Minimum Support ICA Using Order Statistics
Part II: Performance Analysis

Frédéric Vrins & Michel Verleysen
Université catholique de Louvain – Machine Learning Group
www.ucl.ac.be/mlg

Introduction

• Summary of Part I

• In Part I, the R, QR and AvQR-based support estimation have been investigated
• R is the best support convex hull estimator in noise-free case
• But not necessarilly in presence of a.o. outliers : only two extreme sample points are considered ! AvQR is thus better.

• Aim of part II

• For QR and AvQR: how to choose a meaningfull value for m given N without a priori knowledge on the source pdfs ?
• What about a performance analysis based on separation performances viewpoint ?

Choosing m given N : approach

• Idea

The estimation error must be small at least in probability, i.e. find the largest m such that the following relation holds for sufficiently small error threshold $\mathcal{E}$ and sufficiently close to but smaller than one probability threshold P_{thresh} :

$$\Pr\left(\mu[\Omega(Y_i)] - \langle R_m(Y_i)\rangle \leq \mathcal{E}\right) > P_{thresh} \quad (1)$$

• Method

• Problem: left hand side of (1) equals $1 - F_{\langle R_m(Y_i)\rangle}(\mu[\Omega(Y_i)] - \mathcal{E})$ which depends on the output pdf f_{Y_i}, i.e. on the unknown source pdfs !

• Solution: Find a distribution-free lower bound $\mathcal{L}$ for $\Pr(\cdot)$ in (1) and find the largest m such that $\mathcal{L} \geq P_{thresh}$

Choosing m given N : details

• Define the non-blind support estimation error based on quantile difference:

$$\mathcal{E} \doteq \mu[\Omega(X)] - (\xi_q - \xi_p) \quad (2)$$

⇨ p, q in [0,1] control the estimation error $\mathcal{E}$

•Observe that $\langle R_m(X)\rangle \geq R_m(X)\ w.p.\ one$

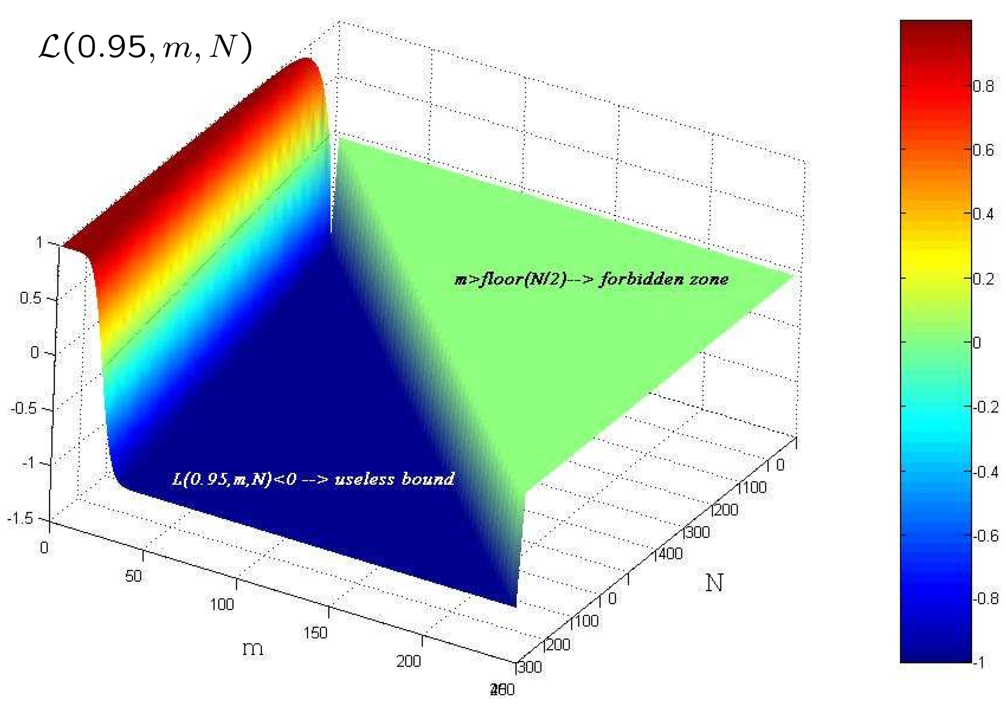
$$\begin{aligned}\Pr\left[\langle R_m(X)\rangle \geq \mu[\Omega(X)] - \mathcal{E}\right] &\geq \Pr\left[R_m(X) \geq \mu[\Omega(X)] - \mathcal{E}\right]\\ &= \Pr\left[R_m(X) \geq \xi_q - \xi_p\right]\\ &\geq \mathcal{L}(q,p,m,N)\end{aligned}$$

