

# Traitement du signal avancé (SIC4102)

Marc Castella  
marc.castella@telecom-sudparis.eu

Télécom SudParis

February 5, 2024

# Part I

## Présentation du module

## Liste des séances programmées

- mer 07/02 10h-13h
- ven 09/02 10h-13h
- mer 28/02 10h-13h
- ven 01/03 10h-13h
- mer 13/03 10h-13h
- mar 20/03 10h-13h
- ven 22/03 10h-13h
- ven 29/03 10h-13h
- mer 03/04 10h-13h
- ven 05/04 10h-13h
- ven 03/05 14:30-16:30 : **contrôle final**
  - ▶ Questions de type QCM,
  - ▶ Exercices simples et questions de cours.
  - ▶ **Pas de CF2.**
  
- TPs: intégrés au cours + travail personnel en autonomie

# Prérequis du module

Liste incomplète de notions utiles:

- algèbre linéaire:
  - ▶ produit matriciel, bases, espace vectoriel, norme, produit scalaire et projection orthogonale
- théorie du signal:
  - ▶ filtre, convolution
  - ▶ signaux à temps discret
  - ▶ échantillonnage
- probabilités et statistiques:
  - ▶ signaux aléatoires, espace des variables aléatoires de carré sommable
  - ▶ bases du cours de statistiques
- bases de programmation Python Scientifique (environnement au choix: iPython, Notebook, Jupyter Lab, ...)

# Présentation du module

## Thème: **Reconstruction** de signaux dans les **modèles linéaires**

- Apprentissage **supervisé**
  - ▶ **approximation** de signaux (choix de critère, régularisation)
  - ▶ **estimation** (régularisation, interprétation bayésienne)
  - ▶ filtre inverse, **filtre de Wiener**; application en communications numériques (**égalisation**)
- Apprentissage **aveugle** et séparation de sources
  - ▶ analyse en composantes principales (PCA)
  - ▶ analyse en composantes indépendantes (**ICA**)
  - ▶ statistiques d'ordre supérieur (cumulants)
  - ▶ principes statistiques et critères de théorie de l'information
- Travaux pratiques en autonomie: modèles linéaires
- Compléments et rappels d'**optimisation** (selon besoins)
- Selon temps disponible: quelques éléments sur le cas non linéaire

## Gradient vector

Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ .

- **Gradient** (column) vector  $\nabla f(x)$ :

$$[\nabla f(x)]_i = \frac{\partial f(x)}{\partial x_i}$$

First-order approximation of  $f$  near  $\bar{x}$ :

$$\begin{aligned} f(x) &= f(\bar{x}) + \nabla f(\bar{x})^\top (x - \bar{x}) \\ &\quad + o(\|x - \bar{x}\|) \end{aligned}$$

- Ex:

$$f(x) = a^\top x$$

$$\nabla f(x) = a$$

$$g(x) = x^\top Mx$$

$$\nabla g(x) = (M + M^\top)x$$

$$= 2Mx \text{ if } M \text{ symmetric.}$$

## Hessian matrix

Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ .

- **Hessian matrix**  $\nabla^2 f(x)$

$$[\nabla^2 f(x)]_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}$$

Second-order approximation of  $f$  near  $\bar{x}$ :

$$f(x) = f(\bar{x}) + \nabla f(\bar{x})^\top (x - \bar{x}) + \frac{1}{2} (x - \bar{x})^\top \nabla^2 f(\bar{x}) (x - \bar{x}) + o(\|x - \bar{x}\|^2)$$

- Ex:

$$f(x) = a^\top x$$

$$\nabla^2 f(x) = 0$$

$$g(x) = x^\top M x$$

$$\nabla^2 g(x) = M + M^\top$$

=  $2M$  if  $M$  symmetric.

