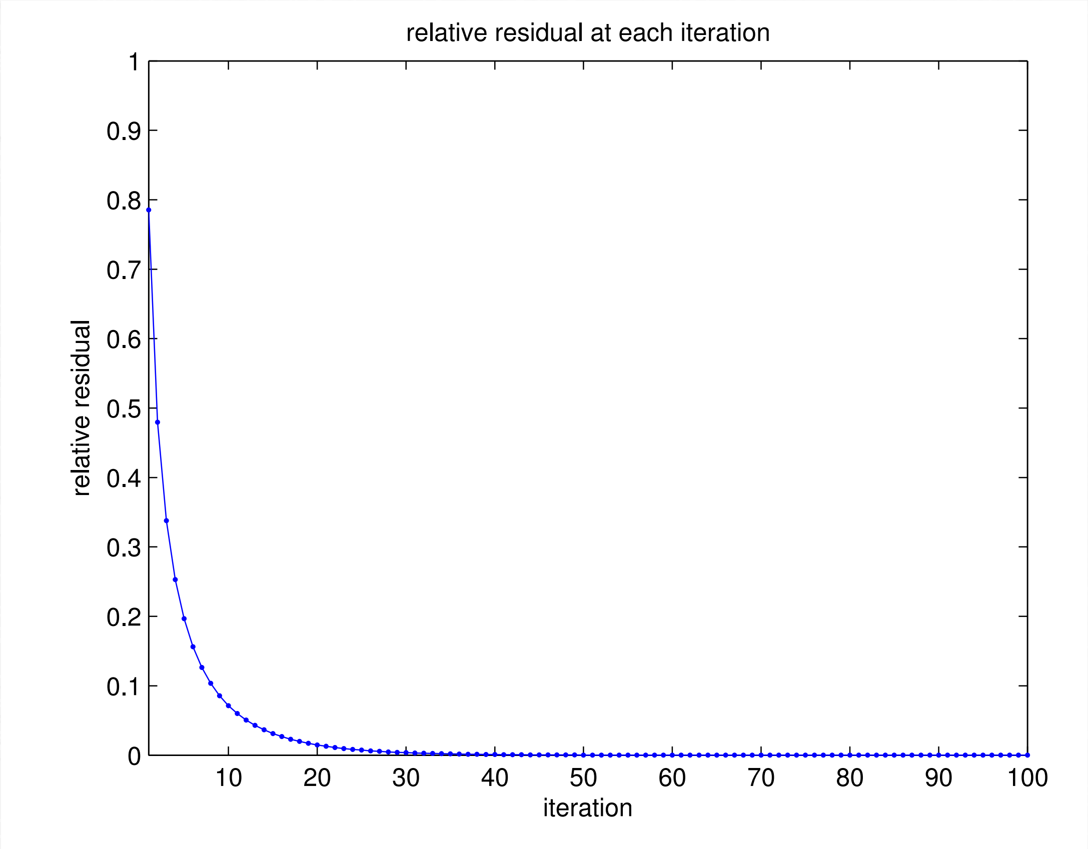
6% error, 0.7% residual

HRI

F=10, OR=2, # regular shifts=***100***
# data= 4 # pixels

Algorithm: alternating projection

(a) error = 0.0001%

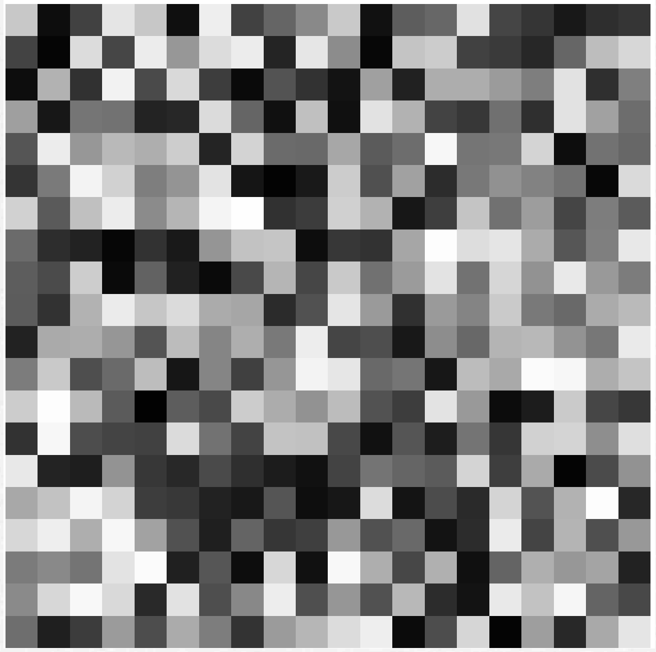


(b) relative residual, residual = 0.00005%

Low resolution illumination

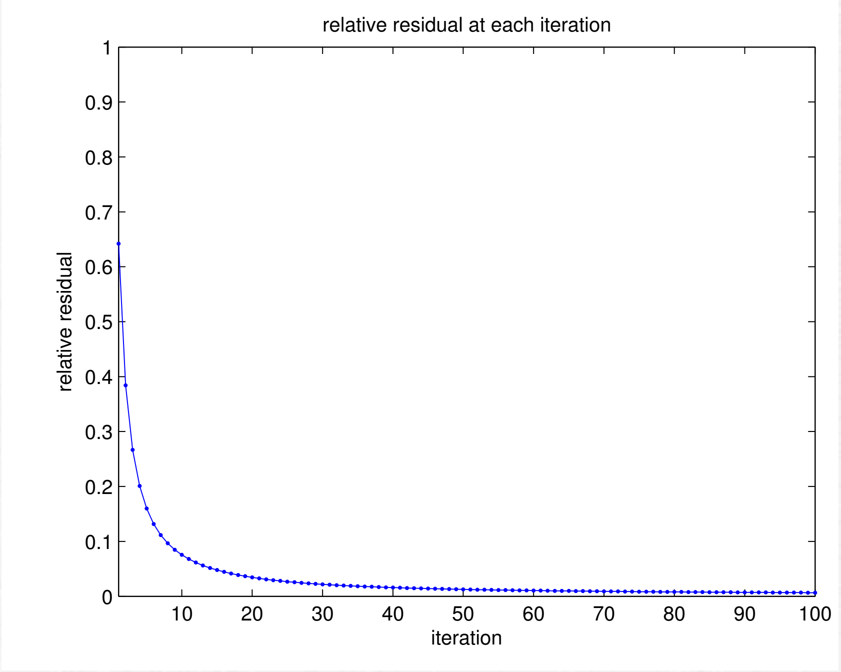
F=5, OR=2, # regular shifts=25
# data= 4 # pixels

Algorithm: alternating projection



LRI: independent 5x5-blocks

(a) error = 14.53%



(b) relative residual, residual = 0.66%

Conclusion

- BLO away the resolvable grid: BLOOMP, BP-BLOT
- Theoretical resolution: 2 or 3 Rayleigh lengths
- Practical resolution: 1 Rayleigh length
- Accuracy: a few percents of Rayleigh length
- # measurements ~ $s^2 x^2_{\mathrm{max}}/x^2_{\mathrm{min}}$, SNR
- Better than (thresholded) BP and analysis approaches such as Frame-adapted BP, SIHT
- BP-BLOT has super-resolution effect
- Roughly translation-invariant coherence pattern, cf. Adcock & Hansen: infinite-dim CS
- Super-resolution with random illumination

References

- F& Liao **Super-resolution by compressive sensing algorithms** IEEE Proc. Asilomar conference on signals, systems and computers, 2012.
- F& Liao **Coherence-Pattern Guided Compressive Sensing with Unresolved Grids** SIAM Journal of Imaging Sciences, Vol. 5, No. 1, pp. 179–202.
- F& Liao **Mismatch and resolution in compressive imaging** Wavelets and Sparsity XIV, Proc. of SPIE Vol. 8138, 2011.

Thank you