⇨ P_{thresh} controls confidence in support estimation

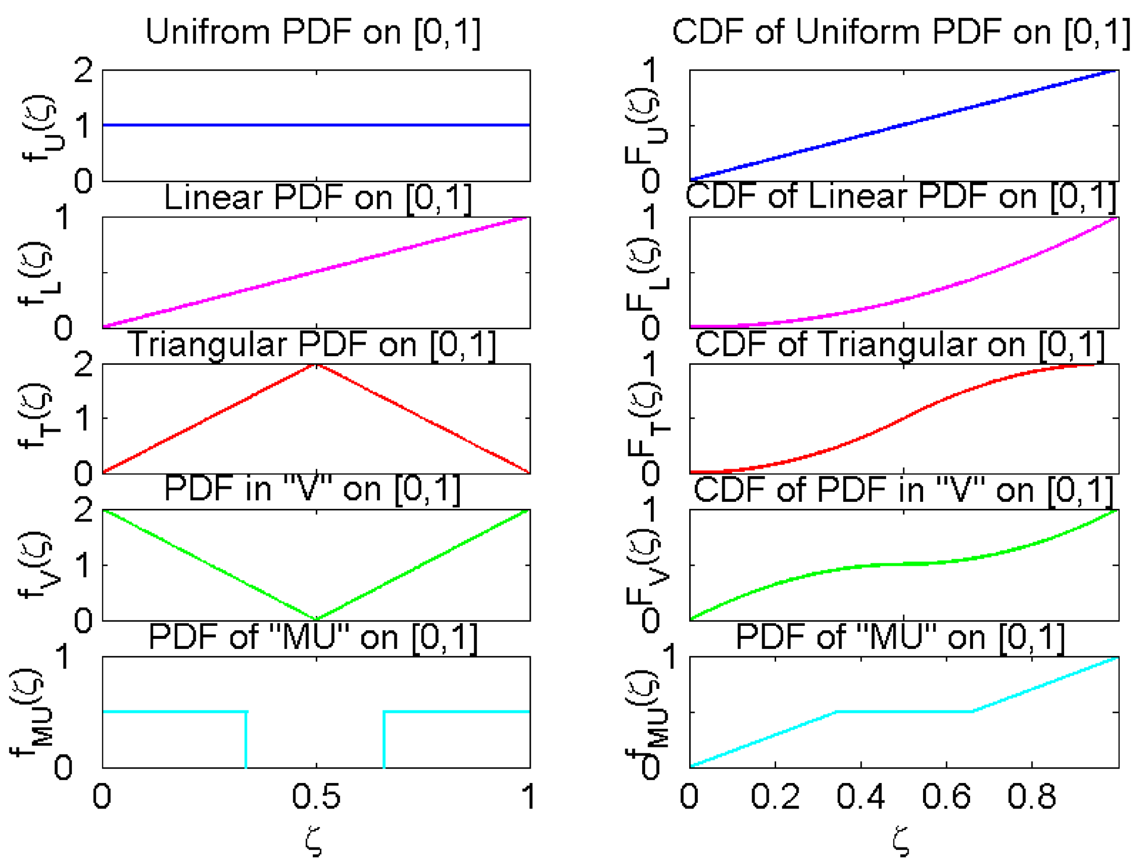
• Fix q close to one, $p = 1 - q$ and the probability threshold P_{thresh}
• Find the largest value m_0 of m ensuring that the lower bound $\mathcal{L}(\cdot)$ given by [Chu 1957] is higher than the desired threshold P_{thresh}
• Then (1) holds. This give confidence in support estimation !