# Part II

## Linear models



## Approximation problem

Given a signal with samples  $\mathbf{y}$  in  $\mathbb{R}^m$ , find its best linear representation on a signal basis  $\mathbf{a}_1, \dots, \mathbf{a}_n$ :

$$\mathbf{y} \approx x_1 \mathbf{a}_1 + \dots + x_n \mathbf{a}_n$$

Norm minimization problem:

$$\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|$$

where  $\mathbf{A} = [\mathbf{a}_1 \ \dots \ \mathbf{a}_n] \in \mathbb{R}^{m \times n}$  with  $m \geq n$ .

Interpretations of  $\mathbf{x}^* = \operatorname{argmin}_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|$ :

- geometric:  $\mathbf{A}\mathbf{x}^*$  is the point in  $\operatorname{Span}(\mathbf{A})$  closest to  $\mathbf{y}$ .
- estimation: linear measurement  $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$  where  $\mathbf{n}$  is a measurement error or additive noise and  $\mathbf{x}$  unknown.
- optimal design/synthesis:  $\mathbf{x}$  are variables,  $\mathbf{A}\mathbf{x}$  is result;  $\mathbf{x}^*$  is design that gives best desired result  $\mathbf{y}$ .

# Norm minimization

$$\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|$$

with  $\mathbf{A} \in \mathbb{R}^{m \times n}$ ,  $m \geq n$ .

- $\|\cdot\|_2$  norm: least-squares approximation. Solution satisfies:

$$\mathbf{A}^\top \mathbf{A} \mathbf{x}^* = \mathbf{A}^\top \mathbf{y}$$

and if  $\text{rank } \mathbf{A} = n$ ,  $\mathbf{x}^* = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{y}$ .

- $\|\cdot\|_\infty$  norm: Chebyshev approximation.
- $\|\cdot\|_1$  norm: minimize sum of absolute residuals.

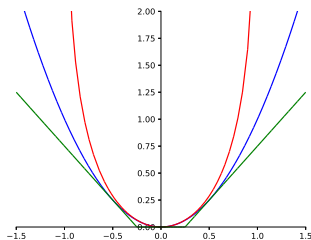
Rq: no analytic solution to last two problems, but can be solved as LPs.

# Penalty function approximation

$$\begin{aligned} \min. & \phi(r_1) + \cdots + \phi(r_m) \\ \text{s.t. } & \mathbf{r} = \mathbf{A}\mathbf{x} - \mathbf{y} \end{aligned}$$

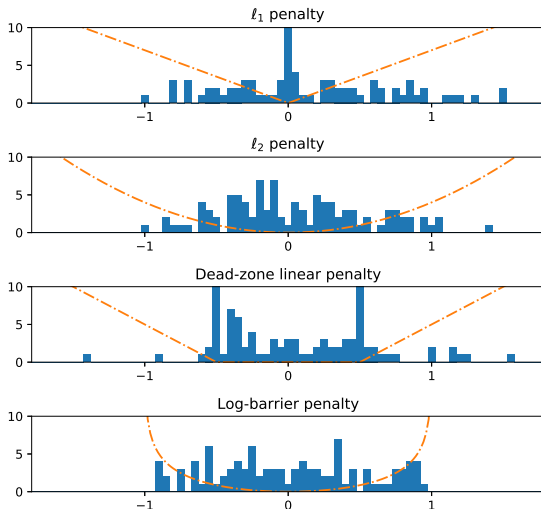
with  $\mathbf{A} \in \mathbb{R}^{m \times n}$  and  $\phi$  a convex penalty function.

