A LEAST ABSOLUTE BOUND APPROACH TO ICA

APPLICATION TO THE MLSP 2006 COMPETITION

John A. Lee

Frédéric Vrins [frederic.vrins@uclouvain.be]

Michel Verleysen

Machine Learning Group

Université catholique de Louvain

Louvain-la-Neuve, Belgium

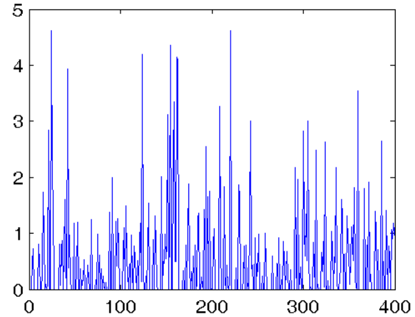
Independent Component Analysis

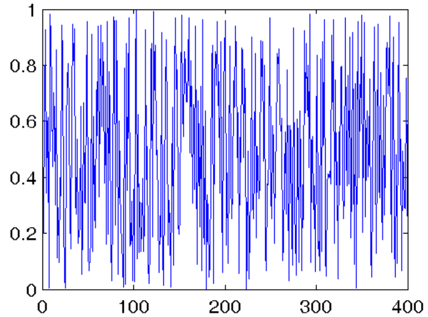
- Mixture model (linear and instantaneous)
  - $\mathbf{x}(\mathbf{t}) = \mathbf{A}\mathbf{s}(\mathbf{t})$ (n sources, n mixtures, 1≤t≤N)
- Aim: identify blindly the mixing coefficients (t omitted)
  - $\mathbf{y} = \mathbf{B}\mathbf{x} = \mathbf{B}\mathbf{A}\mathbf{s} = \mathbf{C}\mathbf{s}$ s.t. $\mathbf{C} = \mathbf{P}\mathbf{D}$ (permutation & scaling)
- Main assumption
  - Elements of $\mathbf{s}$ are statistically independent
- 2-stage separation
  - Whitening using SVD: $\mathbf{z} = \mathbf{V}(\mathbf{x} - \langle\mathbf{x}\rangle)$ s.t. $\mathrm{E}(\mathbf{z}\mathbf{z}^{\mathbf{T}}) = \mathbf{I}$
  - Unmixing: $\mathbf{y} = \mathbf{W}\mathbf{z}$ s.t. $\mathbf{W}\mathbf{W}^{\mathbf{T}} = \mathbf{I}$
- Perf. Index of the Competition

$$\mathrm{SIR} = \frac{1}{n}\sum_{i=1}^{n} 10\log_{10}\frac{\max_j c_{ij}^2}{\sum_j c_{ij}^2 - \max_j c_{ij}^2}$$

2006 MLSP Competition : BSS contest

- Sources: positive, 2 types (50%-50%)
  - Sparse in [0,∞[(super-Gaussian)
  - Uniform in [0,1] (subGaussian)





- 4 sub-problems to solve:

  (for each: 90% of the SIRs from MC runs > 15dB)

  - Large-scale problem (N=5000, largest n, **A** : rand.)
  - Small sample problem (lowest N, n=50, **A** : rand.)
  - Ill-conditioned mixtures (N=5000, largest n, **A** : Hilbert x Givens)
  - Noisy mixtures (N=5000, n=50, **A** : rand., largest σ_{noise})

To fit or not to fit (the method to the data)

- Two main approaches
  - Use general-purpose algorithms (Jade, fastICA,...)
  - Use *dedicated algorithms*, that better match to the a priori knowledge on the sources (Positive, Bounded Srcs). Here: *one-side bounded* sources: let's deal with that knowledge !
- Outline:
  - Ad-hoc criterion *fitted to the sources*
  - Optimization algorithm *fitted to the criterion*
  - Results on MLSP 06 benchmark