Towards a probabilistic relation m(N)

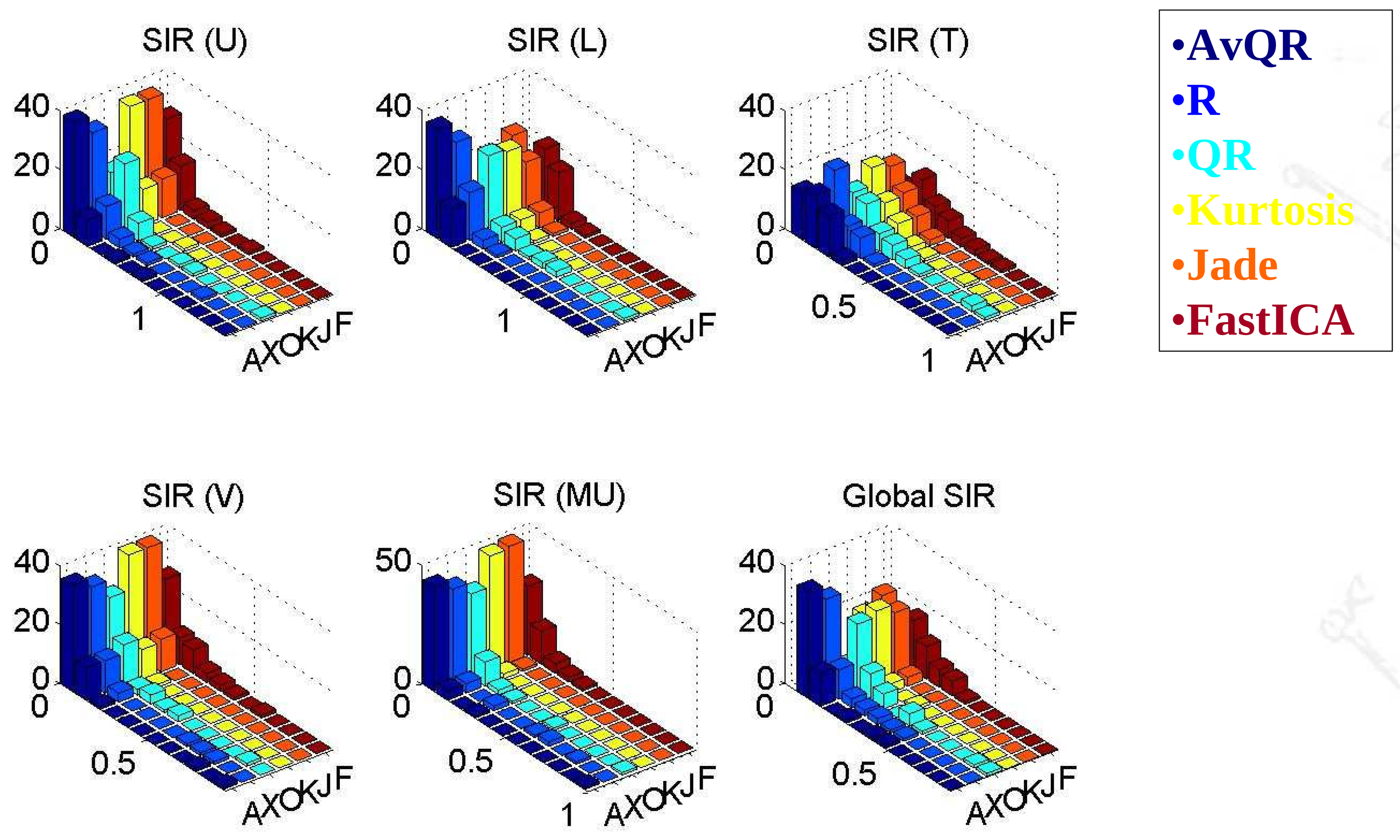
• If $\mathcal{L}(q, m_0, N) \geq P_{thresh}$ then (1) holds in the sense of (2) for all $m \leq m_0$
• Define $\mathcal{L}_+(q,m,N) = \max(0, \mathcal{L}(q,m,N))$



Source pdfs



Summary of results (m=m#(N))



Conclusion

• A blind and meaningfull relation $m = m^\sharp(N)$ can be provided and an empirical law can be derived (see paper for more details)
• With this rule, AvQR-ICA benefits from good separation performances

Main References

• Vrins et al., ICASSP 2005
• Chu, Ann. Math. Stat. 28, 1957
• David, "Order Statistics", 1970