- **quadratic**  $\phi(u) = u^2$
- **dead-zone linear**:  $\phi(u) = [|u| - \alpha]_+$
- **log-barrier**:  $\phi(u) = \begin{cases} -a^2 \log(1 - (\frac{u}{a})^2) & |u| < a, \\ +\infty & \text{otherwise.} \end{cases}$



# Penalty function approximation

Residuals histograms for the different penalties



## Robustness, robust regression

- Outlier: measurement  $y_i = \mathbf{a}_i^\top \mathbf{x} + n_i$  with large noise value  $n_i$ .
- Penalty function that avoid sensitivity to outliers:

$$\phi(u) = \begin{cases} u^2 & |u| < M, \\ M^2 & |u| \geq M. \end{cases}$$

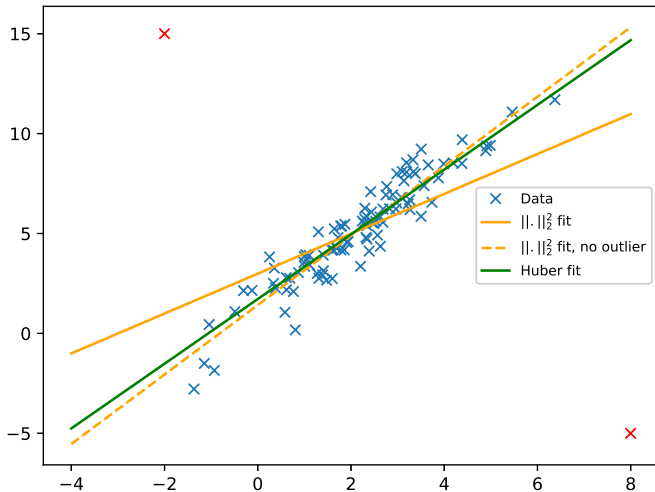
Non convex!

- **Huber** penalty function:

$$\phi_{\text{hub}}(u) = \begin{cases} u^2 & |u| < M, \\ M(2|u| - M) & |u| \geq M. \end{cases}$$

# Robust regression

## Example



# Least-norm problems

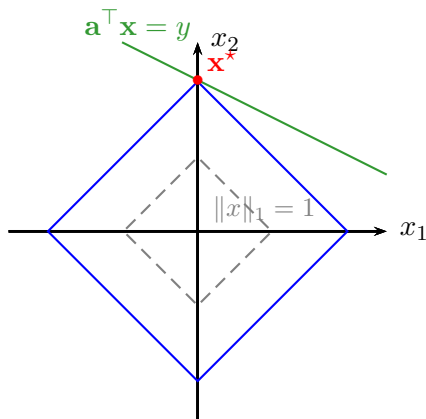
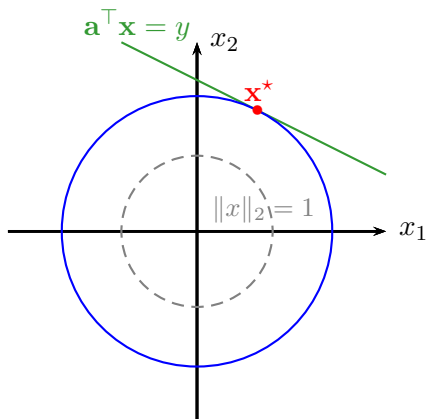
$$\begin{aligned} \min. \quad & \|\mathbf{x}\| \\ \text{s.t.} \quad & \mathbf{Ax} = \mathbf{y} \end{aligned}$$

with  $\mathbf{A} \in \mathbb{R}^{m \times n}$ ,  $m \leq n$ .

- $\|\cdot\|_2$ : least-squares solution of linear equations.  
can be solved via optimality conditions.
- $\|\cdot\|_1$ : minimum sum of absolute values: tends to produce a **sparse** solution.  
can be solved as an LP.

