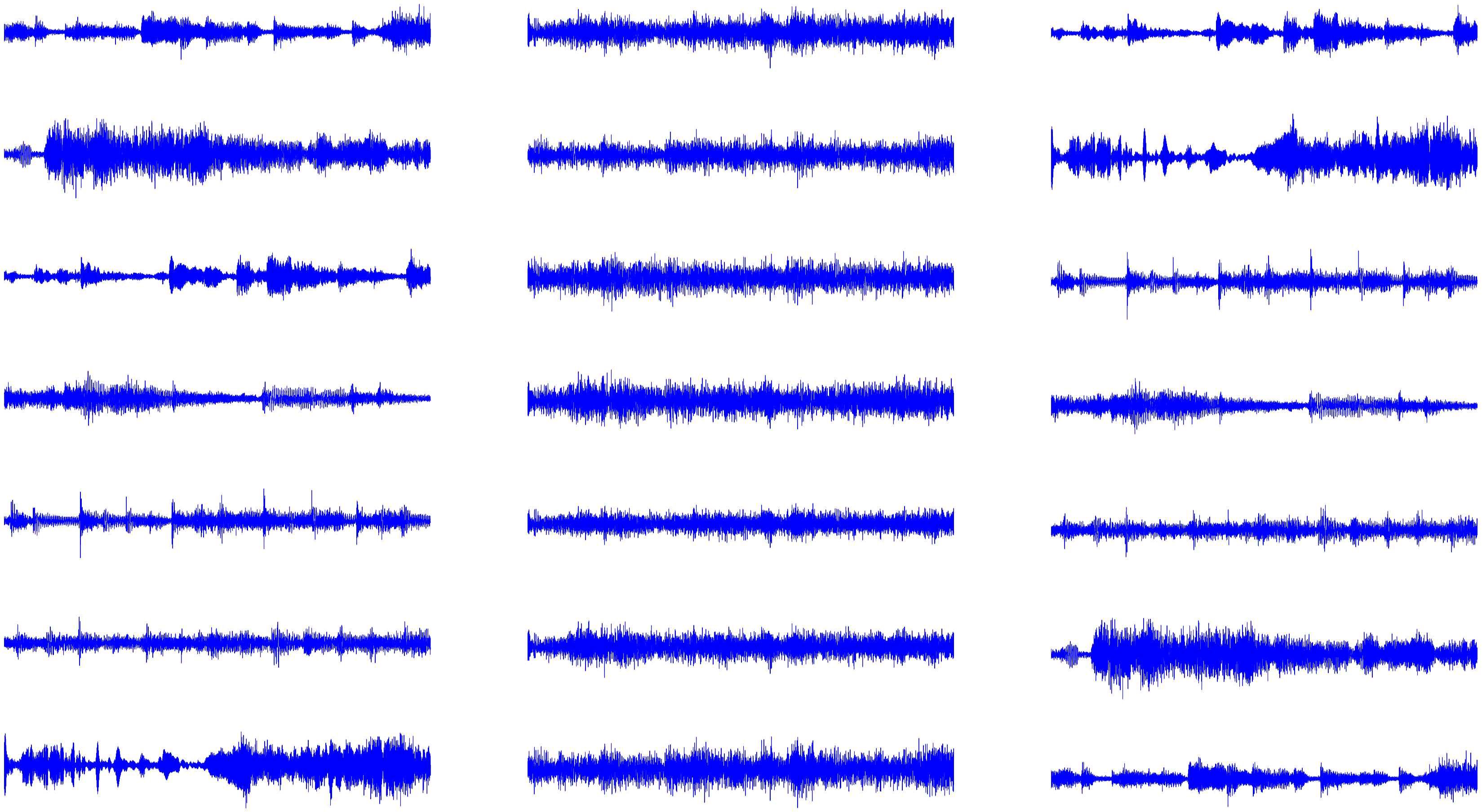
Blind Source Separation, Deconvolution and Localization using Sparse Representations