Minimum Range approach to ICA

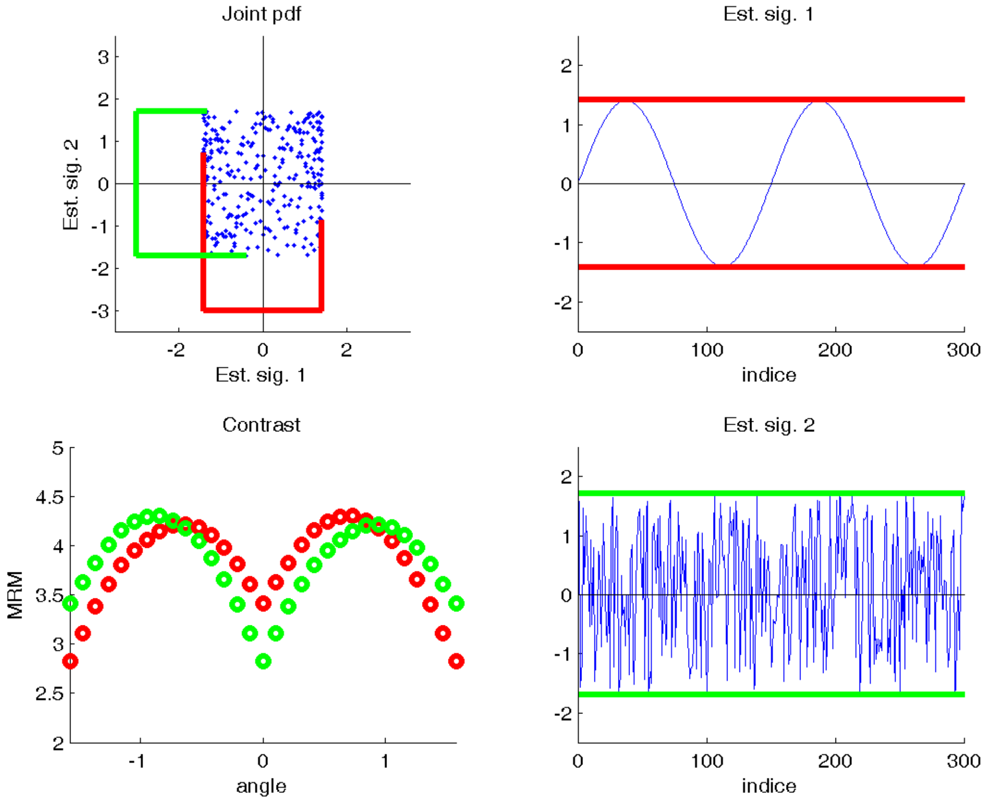
- Previous work about double-bounded sources…

  [Vrins, Jutten & Verleysen, ICASSP 2005]

- Toy illustration:
  - Sine
  - Uniform

<u>Issue #1:</u>
MLSP Contest sources are double-bounded and single-bounded !!!

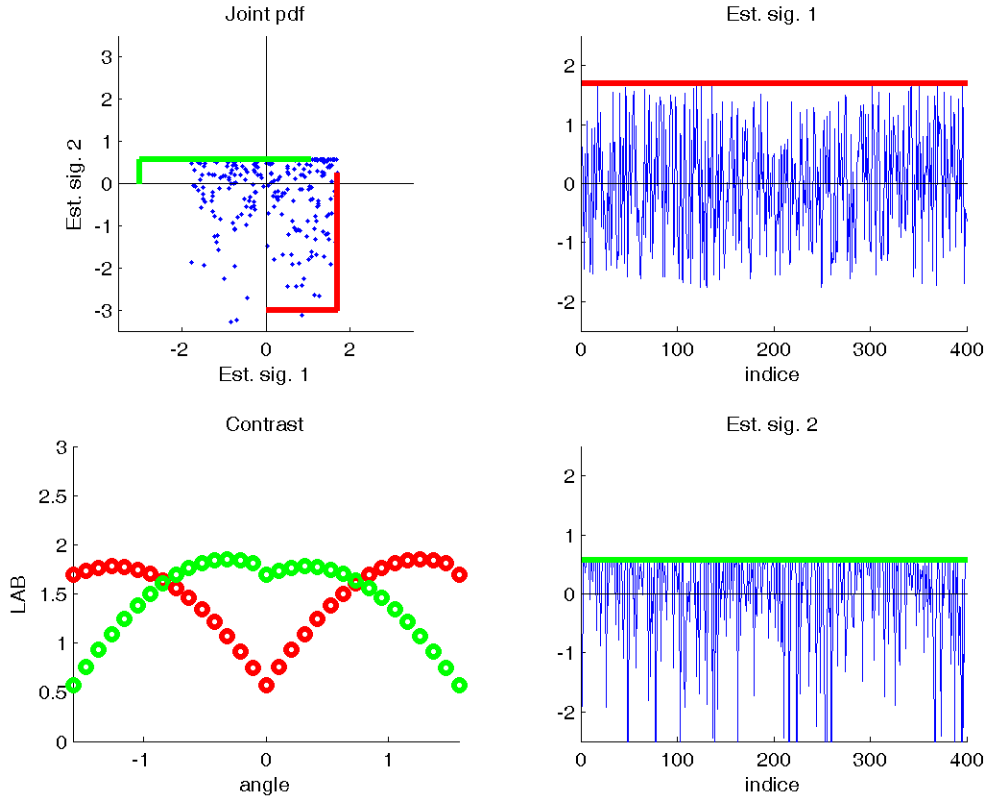
<u>Issue #2:</u>
seeked points are not stationary !!!