Comparison between  $\ell_2$  and  $\ell_1$  norms

$$\begin{aligned} \min. \quad & \|\mathbf{x}\| \\ \text{s.t.} \quad & \mathbf{a}^\top \mathbf{x} = y \end{aligned}$$





# Interpretation of $\ell_1$ as a convex relaxation

- Original  $\ell_0$  cardinality problem:

$$\left[ \begin{array}{l} \min. \|\mathbf{x}\|_0 \\ \text{s.t. } \|\mathbf{x}\|_\infty \leq R \\ \mathbf{x} \in \mathcal{C} \end{array} \right] \Leftrightarrow \left[ \begin{array}{l} \min. \mathbf{1}^\top z \\ \text{s.t. } |x_i| \leq Rz_i \\ x \in \mathcal{C}, \quad z_i \in \{0, 1\} \end{array} \right]$$

- Relaxation of  $\ell_0$  to  $\ell_1$ :

$$\left[ \begin{array}{l} \min. \mathbf{1}^\top z \\ \text{s.t. } |x_i| \leq Rz_i \\ x \in \mathcal{C}, \quad 0 \leq z_i \leq 1 \end{array} \right] \Leftrightarrow \left[ \begin{array}{l} \min. \frac{1}{R} \|\mathbf{x}\|_1 \\ \text{s.t. } x \in \mathcal{C} \end{array} \right]$$

## Iterated weighted $\ell_1$

- Cardinality problem:  $\min. \|\mathbf{x}\|_0$  s.t.  $\mathbf{x} \in \mathcal{C}$
- Iterated weighted  $\ell_1$ : Set  $\mathbf{w} = \mathbf{1}$ .  

$$\min. \|\text{diag}(\mathbf{w})\mathbf{x}\|_1 \quad \text{s.t. } \mathbf{x} \in \mathcal{C}$$

$$w_i := 1/(\epsilon + |x_i|)$$
- Interpretation for  $\mathbf{x} \succeq 0$  (no loss of generality:  $x = x^+ - x^-$ ):  
 Approximate  $\ell_0$  with  $\log(1 + \cdot)$

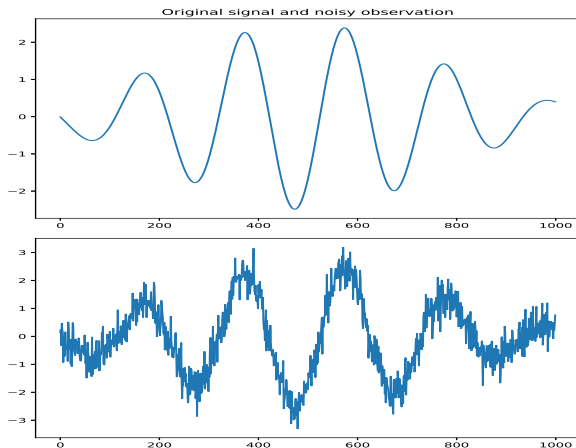
$$\min. \sum_{i=1}^n \log\left(1 + \frac{x_i}{\epsilon}\right) \quad \mathbf{x} \in \mathcal{C}, \mathbf{x} \succeq 0$$

and linearize at current iteration point:

$$\log\left(1 + \frac{x_i}{\epsilon}\right) \approx \log\left(1 + \frac{x_i^{(k)}}{\epsilon}\right) + \frac{x_i - x_i^{(k)}}{\epsilon + x_i^{(k)}}$$

# Signal reconstruction problem

- Observed signal:  $\mathbf{y} = \bar{\mathbf{x}} + \mathbf{n}$  where  $\mathbf{n}$  is noise.



- Objective: reconstruct the unknown signal  $\bar{\mathbf{x}} \in \mathbb{R}^n$

# Signal reconstruction

- Observed signal:  $\mathbf{y} = \bar{\mathbf{x}} + \mathbf{n}$  where  $\mathbf{n}$  is noise.
  - Objective: reconstruct the unknown signal  $\bar{\mathbf{x}} \in \mathbb{R}^n$
- Minimize two objectives (w.r.t.  $\mathbf{x}$ ):
- ▶ data fidelity term:  $\|\mathbf{x} - \mathbf{y}\|$
  - ▶ regularization term:  $\phi(\mathbf{x})$  where  $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$
- Penalized criterion minimization:  $\min. \|\mathbf{x} - \mathbf{y}\|_2^2 + \lambda\phi(\mathbf{x})$
- Examples:

