SWM: a class of convex contrasts for source separation

Frédéric Vrins, Michel Verleysen
Machine Learning Group
Université catholique de Louvain – UCL
Belgium

Christian Jutten
Images and Signals Laboratory
INP Grenoble
France

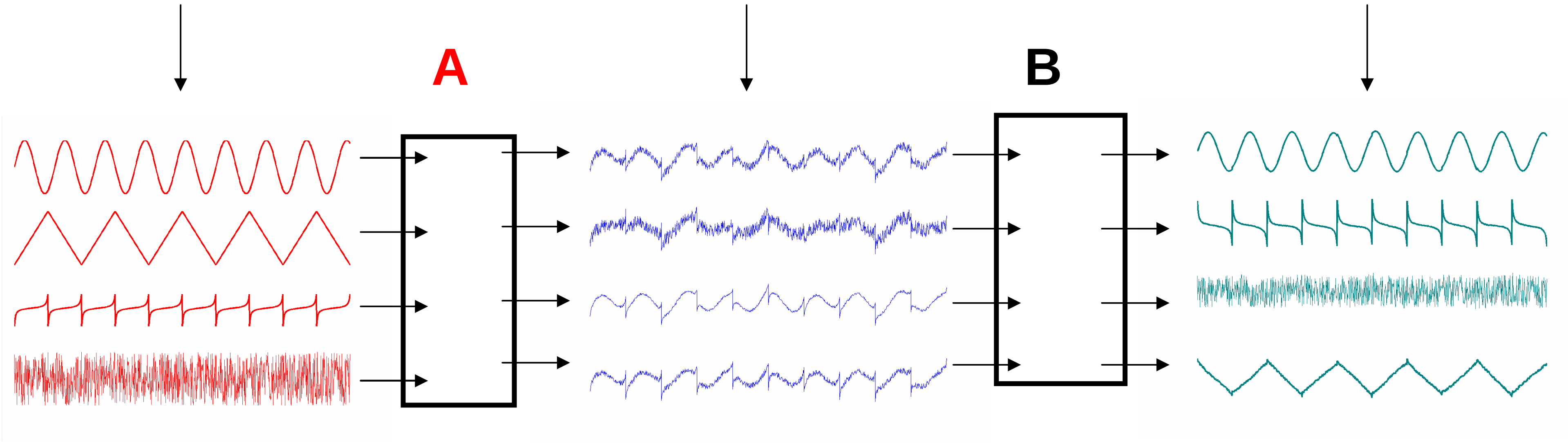
ICASSP 2005 (Philadelphia, USA) – March, 2005

Overview

- Introduction
- Contrast properties
- Spurious minima of popular contrasts in MM cases
- SWM: genesis of the criterion
- Discriminancy of the SWM contrast for bounded sources
- Practical considerations
- Conclusion and future work

Blind source separation: principle

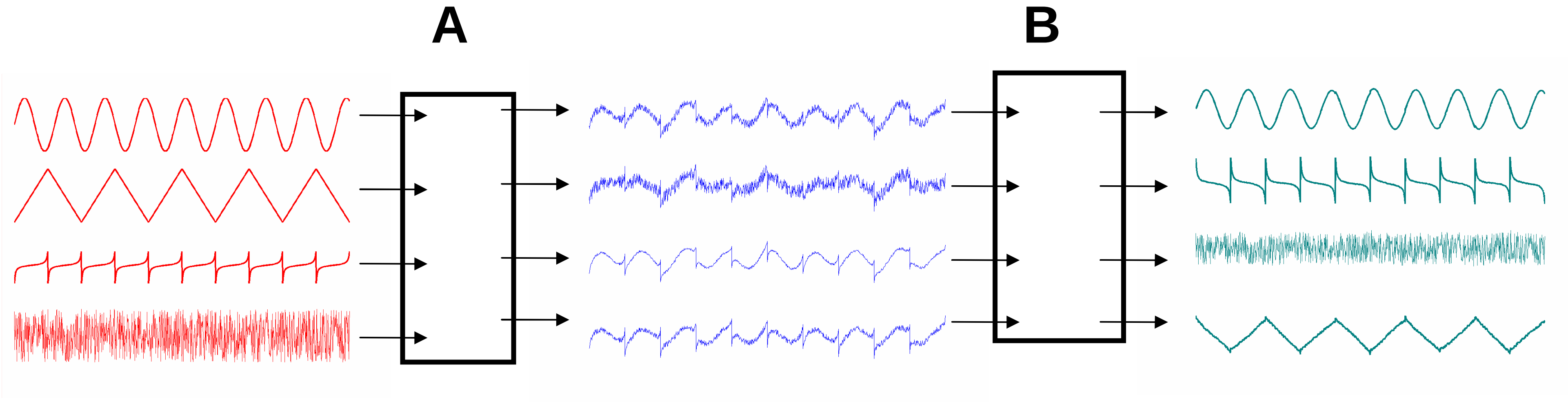
Sources **s** Mixtures **x** Outputs **y**