Michael Zibulevsky

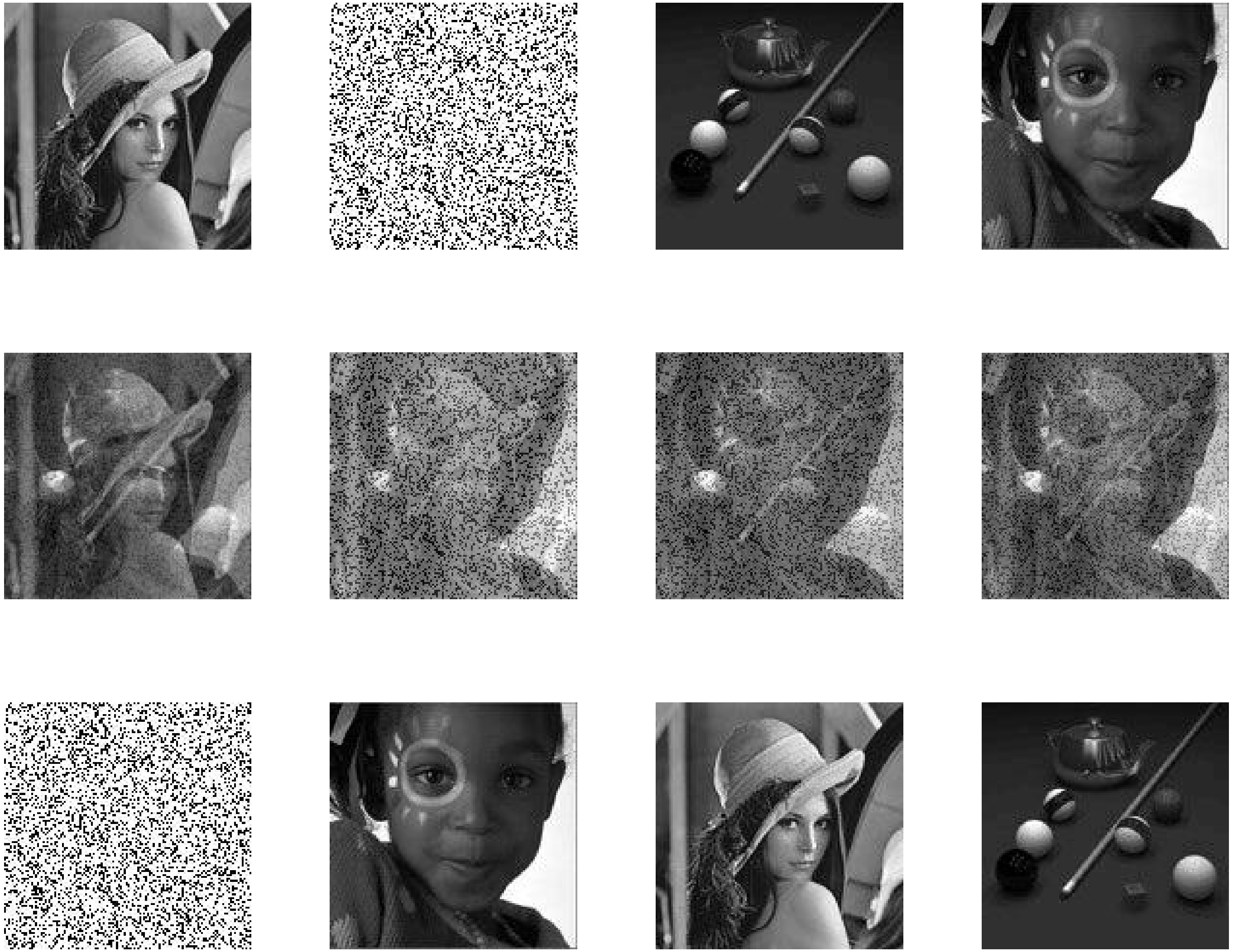
Department of Electrical Engineering
Technion – Israel Institute of Technology

Separation of Musical Sounds



Sources Mixtures Separated

Separation of images



Top – sources, middle – mixtures, bottom – separated

Application areas

- Electroencephalography (EEG)
- Magnetoencephalography (MEG)
- Electrocardiography (ECG)
- Functional MRI
- Hyperspectral imaging (satellite, bio, ...)
- Acoustics
- Radio
- Radar

Problem Formulation

Observed sensor signals:

$$x(t) = As(t) + \xi(t)\ ,$$

$s(t)$ – unknown source signals

A – unknown mixing matrix

$\xi(t)$ – white Gaussian noise

Sparse Representation of Sources

$$s_i(t) = \sum_{k=1}^{K} c_{ik} \varphi_k(t),$$

$\varphi_k(t)$ – *atoms* of an [overcomplete] signal *dictionary*

Coefficients c_{ik} are supposed to be sparse.

Soft sparseness:

$$p_i(c_{ik}) \propto e^{-h(c_{ik})}$$

$$h(c) = |c|^{\gamma} \ , \quad \gamma \leq 1 \ ,$$

$$
\begin{aligned}
x(t) &= As(t) + \xi(t) \\
s_i(t) &= \sum_{k=1}^{K} c_{ik} \varphi_k(t)
\end{aligned}
$$

Matrix Notation

$$
\begin{aligned}
X &= AS + \xi \\
S &= C\Phi \\
X &\approx AC\Phi
\end{aligned}
$$

matrix C should be sparse

Maximum a Posteriori Approach

Inspired by the idea of Basis Pursuit

(Chen, Donoho, Sanders)

$$\min_{A,C} \; \frac{1}{2\sigma^2}\|AC\Phi - X\|_F^2 + \sum_{j,k} h(c_{jk})$$

Subject to $\|A\| \leq 1$

where $\|U\|_F^2 = \sum_{i,j} U_{ij}^2$ Frobenius matrix norm

Non-Overcomplete Dictionary

Dual basis

$$\Psi = \Phi^{-1}$$

Decomposition coefficients of the sensor signal

$$Y = X\Psi$$

Mixing model

$$Y = AC$$

Unmixing matrix $W \approx A^{-1}$

$$\tilde{C} = WY$$

Quasi-maximum likelihood estimator

$$L(W;Y) = -\log|\det W| + \frac{1}{K}\sum_{j,k} h(c_{jk})$$

$$\min_{W} \; L(W;Y)$$

Relative Optimization of Quasi-Likelihood Function

Relative Optimization algorithm

- Start with an initial source estimate $C^1 = Y$;
- **For** $k = 1, 2, ...,$ until convergence
  - Start with separation matrix $V = I$
  - Get V^{k+1} by optimization step for $L(V; C^k)$
  - Update the source estimate
    $C^{k+1} = V^{k+1} C^k$;
- **End**

Relative optimization properties

- The relative (natural) gradient method (Cichocki, Amari, Cardoso) is a particular instance of relative optimization.

- Current step does not depend on mixing matrix – only on current source estimate.

  This means that ill-conditioning of mixing matrix is irrelevant when approaching solution.

Fast Relative Newton Method

- Newton system can be approximately solved as a set of 2x2 systems.
- Complexity as in gradient computation.
- To guarantee descent direction, invert signs of negative eigenvalues
- Block-coordinate version converges in 10-20 iterations to accuracy of 10^{-8}

Smoothing Method of Multipliers

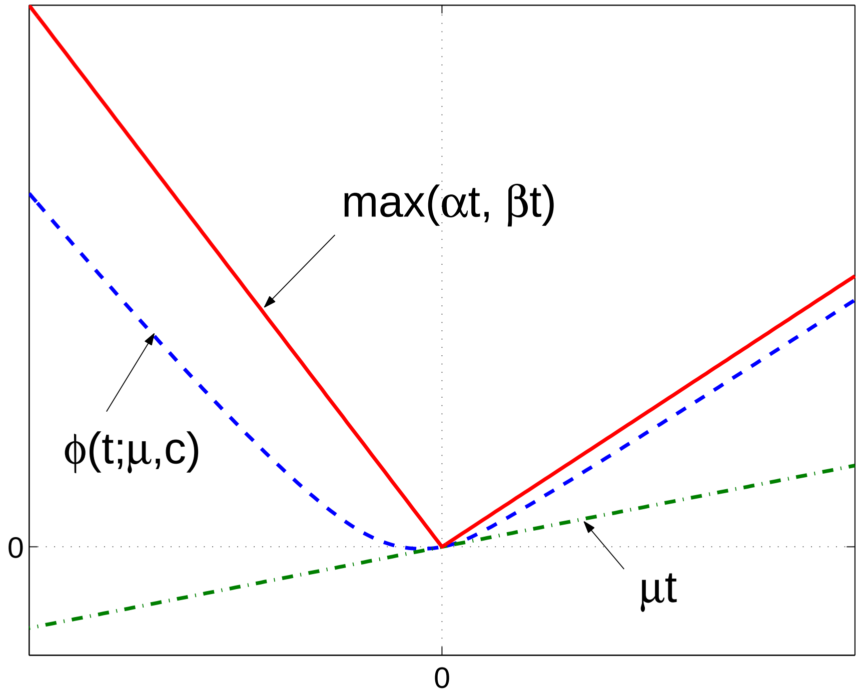
The problem:

$$\min\ F(x)$$

$$F(x) = f(x) + \sum_{i=1}^{m} |g_i(x)|,$$

$f(x)$ and $g_i(x)$ – smooth functions;

Smoothing absolute value function, we control slope μ at the origin



$$\min f(x) + \sum_{i=1}^{m} \varphi(g_i(x), u_i, c), \quad -1 \le u_i \le 1.$$

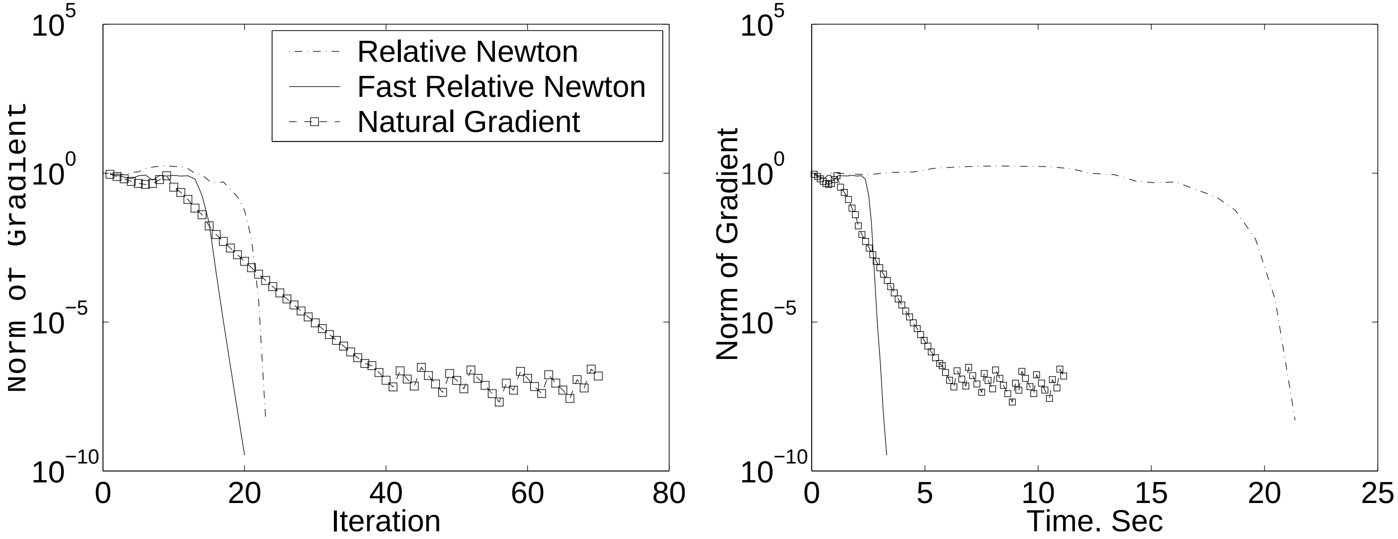
We can get an exact solution of the original problem, using correct slopes u_i. The method finds them iteratively.

SIMULATIONS

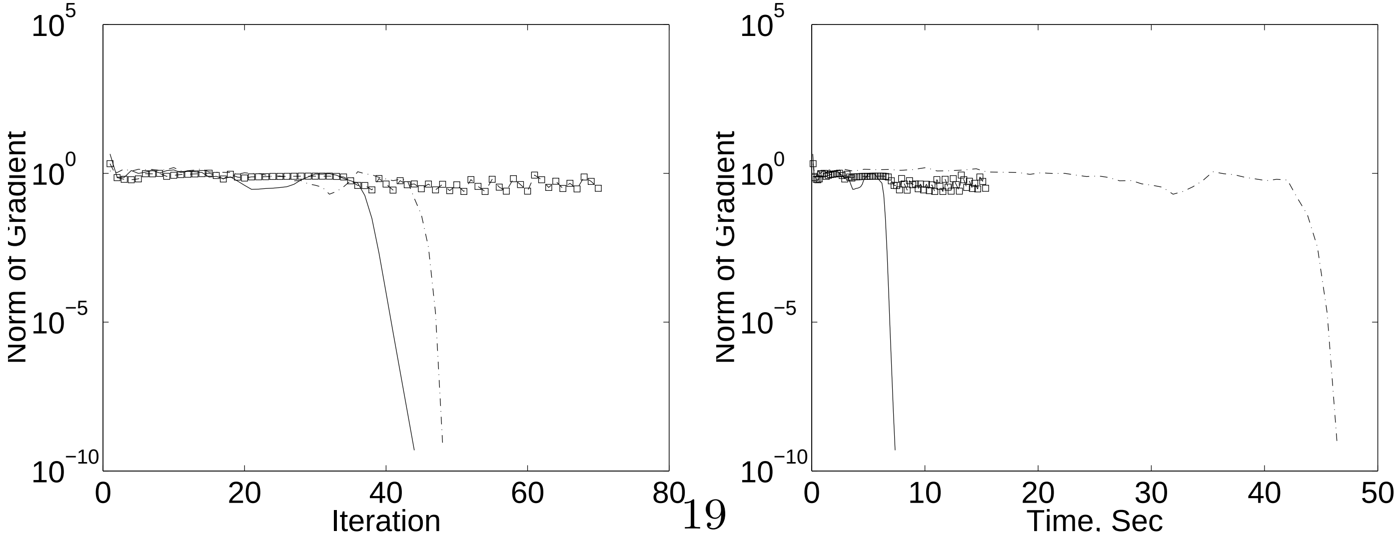
Data

- Ten sparse i.i.d. Gaussian sources (by sprandn.m)
- Four natural images
- Random mixing matrices with i.i.d. entries

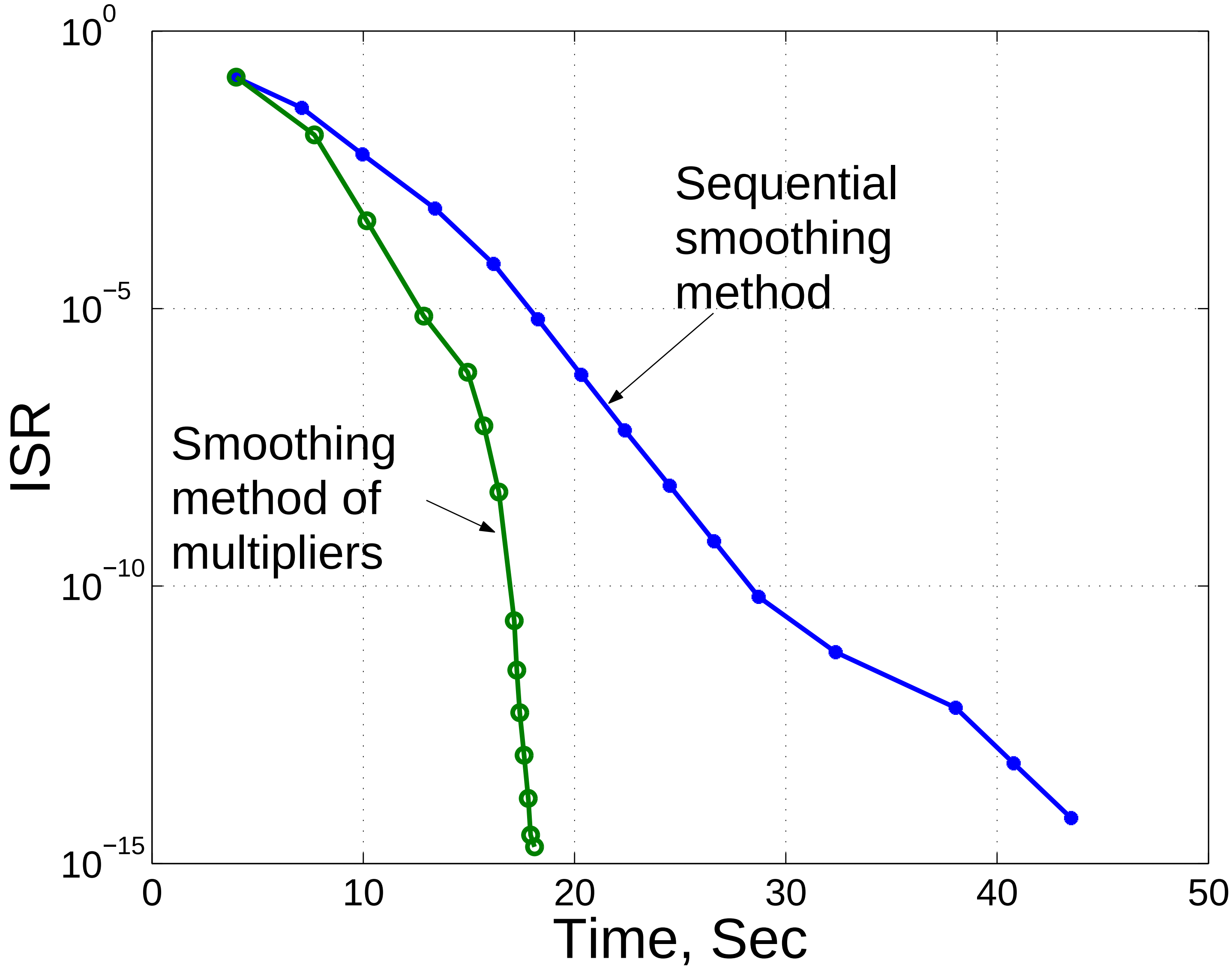
Convergence: smoothing parameter $\lambda = 1$

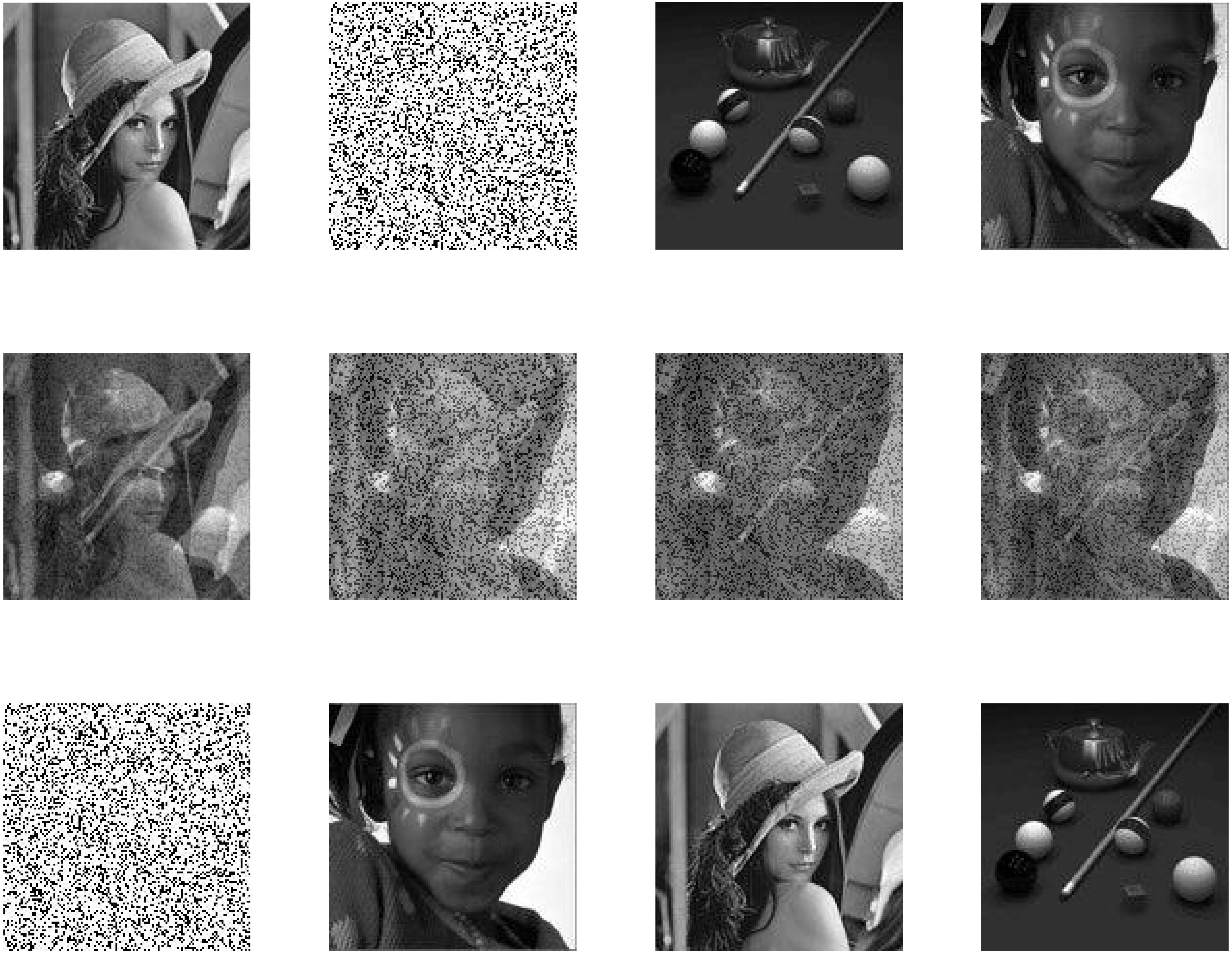


Convergence: smoothing parameter $\lambda = 0.01$

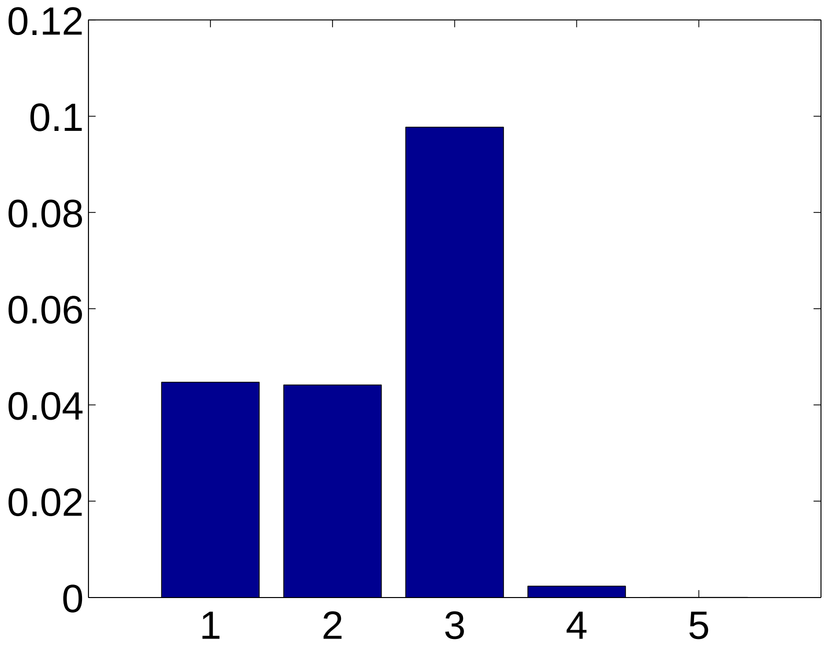


ISR progress with iterations / CPU time





ISR of image separation



1 – INFOMAX (Bell-Sejnowski);

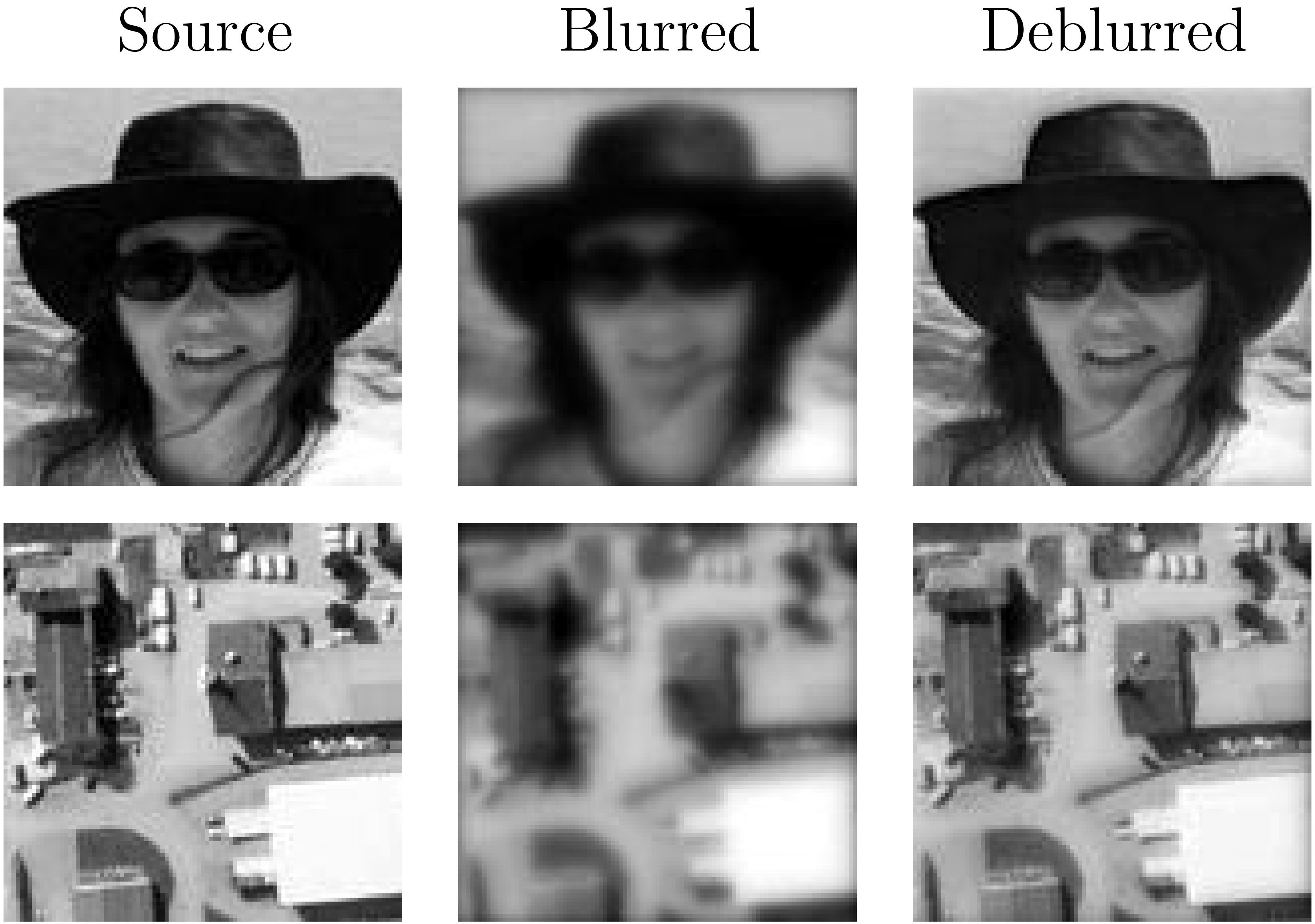
2 – Fast ICA (Hyvarinen);

3 – JADE (Cardoso);

4 – Sequential smoothing with $c = 100$;

5 – Smoothing method of multipliers, ISR$= 10^{-15}$

Deconvolution



Source

Blurred

Deblurred

Signal Reconstruction in Sensor Arrays Using Temporal-Spatial Sparsity

General Approach

- Divide the space into discrete grid of locations (much more possible locations than active sources).
- Formulate solution through inverse problem framework. The problem is ill-posed.
- Use l_1-norm regularization, which is known to effectively enforce sparsity.

Observation Model

Source signals of duration T from several of m grid nodes arrive to each of n sensors with different delays and attenuations

$$Y = \mathcal{A}S + N$$

S – $m \times T$ source matrix (signals in rows);

$\mathcal{A}$ – *forward operator*, which *shifts*, *attenuates* and *sums* incoming signals, modeling the real environment.

N – noise;

Y – $n \times \tilde{T}$ sensors measurements matrix;

Convex objective function

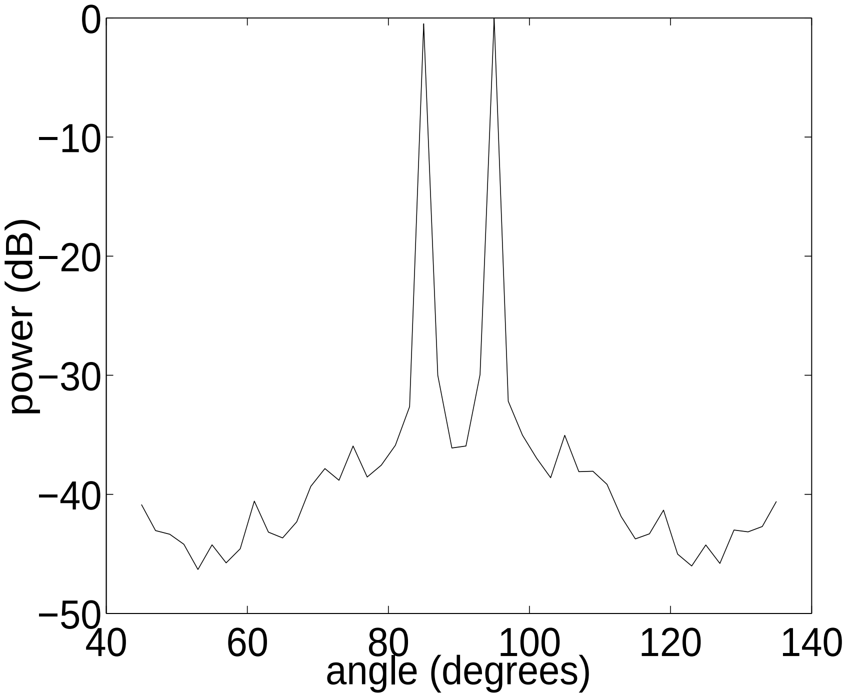
$$F(C) = \frac{1}{2}\|Y - \mathcal{A}(C\Phi)\|_F^2 + \mu_1 \sum_{i,j} |c_{ij}| + \mu_2 \sum_{i=1}^{m} \|c_i\|_2$$

- Term $\mu_1 \sum_{i,j} |c_{ij}|$ enforces sparsity of the coefficients;

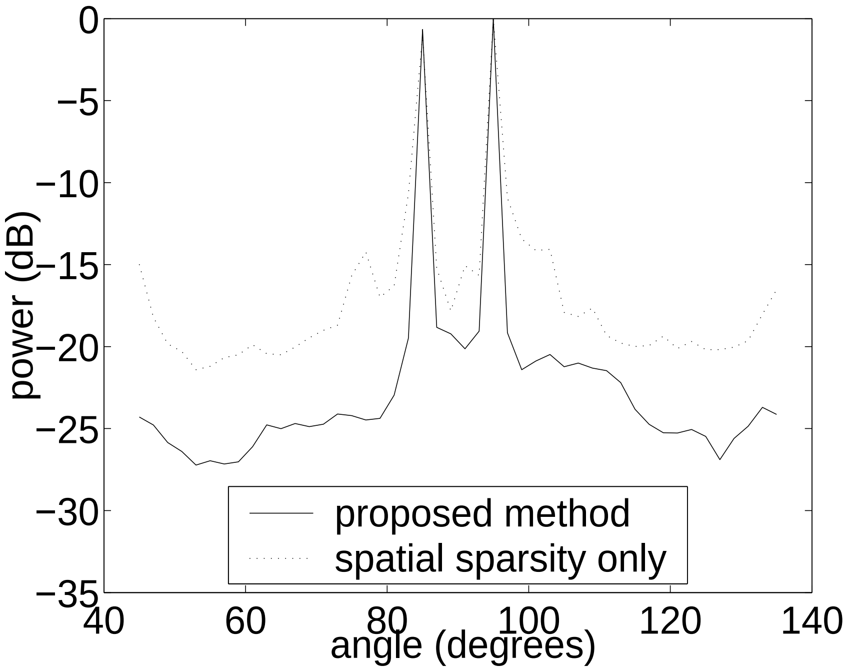
- Term $\mu_2 \sum_{i=1}^{m} \|c_i\|_2$ enforces solutions with the source signals concentrated in a small number of locations; c_i – i-th row of the matrix C (the i-th source' coefs)

- Preconditioned Truncated Newton method is used to solve this large-scale optimization problem

Direction of Arrival (DOA) Estimation

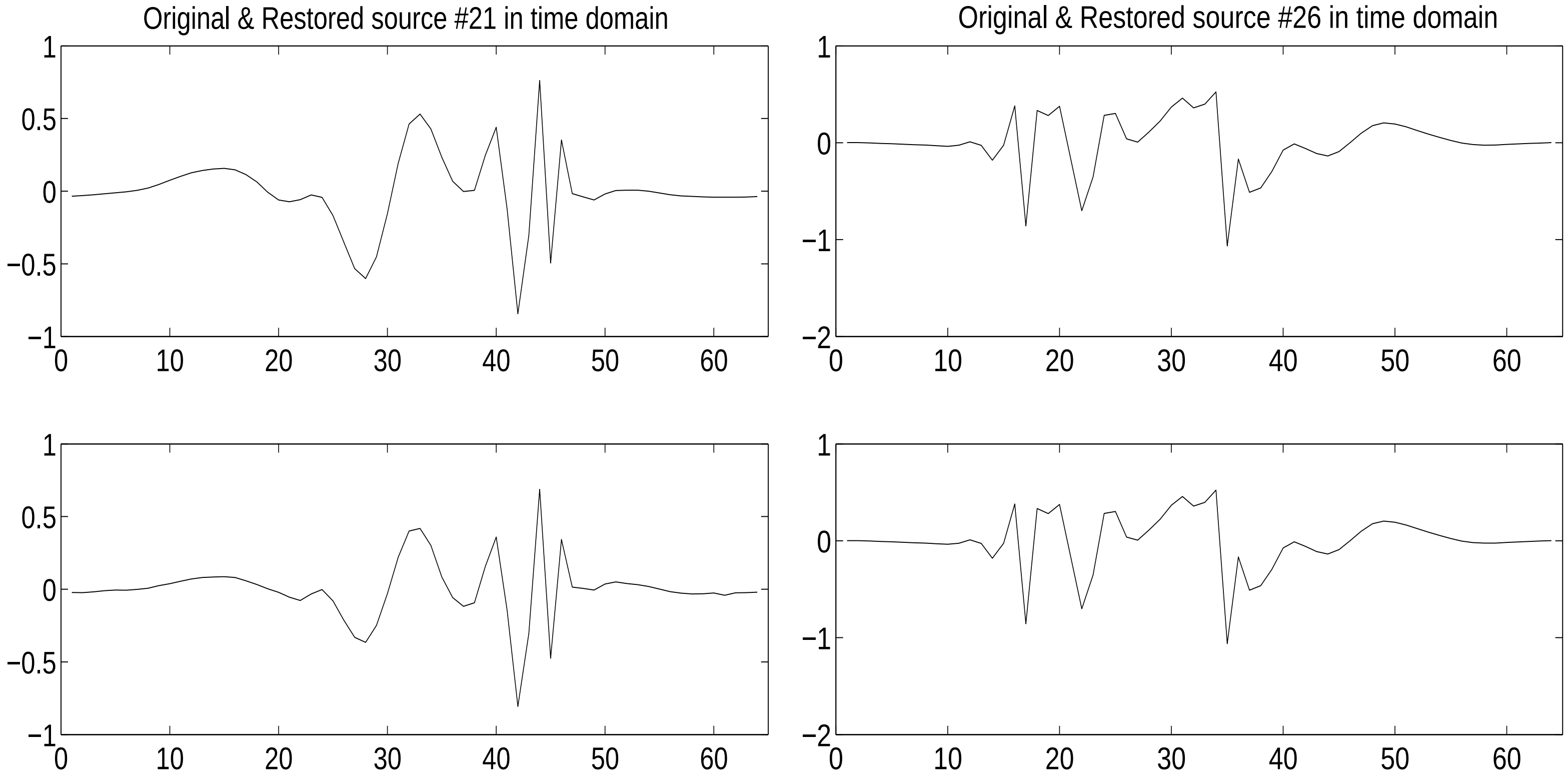


No noise



SNR=5 dB

Source Reconstruction, SNR = 5dB



Top: sources from 2 active directions,

bottom: restored sources

Conclusions

- Sparse source representation dramatically improves quality of separation.
- Quasi-ML separation of sparse sources is practically perfect when the nonlinearity approaches the absolute value function.
- Relative Newton iteration has a complexity of gradient step
- Smoothing method of multipliers provides accurate solution with moderate smoothing parameter.

- Separation/localization of sources with temporal- spatial sparsity can be formulated as a *convex* optimization problem with L_1 and L_2 norm terms.
- This method achieves super-resolution in source localization.
- The separation is robust to low SNR and very limited data