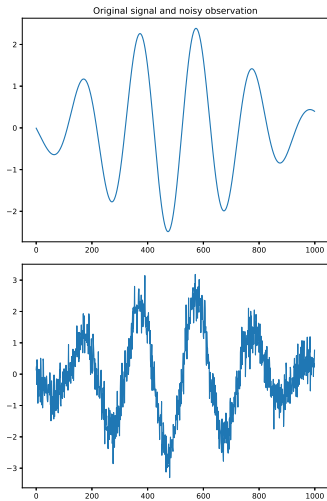
quadratic smoothing

$$\phi(\mathbf{x}) = \sum_{i=1}^{n-1} (x_{i+1} - x_i)^2$$

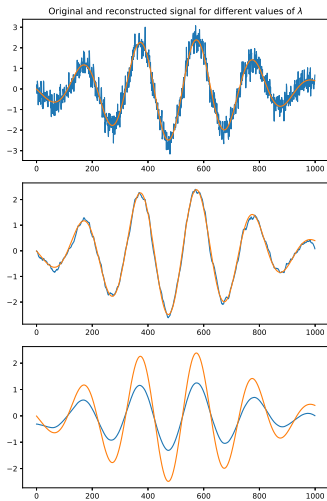
total variation denoising

$$\phi(\mathbf{x}) = \sum_{i=1}^{n-1} |x_{i+1} - x_i|$$

# Quadratic smoothing (example 1)

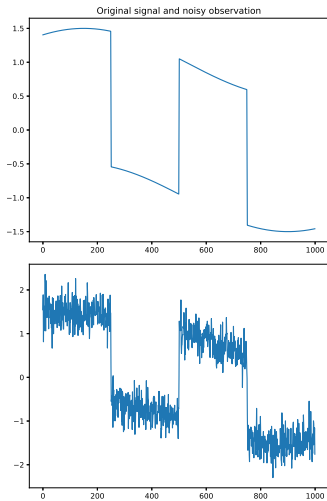


original and noisy signal

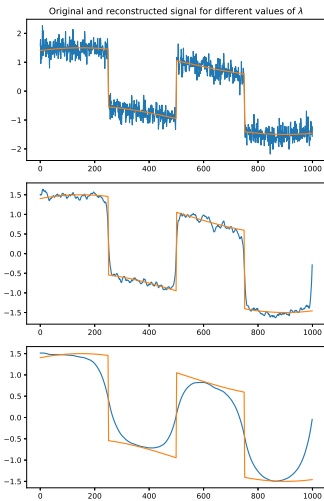


three reconstructed signals  
for different  $\lambda$

# Quadratic smoothing (example 2)

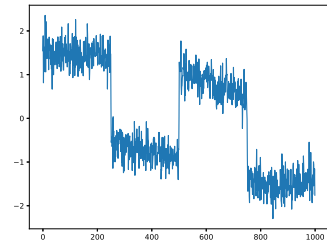
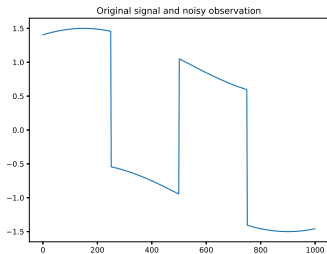


original and noisy signal

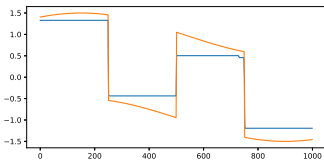
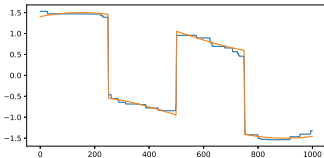
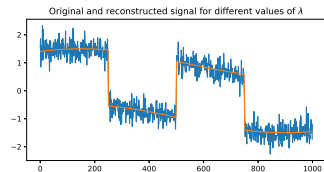


three reconstructed signals  
for different  $\lambda$

# Total variation denoising



original and noisy signal



three reconstructed signals  
for different  $\lambda$

## Maximum likelihood estimation

- Gaussian noise: least-squares, best linear unbiased estimator (BLUE)
- Laplacian noise:  $l_1$  norm approximation
- uniform noise:  $l_\infty$  norm