UNKNOWN **KNOWN** **TO ESTIMATE**

$$\mathbf{y} = \mathbf{BAs} = \mathbf{Cs}$$

Blind source separation



y=BAs=Cs

Aim: find **B*** such that **C*=B*A=PD** i.e. such that **C*** is non-mixing

Knowing: that the **s** are independent and **x=As**

Overview

- Introduction
- Contrast properties
- Spurious minima of popular contrasts in MM cases
- SWM: genesis of the criterion
- Discriminancy of the SWM contrast for bounded sources
- Practical considerations
- Conclusion and future work

Contrast properties

- The global maxima of a contrast corresponds to non-mixing matrices **C**: the sources are extracted
- In most of cases, gradient-based algorithms are used to maximize the contrast functions

→ one may be stucked in local maxima corresponding to mixing matrices **C**

- Hence: it is important to analyze the « discriminancy » property of a contrast :
  - Do all local maxima of the popular contrasts correspond to non-mixing **C** matrices ?

Overview

- Introduction
- Contrast properties
- Spurious minima of popular contrasts in MM cases
- SWM: genesis of the criterion
- Discriminancy of the SWM contrast for bounded sources
- Practical considerations
- Conclusion and future work

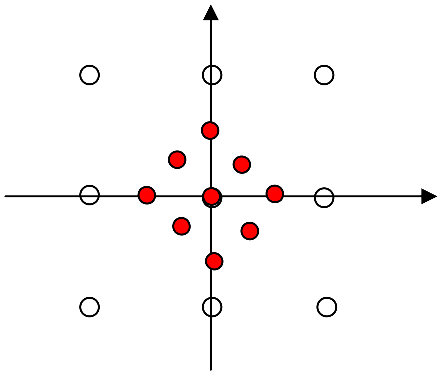
Spurious minima: ML contrast

✧ ML spurious maxima = KL local minima (local matching of output and source distributions)

[Cardoso, in Unsup. Adapt. Filt., 00]

$$\max_{\theta} E \log p_{\theta}(\mathbf{x}|\mathbf{x} = \mathbf{A}\mathbf{s}) \quad = \quad \min_{\theta} KL(p_X; p_{\theta})$$

Where we assume an exact model for the source distributions : $p_{\mathbf{s}} = \prod_i p_{s_i}$



○ Exact sources (model)
● Outputs (estimations)

Spurious minima: entropy-based contrast

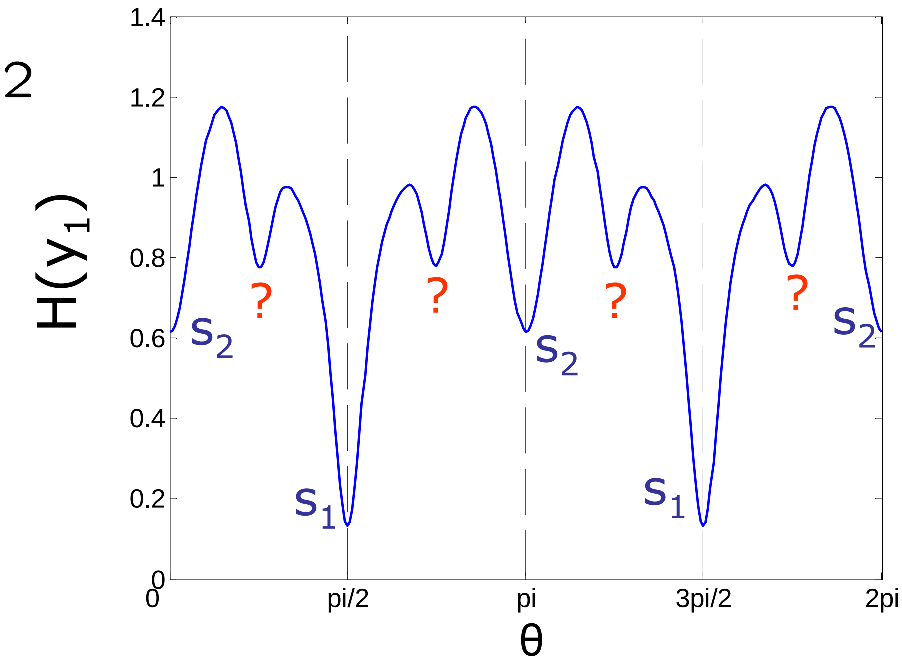
- ✧ -H(y) is a contrast → spurious entropy minima ?