Issue #1: Least Absolute Bound ICA

- Idea: measure the range on one side only

- Toy illustration:
  - MLSP sources



Issue #1: Solved ☺
OK for double- and single-bounded sources

Issue #2: Pending… ☹
LAB is not everywhere differentiable either

R vs LAB: estimation of extreme pts

- Define
  - $\mathbf{y}' = \text{sort}(\mathbf{y}) : \forall \mathbf{1} \leq \mathbf{t} \leq \mathbf{N}-\mathbf{1},\ \mathbf{y}'(\mathbf{t}) \leq \mathbf{y}'(\mathbf{t}+\mathbf{1})$
- Minimum Range Measure
  - $R(\mathbf{Y}) \approx \mathbf{y}'(\mathbf{N}) - \mathbf{y}'(\mathbf{1})$ (only 2 points are involved in est.)
  - \+ robust : $R(\mathbf{Y}) \approx \langle \mathbf{y}'(\mathbf{N}-\mathbf{q}+\mathbf{1}), \ldots, \mathbf{y}'(\mathbf{N})\rangle - \langle \mathbf{y}'(\mathbf{1}), \ldots, \mathbf{y}'(\mathbf{q})\rangle$
- Least Absolute Bound
  - $LAB(\mathbf{Y}) \approx \min\{-\mathbf{y}'(\mathbf{1}), \mathbf{y}'(\mathbf{N})\}$ (a single point is used !)
  - \+ robust: $LAB(\mathbf{Y}) \approx \min\{-\langle \mathbf{y}'(\mathbf{1}), \ldots, \mathbf{y}'(\mathbf{q})\rangle, \langle \mathbf{y}'(\mathbf{N}-\mathbf{q}+\mathbf{1}), \ldots, \mathbf{y}'(\mathbf{N})\rangle\}$
- N large enough : $\mathbf{y}'(\mathbf{q}) \approx \inf \mathbf{Y},\ \mathbf{y}'(\mathbf{N}-\mathbf{q}+\mathbf{1}) \approx \sup \mathbf{Y}$

Choosing q: [Vrins & Verleysen, ICA 2006]

Issue #2: Algo (LAB minimization)

- Mean subtraction
- Robust whitening:
  - SVD whitening directly on centered mixtures
  - Discard badly whitened mixtures keep only, let's say, $\mathbf{z_1, \ldots, z_{n-p}}$
- Deflation (for i=1:n-p)
  - LAB is minimized for directions $\mathbf{w}_i$ orthogonal to 'edges' of the joint support
    - Go away from corners
  - Random $\mathbf{w}_i^{(k=1)}$ perp to $\mathbf{w}_1, \ldots, \mathbf{w}_{i-1}$ and iterate on k ('till convergence)
    - Closest corner direction $\mathbf{u}^{(k)}$ given by normalized $\mathbf{z}$-coordinates of $\min[-\min(\mathbf{w}_i^{(k)}\mathbf{z}), +\max(\mathbf{w}_i^{(k)}\mathbf{z})]$
    - Update $\mathbf{w}_i^{(k)}$ in opposite direction (rotation towards $-\mathbf{u}^{(k)}$)
  - Slightly relaxed orthogonality constraint : almost no error accumulation ☺