## Maximum a posteriori estimation

- Gaussian noise:
- Gaussian regularization: RIDGE estimator (link with MMSE/Wiener filtering)
- Laplace regularization:  $l_1$  prior term
- uniform prior:  $l_\infty$  constraint

## Perfect (non noisy) observations:

- Gaussian prior: least  $l_2$  solution
- Laplace prior: least  $l_1$  solution, link with basis pursuit/compressive sensing



# Maximum likelihood

## Parametric estimation

- Observation:  $\mathbf{y} \in \mathbb{R}^m$
- $\boldsymbol{\theta} \in \mathbb{R}^n$ ; unknown parameter to be estimated
- Parametric probability law:  $\mathbf{y} \sim p_{\boldsymbol{\theta}}(\mathbf{y})$

Maximum likelihood. For a given realization of  $\mathbf{y}$

- likelihood:  $\boldsymbol{\theta} \mapsto p_{\boldsymbol{\theta}}(\mathbf{y})$
- log-likelihood:  $\mathcal{L}(\boldsymbol{\theta}) = \log p_{\boldsymbol{\theta}}(\mathbf{y})$
- maximum likelihood estimator:

$$\begin{aligned}\hat{\boldsymbol{\theta}}_{\text{ML}} &= \text{Arg max}_{\boldsymbol{\theta}} p_{\boldsymbol{\theta}}(\mathbf{y}) \\ &= \text{Arg max}_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta})\end{aligned}$$

Remark: If above maximization is a convex problem, constraint  $\boldsymbol{\theta} \in C$  with  $C$  convex can be added.

# Observation model

Assumed model:

$$\mathbf{y} = \mathbf{A}\boldsymbol{\theta} + \mathbf{b} \quad \text{where:}$$

$$\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} \in \mathbb{R}^m \quad \mathbf{A} = \begin{bmatrix} \mathbf{a}_1^\top \\ \vdots \\ \mathbf{a}_m^\top \end{bmatrix} \in \mathbb{R}^{m \times n} \quad \mathbf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} \in \mathbb{R}^m, \text{ i.i.d.}$$

- ▶ **linear** model with **additive noise** reads also ( $\forall i = 1, \dots, m$ )  
 $y_i = \mathbf{a}_i^\top \boldsymbol{\theta} + b_i$  with  $b_i$  i.i.d. noise with pdf  $b_i \sim p_b(\cdot)$  and  $\mathbf{a}_i$  given.

- ▶ Likelihood:  $p_{\boldsymbol{\theta}}(\mathbf{y}) = \prod_{i=1}^m p_b(y_i - \mathbf{a}_i^\top \boldsymbol{\theta})$

- ▶ Log-likelihood:  $\mathcal{L}(\boldsymbol{\theta}) = \sum_{i=1}^m \log p_b(y_i - \mathbf{a}_i^\top \boldsymbol{\theta})$

## Gaussian noise

Assume noise  $\sim \mathcal{N}(0, \sigma^2)$ , that is  $p_b(\cdot) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\cdot)^2}{2\sigma^2}}$ .