$$y_1 = \sin\theta s_1 + \cos\theta s_2$$

$$H(s_1) \; < \; H(s_2)$$

$$E\{\mathbf{ss}^T\} = \mathbf{I}$$

S1 and s2 are bimodal



- ✧ Explaination ?

  → No comparison between densities (like in ML)!

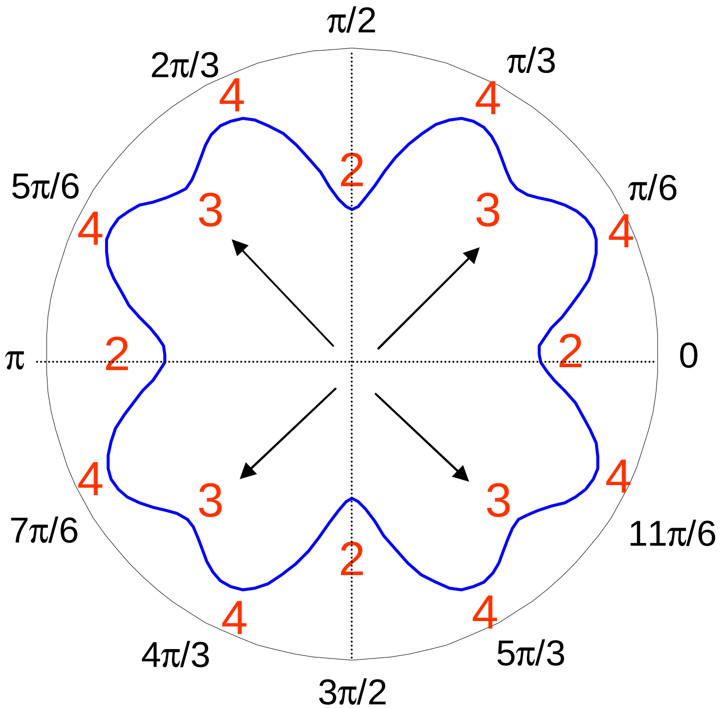
  → But the number of modes change !

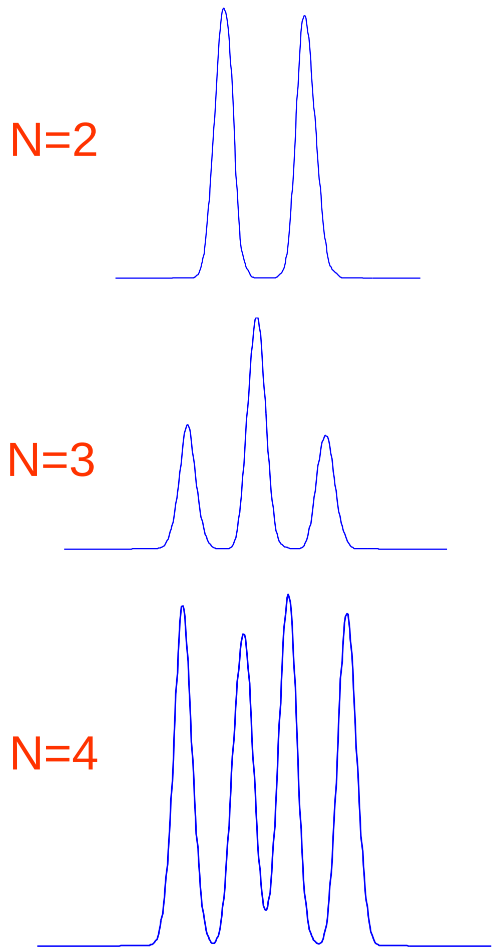
  Intuitive justification : [Vrins et al., Sig. Proc., In Press]

Marginal entropy and Modality

- A Link must exist !

H(sinθ X+cos θ Y) where pX=pY (N=2)





Similar results holds for MI with orthogonality constraint

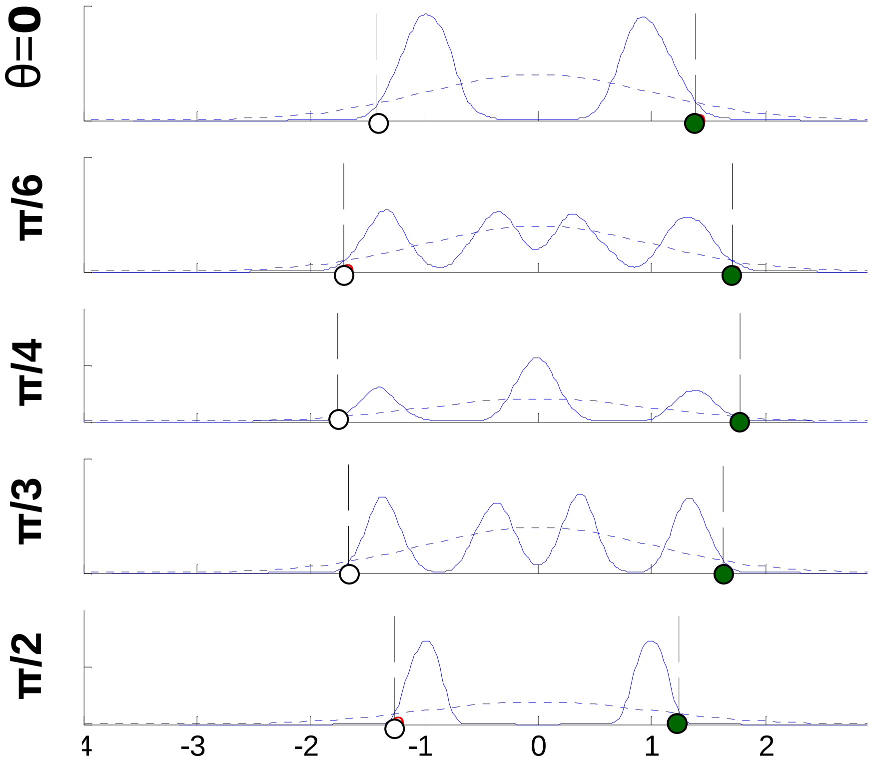
Overview

- Introduction
- Contrast properties
- Spurious minima of popular contrasts in MM cases
- SWM: genesis of the criterion
- Discriminancy of the SWM contrast for bounded sources
- Practical considerations
- Conclusion and future work

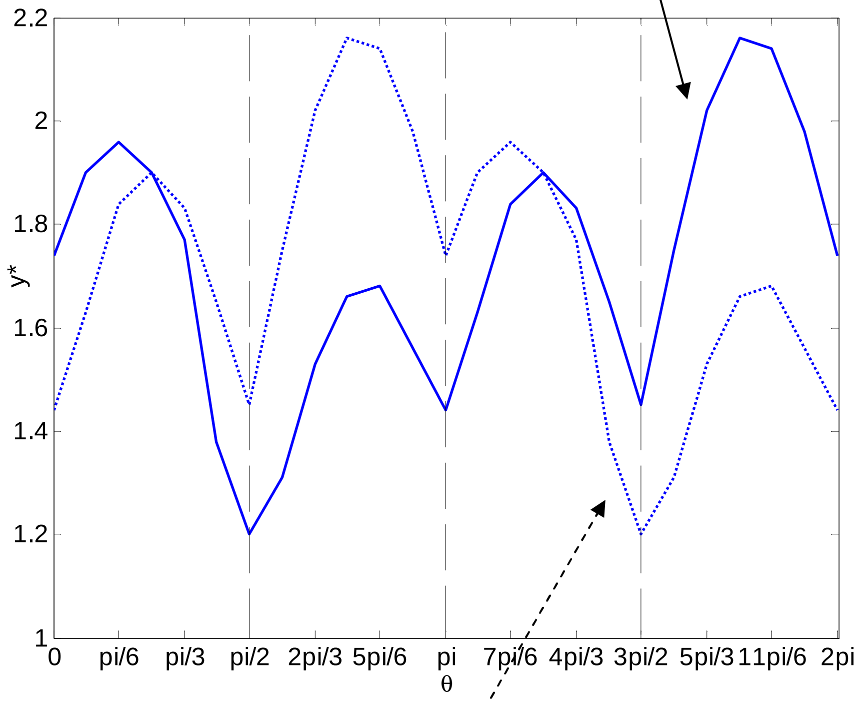
Support Width Measure genesis

Minus « support width » of an output pdf [Vrins et al., IEEE SPL, 05]

$$y_1 = \sin\theta s_1 + \cos\theta s_2$$



Evol. of right root ●



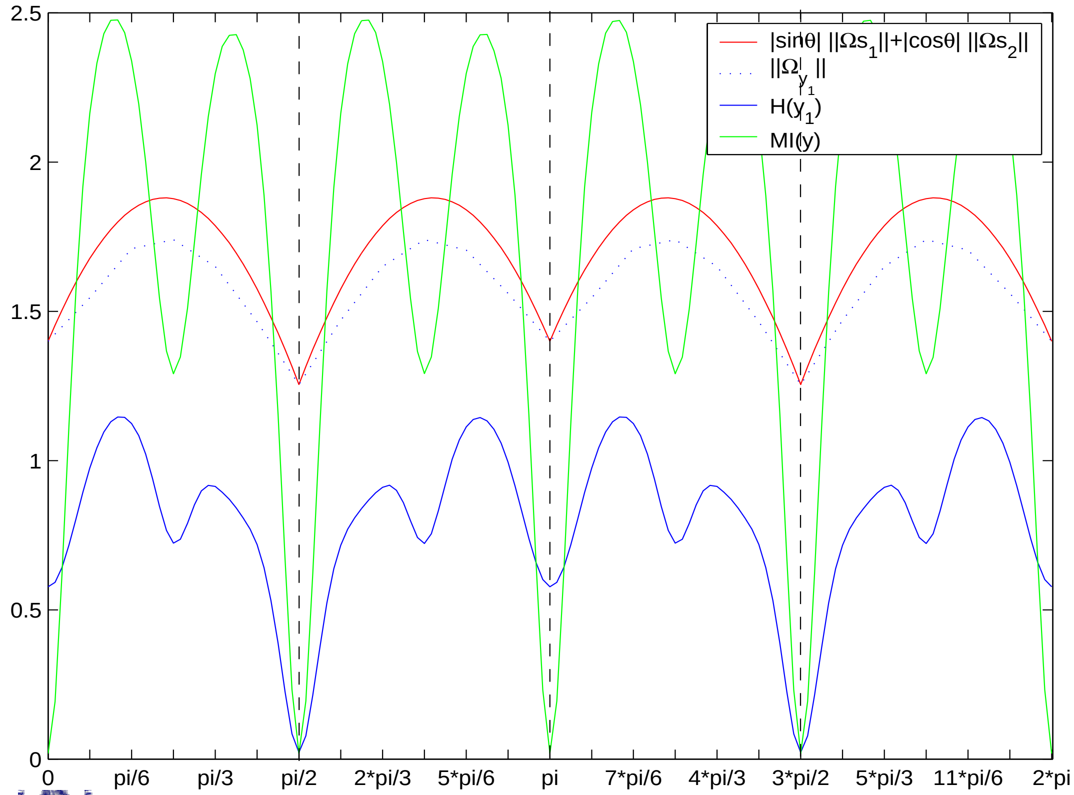
Evol. of abs. left root ○

Overview

- Introduction
- Contrast properties
- Spurious minima of popular contrasts in MM cases
- SWM: genesis of the criterion
- Discriminancy of the SWM contrast for bounded sources
- Practical considerations
- Conclusion and future work

Convexity of SWM in the bounded case

- A theoretical proof in the bounded sources case (see paper for a sketch of proof)
  - Using : $yi=ci1.s1+ci2.s2 \rightarrow ||p_{yi}||=||p_{ci1.s1}*p_{ci2.s2}||=||p_{ci1.s1}||+||p_{ci2.s2}||$



Mutual information

Entropy

|sinθ|.||s1||+|cos θ |.||s2||

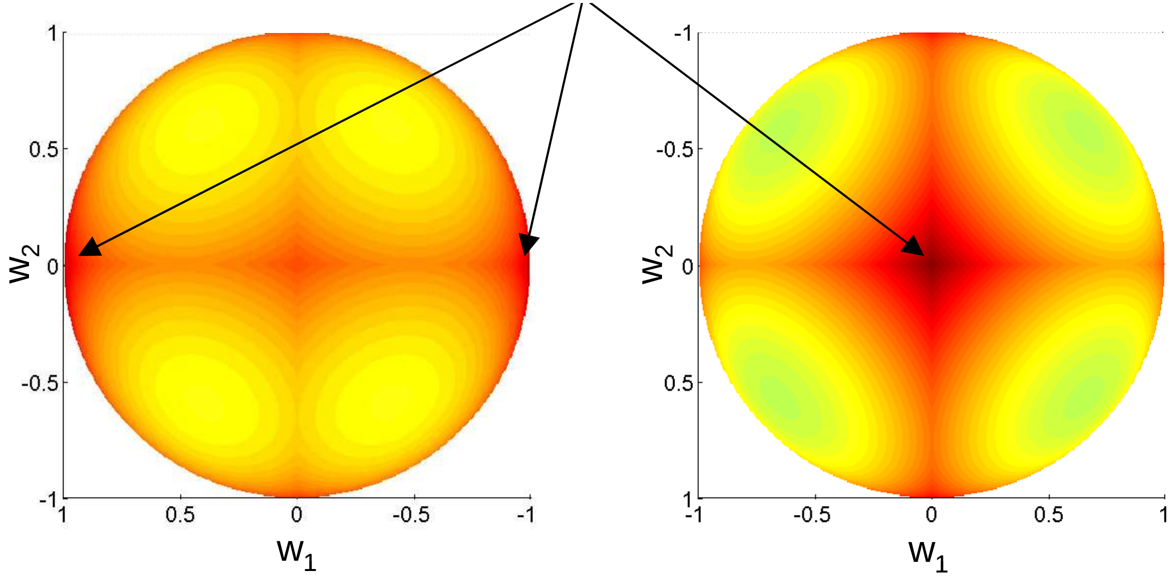
max|y1|-min|y1|

SWM is a discriminant contrast

- A simple 3D example

$$y_1 \;=\; w_1.s_1 + w_2.s_2 + w_3.s_3 \;\; s.t. \;\; \sum_i w_i^2 = 1$$

Global minima=lowest SW source (Orthog. Contrast : **Eyy**T=**I**)



(a) : ||- s_3||>||- s_2||>||- s_1||

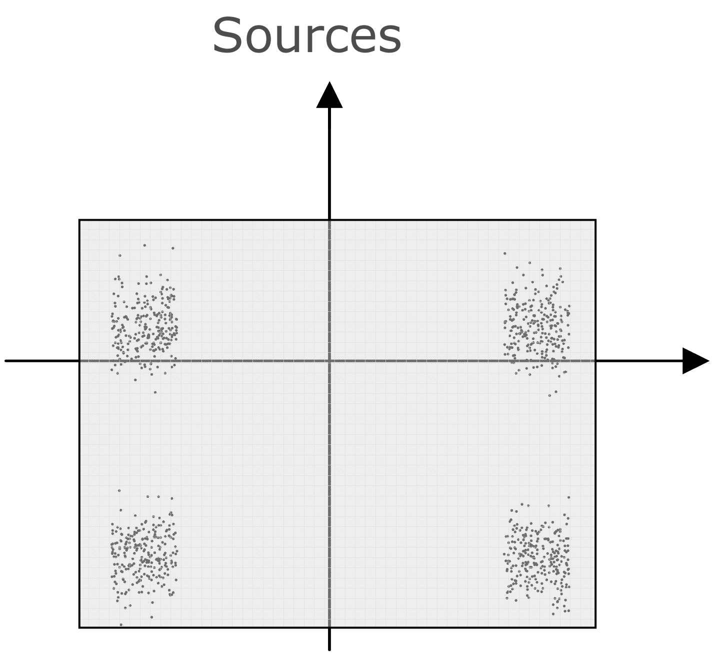
(b) : ||- s_1||>||- s_2||>||- s_3||

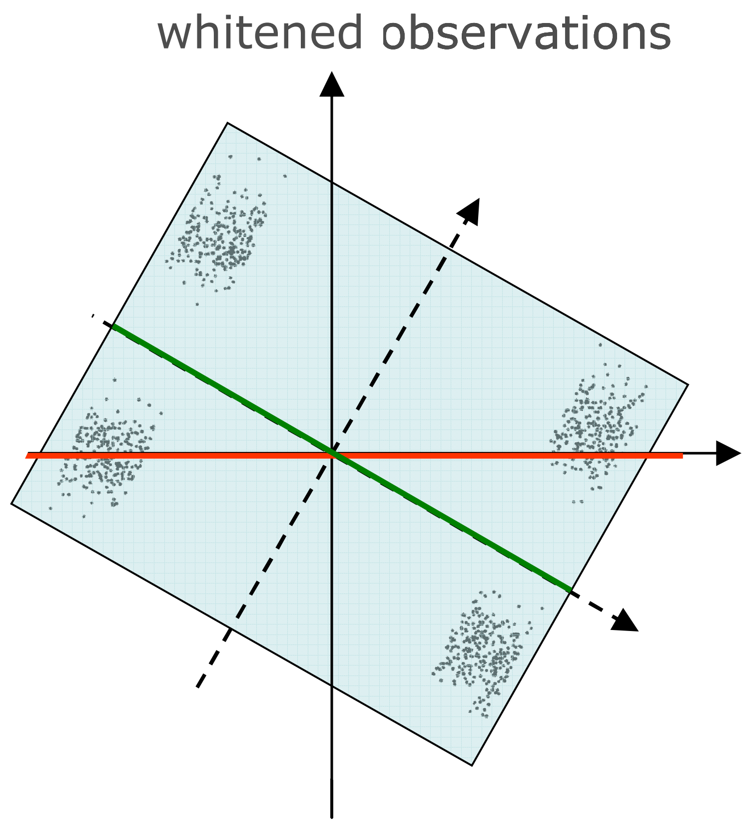
Overview

- Introduction
- Contrast properties
- Spurious minima of popular contrasts in MM cases
- SWM: genesis of the criterion
- Discriminancy of the SWM contrast for bounded sources
- Practical considerations
- Conclusion and future work

SWM : link to geometric ICA in 2D

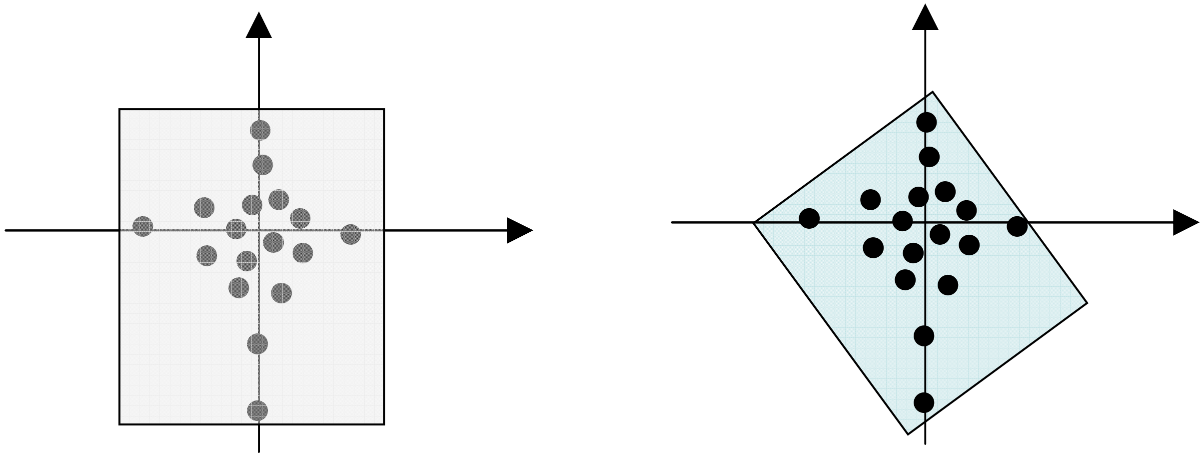
- If the observations are whitened, minimizing SW = finding the rotation that minimize the support of an a priori chosen ouput.





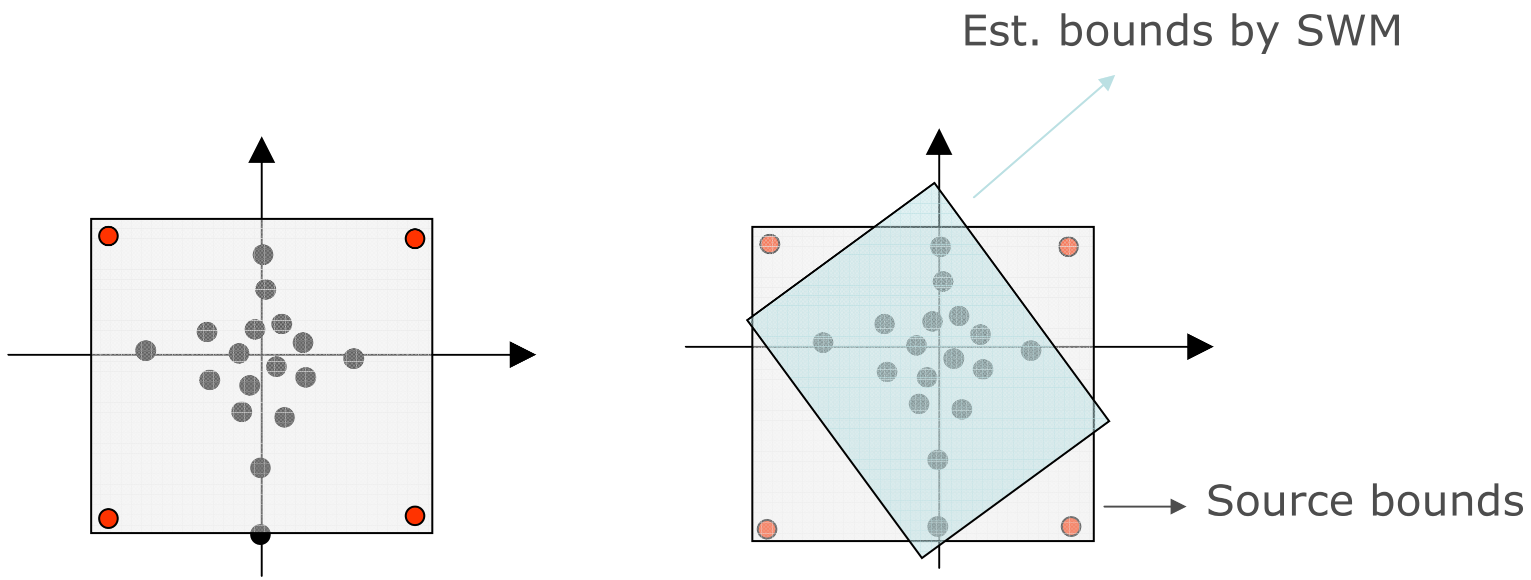
SWM : practical considerations

- A estimation of the SWM is required if sources unknown !
- SWM convex if SW correctly estimated
- In practice, SW estimation is:
  - Reliable if many points in the bound areas (precise estimation); this is the case for sub-Gaussian signals (e.g. sine wave)
  - NOT reliable if the sample points are mainly located in the 'center' of the ditribution; this is the case for sup-G signals.



SWM : practical considerations

✧ For super-Gaussian signals, essential information is missing most of time (less probable) than for sub-Gaussian signals (red points) !



Overview

- Introduction
- Contrast properties
- Spurious minima of popular contrasts in MM cases
- SWM: genesis of the criterion
- Discriminancy of the SWM contrast for bounded sources
- Practical considerations
- Conclusion and future work

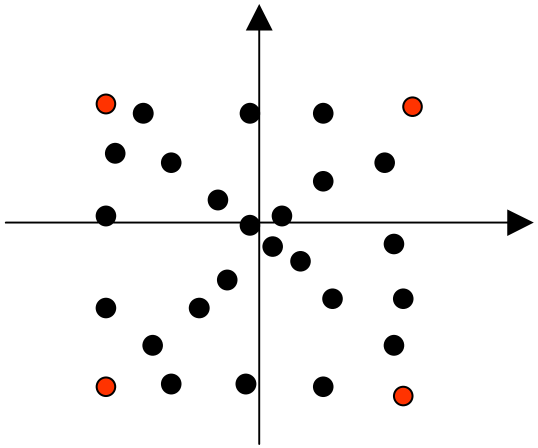
SWM : conclusion and future work

- Several popular BSS contrasts are not discriminant when dealing with mutlimodal sources (entropy, MI, Likelihood,…)
- SWM is a simple criterion which is discriminant when dealing with bounded sources: no local maxima exist (in theory)
- Few sample points are exploited by the criterion, which can cause a lack of robustness for sources having few points *near the border* of their pdf … However :

SWM : Further applications

- This apparent weakness may become in advantage, as e.g. for extraction of correlated grayscale images : the pdf are bounded in [0 255], and the pattern inside the distribution (e.g. the square) has no matter.

Specific structure inside the square → a dependence exist !



s1 and s2 are correlated,
but can be extracted using SWM
[Vrins et al., IWANN 05, To appear]