If $\arccos \frac{\mathbf{w}_i^{(k)}\mathbf{w}_r}{\|\mathbf{w}_i^{(k)}\|.\|\mathbf{w}_r\|} < \pi/6$ then reinitialize

Then, put: $\mathbf{W}=[\mathbf{w}_1, \ldots, \mathbf{w}_{n-p}, \mathbf{e}_{n-p+1}, \ldots, \mathbf{e}_n]$

Algorithm: LAB minimization

- Properties of exact criterion ≈ properties of the estimated one
  - none of the minima are spurious : `convex contrast'
- Update rule: $\mathbf{w_i^{temp}} = \cos(\alpha^{(\mathbf{k})})\mathbf{w_i^{(k)}} - \sin(\alpha^{(\mathbf{k})})\mathbf{u^{(k)}}$

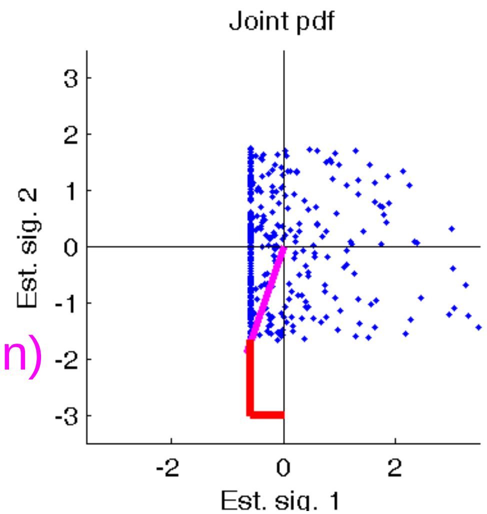
If $LAB(\mathbf{w}_i^{\text{temp}}) < LAB(\mathbf{w}_i^{(k)})$, then $\mathbf{w_i^{(k+1)}} \leftarrow \mathbf{w_i^{temp}}, \alpha^{(k+1)} \leftarrow 1.01\alpha^{(k)}$

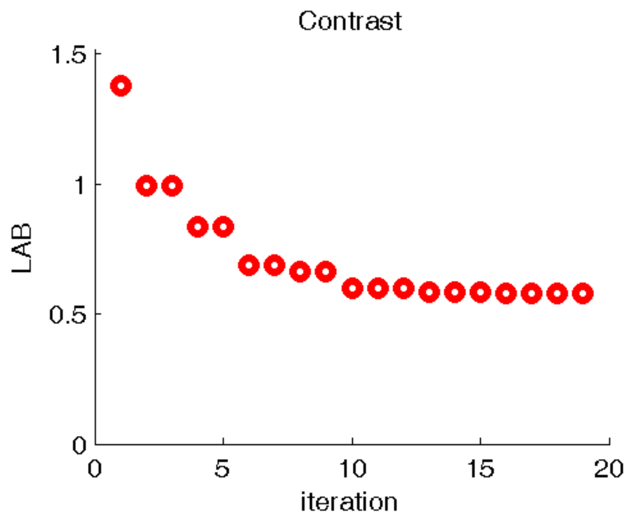
else: $\alpha^{(k+1)} \leftarrow \alpha^{(k)}/1.2$

2D demo:

LAB

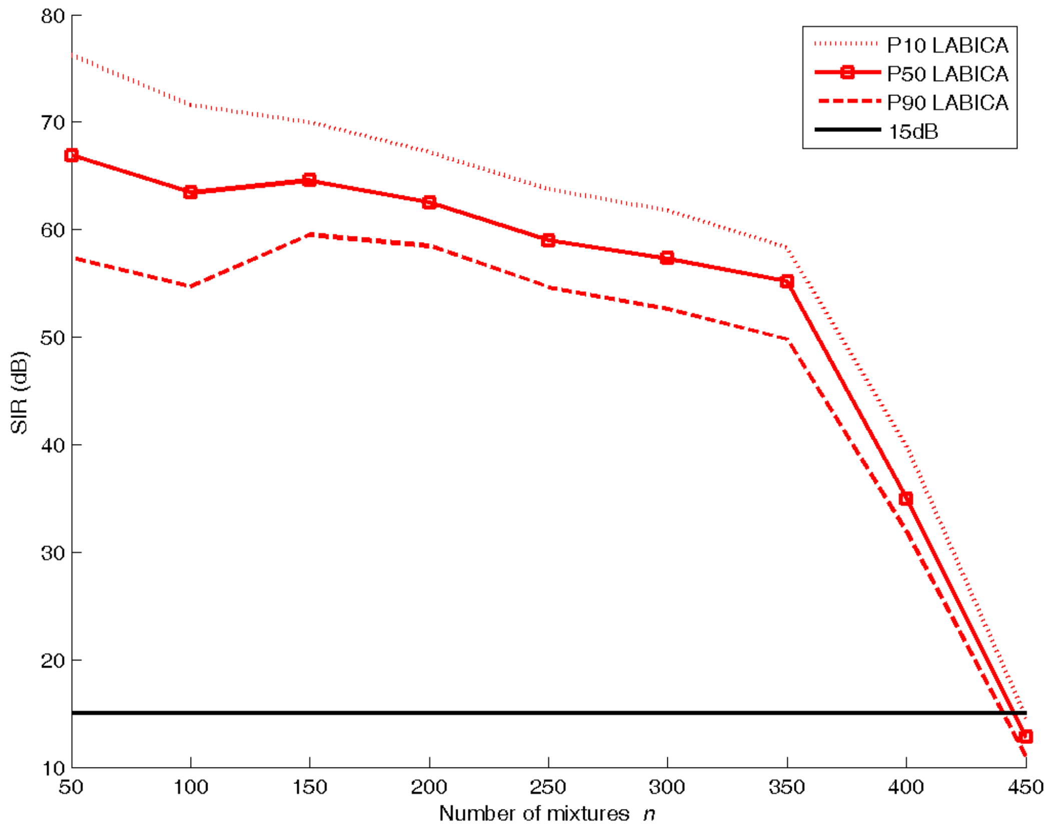
2 **u**
(corner direction)





Problem 1: large-scale mixtures

- Random mixing matrix
  - N=5000, n>50 (increasing number of sources)
- Results:
  - 20 Monte Carlo



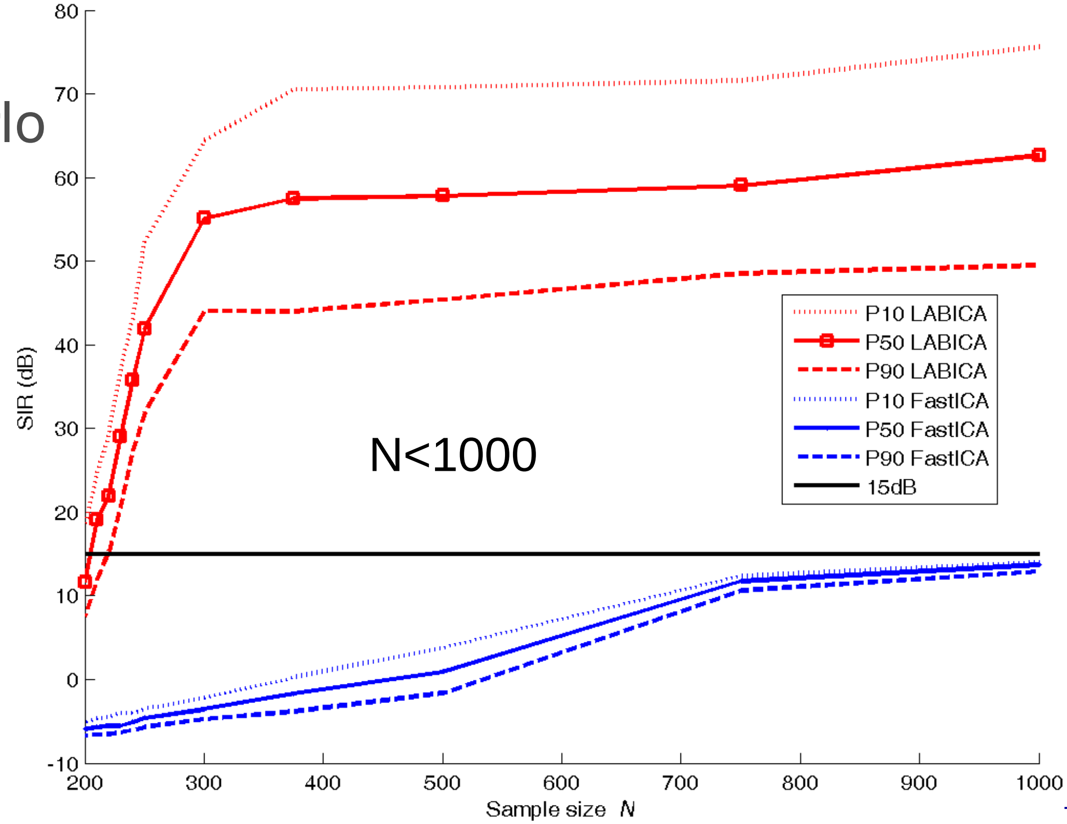
50 or 350 mixtures: almost the same separation quality ☺

More than 400 sources can be separated ☺

Problem 2: Small sample problem

- Random mixing matrix
  - N<5000 (decreasing sample size), n=50
- Results:
  - 100 Monte Carlo

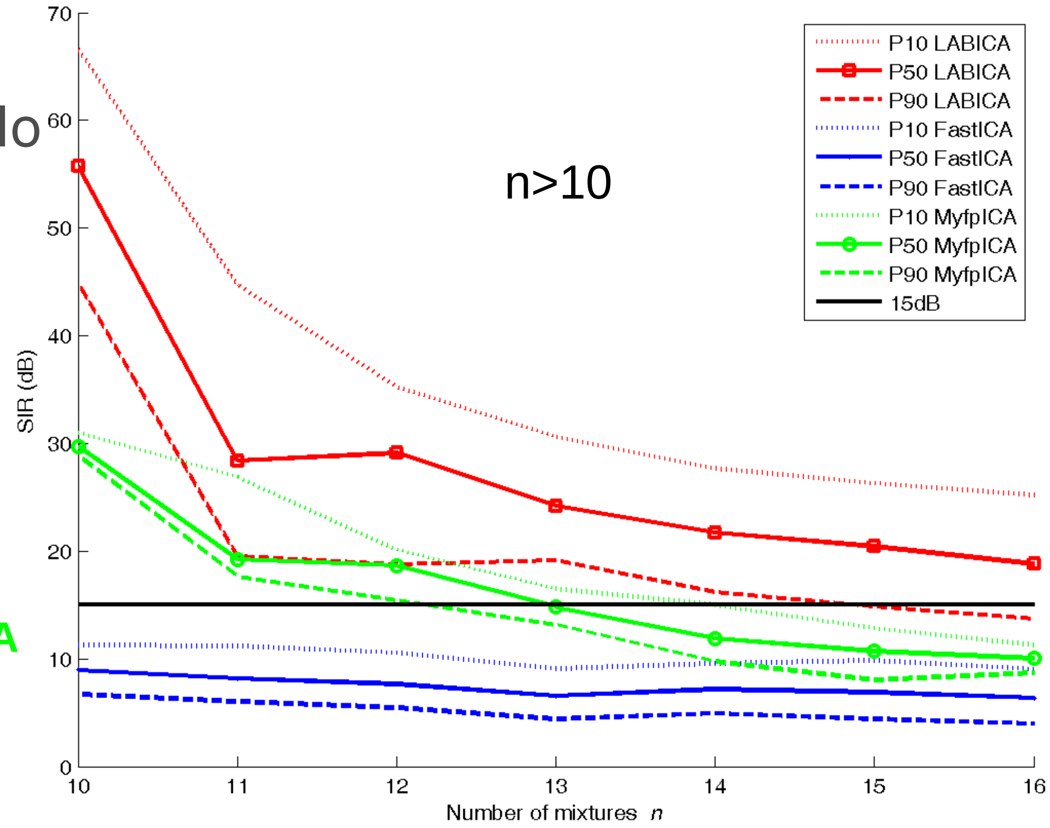
Less than 250 observations are needed to separe 50 mixtures ☺



Problem 3: Ill-conditioned Mixtures

- Mixing matrix = Hilbert times ramdom Givens
  - N=5000, n>2 (increasing number of sources)
- Results:
  - 100 Monte Carlo

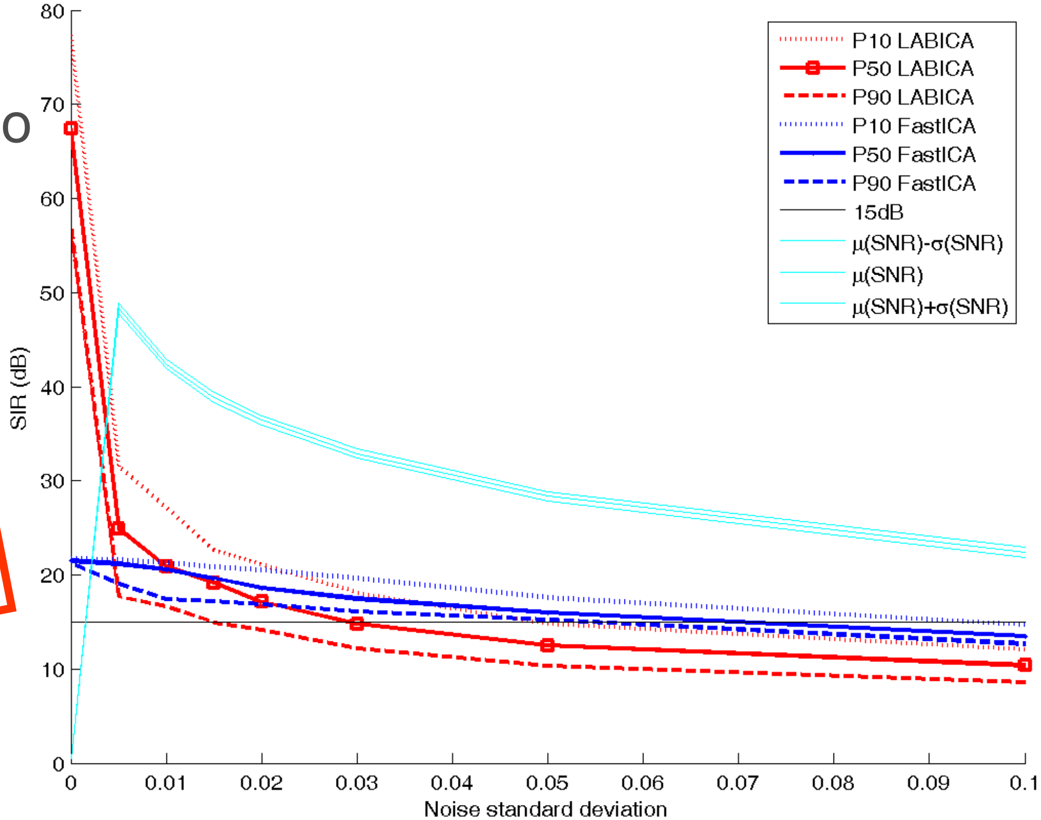
Whitening is of prime importance !!!



Myfpica = Fastica with
•Kurtosis-driven nonlin.
•Same whitening as LABICA

Problem 4: Noisy Mixtures

- Random mixing matrix + Gaussian noise
  - N=5000, n=50, increasing noise variance
- Results:
  - 100 Monte Carlo



LABICA is very sensitive to noise ☹ (pdf edges get 'blurred')

Conclusions

- Semi-blind better than blind : use the available knowledge
- Robust Whitening = important issue
  - use SVD instead of EVD
  - Smooth orthogonality constraint (avoid cumulation of errors)
- Accurate estimates of n-p srcs + bad estimates of p srcs beter than Approximative estimation of the n sources
- Non-differentiable contrasts can still efficiently be optimized
- ‘Difficult’ BSS problems can be solved because semi-blind
  - Many sources/mixtures
  - Few observations, Ill conditioned mixing matrices
  - Noise → robust method for edge estimation is a challenge
- Algo:
  - Refine and improve the optimization process (it was developed rather quickly, in the fever of the contest)

Thanks for your attention!

If you have any question, please ask…

More info on www.ucl.ac.be/mlg