▷ Log-likelihood:

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta}) &= \sum_{i=1}^m \log p_b(y_i - \mathbf{a}_i^\top \boldsymbol{\theta}) \\ &= \sum_{i=1}^m \left[ -\frac{1}{2} \log(2\pi\sigma^2) - \frac{(y_i - \mathbf{a}_i^\top \boldsymbol{\theta})^2}{2\sigma^2} \right] \\ &= \underbrace{-\frac{m}{2} \log(2\pi\sigma^2)}_{\text{cst.}} - \frac{1}{2\sigma^2} \underbrace{\sum_{i=1}^m (y_i - \mathbf{a}_i^\top \boldsymbol{\theta})^2}_{\|\mathbf{y} - \mathbf{A}\boldsymbol{\theta}\|_2^2} \end{aligned}$$

▷ Maximum likelihood estimator is least-squares / least  $\ell_2$  norm approximation:

$$\hat{\boldsymbol{\theta}}_{\text{ML}} = \text{Arg min}_{\boldsymbol{\theta}} \|\mathbf{y} - \mathbf{A}\boldsymbol{\theta}\|_2^2 := \hat{\boldsymbol{\theta}}_{\text{ls}}$$

See properties of least-squares / "BLUE" estimator in exercise.

## Laplace noise

Assume noise has density  $p_b(\cdot) = \frac{1}{2\alpha} e^{-\frac{|\cdot|}{\alpha}}$  ( $\alpha > 0$  given)

▶ Log-likelihood:

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta}) &= \sum_{i=1}^m \log p_b(y_i - \mathbf{a}_i^\top \boldsymbol{\theta}) \\ &= \sum_{i=1}^m \left[ -\log(2\alpha) - \frac{|y_i - \mathbf{a}_i^\top \boldsymbol{\theta}|}{\alpha} \right] \\ &= \underbrace{-m \log(2\alpha)}_{\text{cst.}} - \frac{1}{\alpha} \underbrace{\sum_{i=1}^m |y_i - \mathbf{a}_i^\top \boldsymbol{\theta}|}_{\|\mathbf{y} - \mathbf{A}\boldsymbol{\theta}\|_1} \end{aligned}$$

▶ Maximum likelihood estimator is least  $\ell_1$  norm approximation:

$$\hat{\boldsymbol{\theta}}_{\text{ML}} = \text{Arg min}_{\boldsymbol{\theta}} \|\mathbf{y} - \mathbf{A}\boldsymbol{\theta}\|_1$$

## Uniform noise

Assume uniform noise with density  $p_b(\cdot) = \frac{1}{2\alpha} \mathbb{1}_{[-\alpha, \alpha]}(\cdot)$  ( $\alpha > 0$  given)

▷ Likelihood:

$$\begin{aligned} p_{\boldsymbol{\theta}}(\mathbf{y}) &= \prod_{i=1}^m \frac{1}{2\alpha} \mathbb{1}_{[-\alpha, \alpha]}(y_i - \mathbf{a}_i^{\top} \boldsymbol{\theta}) \\ &= \begin{cases} \frac{1}{(2\alpha)^m} & \forall i, -\alpha \leq y_i - \mathbf{a}_i^{\top} \boldsymbol{\theta} \leq \alpha \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

▷ Maximum likelihood estimator such that  $-\alpha \leq y_i - \mathbf{a}_i^{\top} \boldsymbol{\theta} \leq \alpha$  for all  $i = 1, \dots, m$  is given by  $\ell_{\infty}$  norm minimization:

$$\begin{aligned} \hat{\boldsymbol{\theta}}_{\text{ML}} &= \text{Arg min}_{\boldsymbol{\theta}} \|\mathbf{y} - \mathbf{A}\boldsymbol{\theta}\|_{\infty} \\ &\text{given by: } \|\mathbf{y} - \mathbf{A}\hat{\boldsymbol{\theta}}_{\text{ML}}\|_{\infty} \leq \alpha \end{aligned}$$

# Maximum a posteriori

- Observation:  $\mathbf{y} \in \mathbb{R}^m$
- $\boldsymbol{\theta} \in \mathbb{R}^n$ ; unknown parameter to be estimated is a **random variable**.

Prior density:  $p(\boldsymbol{\theta}) = \int p(\boldsymbol{\theta}, \mathbf{y}) d\mathbf{y}$

- Posterior density:  $p(\boldsymbol{\theta}/\mathbf{y}) = \frac{p(\boldsymbol{\theta}, \mathbf{y})}{p(\mathbf{y})} = \frac{p_{\boldsymbol{\theta}}(\boldsymbol{\theta})p(\mathbf{y}/\boldsymbol{\theta})}{p_{\mathbf{y}}(\mathbf{y})}$
- Maximum a posteriori. For a given realization of  $\mathbf{y}$ :

$$\begin{aligned}\hat{\boldsymbol{\theta}}_{\text{MAP}} &= \text{Arg max}_{\boldsymbol{\theta}} p(\boldsymbol{\theta}/\mathbf{y}) \\ &= \text{Arg max}_{\boldsymbol{\theta}} p(\boldsymbol{\theta})p(\mathbf{y}/\boldsymbol{\theta}) \quad \text{taking log:}\end{aligned}$$

$$\hat{\boldsymbol{\theta}}_{\text{MAP}} = \text{Arg max}_{\boldsymbol{\theta}} \left[ \log p(\boldsymbol{\theta}) + \log p(\mathbf{y}/\boldsymbol{\theta}) \right]$$

# Maximum a posteriori

- Observation:  $\mathbf{y} \in \mathbb{R}^m$
- $\boldsymbol{\theta} \in \mathbb{R}^n$ ; unknown parameter to be estimated is a **random variable**.

Prior density:  $p(\boldsymbol{\theta}) = \int p(\boldsymbol{\theta}, \mathbf{y}) d\mathbf{y}$

- Posterior density:  $p(\boldsymbol{\theta}/\mathbf{y}) = \frac{p(\boldsymbol{\theta}, \mathbf{y})}{p(\mathbf{y})} = \frac{p_{\boldsymbol{\theta}}(\boldsymbol{\theta})p(\mathbf{y}/\boldsymbol{\theta})}{p_{\mathbf{y}}(\mathbf{y})}$
- Maximum a posteriori. For a given realization of  $\mathbf{y}$ :

$$\begin{aligned}\hat{\boldsymbol{\theta}}_{\text{MAP}} &= \text{Arg max}_{\boldsymbol{\theta}} p(\boldsymbol{\theta}/\mathbf{y}) \\ &= \text{Arg max}_{\boldsymbol{\theta}} p(\boldsymbol{\theta})p(\mathbf{y}/\boldsymbol{\theta}) \quad \text{taking log:} \\ \hat{\boldsymbol{\theta}}_{\text{MAP}} &= \text{Arg max}_{\boldsymbol{\theta}} \left[ \log p(\boldsymbol{\theta}) + \log p(\mathbf{y}/\boldsymbol{\theta}) \right]\end{aligned}$$

- $p(\mathbf{y}/\boldsymbol{\theta}) \leftrightarrow p_{\boldsymbol{\theta}}(\mathbf{y})$  is the likelihood term.

# Observation model

Assumed model:

$$\mathbf{y} = \mathbf{A}\boldsymbol{\theta} + \mathbf{b} \quad \text{where:}$$

$$\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} \in \mathbb{R}^m \quad \mathbf{A} = \begin{bmatrix} \mathbf{a}_1^\top \\ \vdots \\ \mathbf{a}_m^\top \end{bmatrix} \in \mathbb{R}^{m \times n} \quad \mathbf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} \in \mathbb{R}^m, \text{ i.i.d.}$$

- ▶ **linear model with independent additive noise** reads also  $(\forall i = 1, \dots, m) y_i = \mathbf{a}_i^\top \boldsymbol{\theta} + b_i$  with  $b_i$  i.i.d. noise with pdf  $b_i \sim p_b(\cdot)$  and  $\mathbf{a}_i$  given.

- ▶ Joint distribution: 
$$p(\boldsymbol{\theta}, \mathbf{y}) = p(\boldsymbol{\theta}) \prod_{i=1}^m p_b(y_i - \mathbf{a}_i^\top \boldsymbol{\theta})$$

- ