

BLOCK-COORDINATE

RELATIVE NEWTON METHOD

FOR BLIND SOURCE SEPARATION

Alexander Bronstein
Michael Bronstein
Michael Zibulevsky

© 2003

The observed mixtures:

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

$s(t)$ unknown sources

A unknown invertible mixing matrix

The BSS problem:

$$\hat{s}(t) = Wx(t),$$

where $W \approx A^{-1}$

The minus-log-likelihood function

$$L(W; X) = -\log |\det W| + \frac{1}{T} \sum_{i,t} h_{\lambda} (W_i x(t)),$$

where

$W \approx A^{-1}$

unmixing matrix

W_i

i -th row of W

$h_{\lambda}(\cdot) \approx |\cdot|$

smooth convex approximation of the module function
with the smoothness parameter λ .

The relative Newton algorithm [1] for the minimization of the objective function:

1. Start with an initial estimate, $W^{(1)}$ of the separation matrix;
2. **FOR** $k = 1, 2, \dots$ until convergence
3. Compute current source estimate $U^{(k)} = W^{(k)}X$;
4. Starting with $V = I$, compute $V^{(k+1)}$ producing single Newton step of $L(V; U^{(k)})$
5. Update the estimated separation matrix $W^{(k+1)} = V^{(k+1)} W^{(k)}$;
6. **END**

[1] M. Zibulevsky, "Relative Newton method for quasi-ML blind source separation", Submitted to JMLR, 2002.

Newton method requires the **gradient** and the **Hessian** of the objective function.

The gradient is given by

$$\nabla L(w; X) = -W^T + \frac{1}{T} h'_\lambda(WX) X^T$$

where

$w = \text{vec}(W)$ is a row-stack representation of the matrix W

The Hessian of the first term, $-\log \det W$, is given by:

$$H_k = \text{vec}^T \left(A^j A_i \right)^T ; \quad k = (i-1)N + j$$

where

A_i i -th row of A

A^j j -th column of A

The Hessian of the second term, $\frac{1}{T} \sum_{i,t} h_\lambda (W_i x(t))$, is a block-diagonal

matrix with the following $N \times N$ blocks:

$$B^m = \frac{1}{T} \sum_{i,t} h''_\lambda (W_m x(t)) x(t) x^T(t) ; \quad m = 1, \dots, N$$

Given $V = I$ and assuming independent and zero-mean sources, the first term of the Hessian becomes

$$H_k = \text{vec}^T \left(e_i e_j^T \right)$$

where

e_i standard basis vector, containing 1 at the i -th place

For the second term of the Hessian, one can use the following **diagonal** approximation:

$$B_{ii}^m = \frac{1}{T} \sum_{i,t} h_{\lambda}''(u_m(t)) u_i(t) \quad ; \quad m = 1, \dots, N$$

where

$u_m(t)$ are current estimates of the sources

Such an approximation is more accurate as the sample size T grows.

This greatly simplifies the Newton direction computation, requiring only $N(N-1)/2$ systems of 2×2 linear equations to be solved.

1. Start with an initial estimate, $W^{(1)}$ of the separation matrix;
2. **FOR** $k = 1, 2, \dots$ until convergence
3. **FOR** $i = 1, \dots, M$
4. **FOR** $j = i, \dots, M$
5. Efficiently update $U^{(k)} = W^{(k)}X$;
6. Starting with $V = I$, compute $V^{(k+1)}$ producing one block-coordinate Newton step with the blocks W_{ij} , W_{ji} .
7. Update the estimated separation matrix $W^{(k+1)} = V^{(k+1)} W^{(k)}$;
8. **END**
9. **END**
10. **END**

- $N(N-1)/2K^2$ inner iterations of the block-coordinate method update the entire matrix W and are equivalent to a single (“full”) relative Newton step.
- For $K=N$, the relative Newton step is obtained.
- The same number of 2×2 equations must be solved for any K .
- The cfunction, its gradient and Hessian shall be computed efficiently.

“Full” matrix computation

$$U = \begin{matrix} N & & T \\ \boxed{W} & \cdot & \boxed{X} \\ N & & N \end{matrix}$$

$$h(U) = h \begin{matrix} T & & N \\ \boxed{U} & & \end{matrix}$$

$$\det \begin{matrix} N \\ \boxed{W} \end{matrix}$$

Block-coordinate computation

$$U = \begin{matrix} K & & T \\ \boxed{W_{ij}} & \cdot & \boxed{X} \\ & & \end{matrix}$$

$$h(U) = h \begin{matrix} T & & K \\ \boxed{U} & & \end{matrix}$$

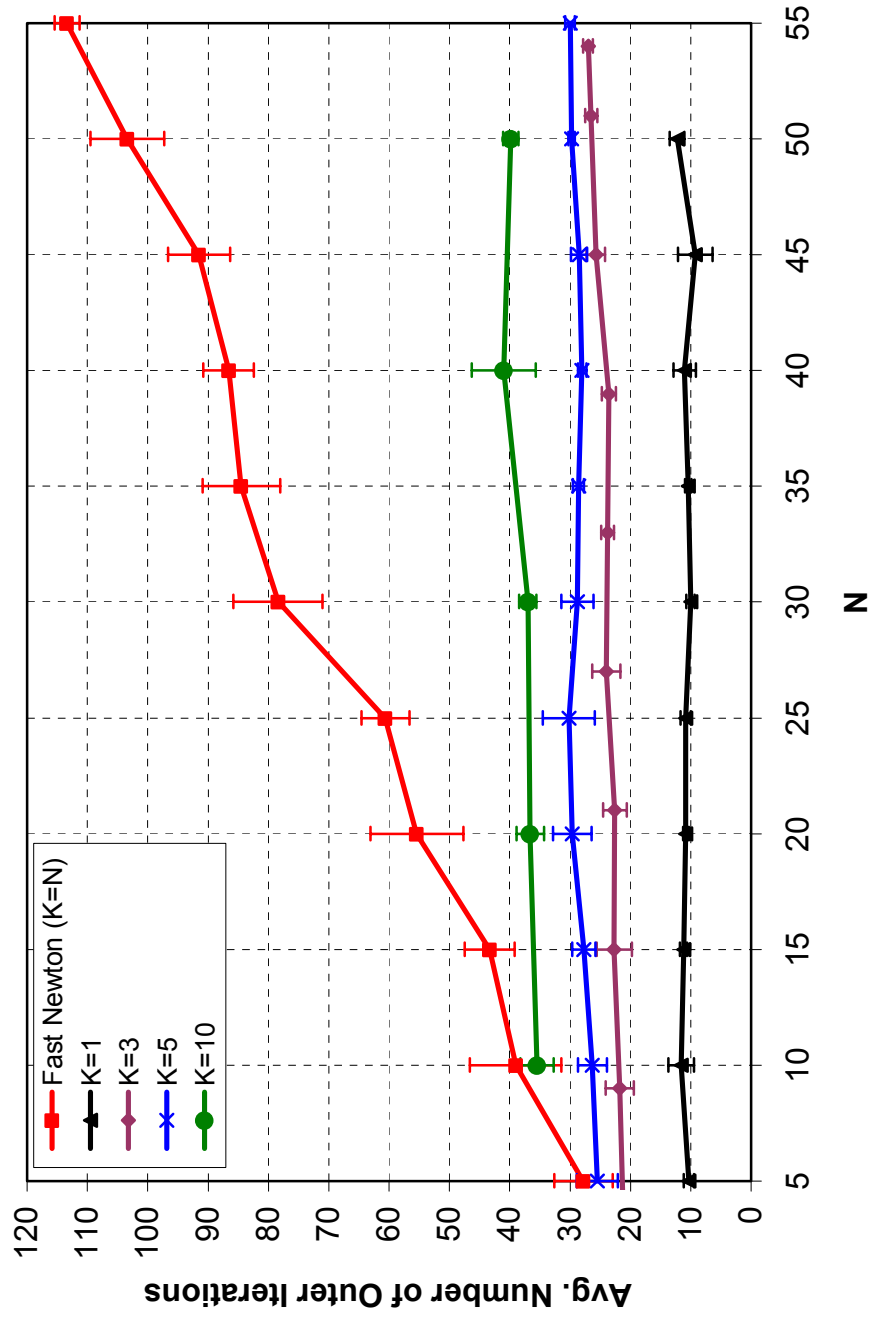
$$\det W = \det \begin{matrix} 2K \\ \begin{matrix} I & W_{ij} \\ W_{ji} & I \end{matrix} \end{matrix}$$

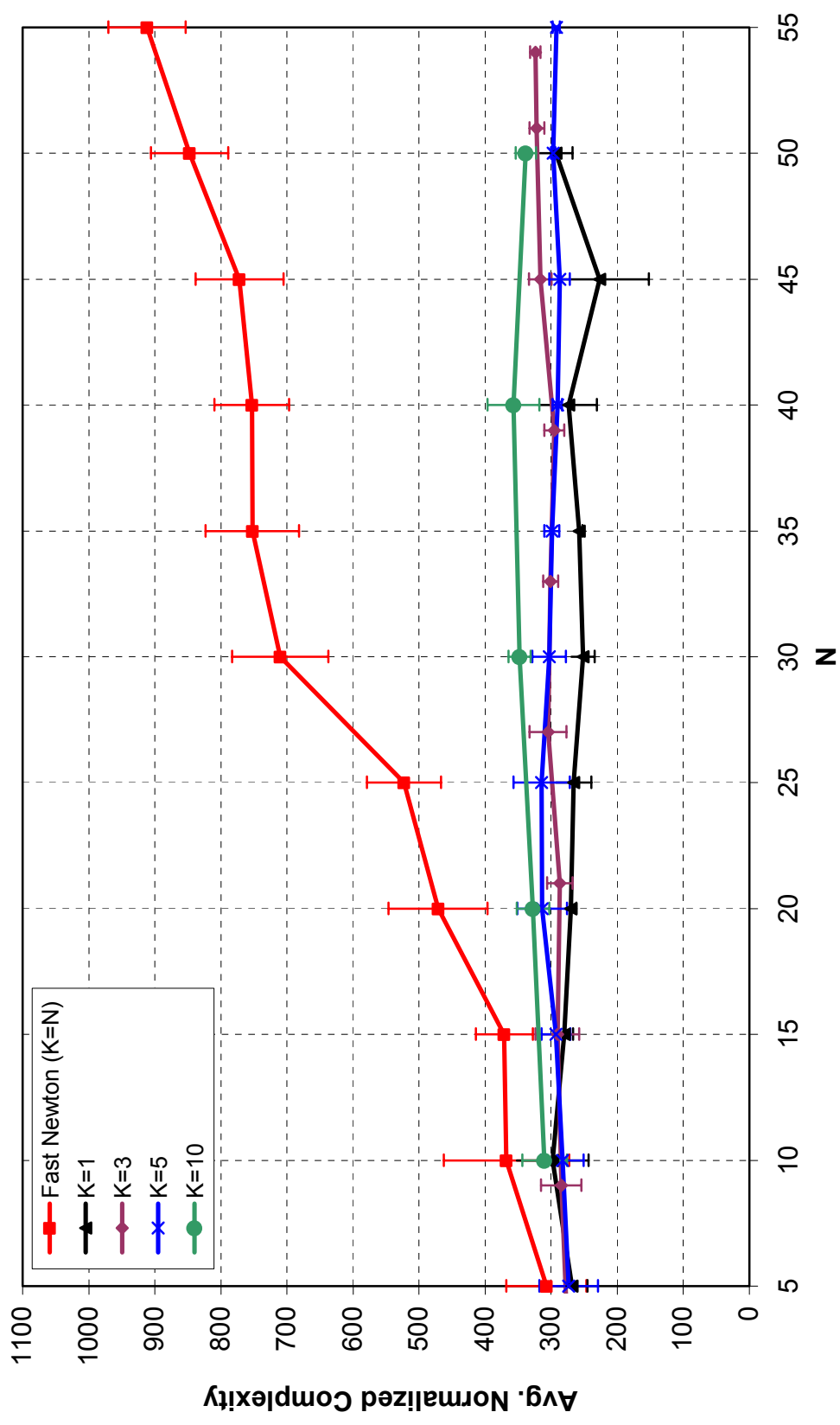
- The block-coordinate method is advantageous over the “full” relative Newton if it reduces the number of outer iterations by more than

$$\beta > 1 + \frac{\alpha + 1}{3K}$$

where α is the computational complexity of $h_\lambda, h'_\lambda, h''_\lambda$
($\alpha \approx 10$ multiplications)

The algorithms were compared on problems of different size with constant $T = 1000$, $\lambda = 0.1$. Sparse sources were used.





- The block-coordinate method exhibited superior performance over the relative Newton method.
- In preliminary studies, the block-coordinate method showed an approximately **constant** number of outer iterations, independent on N .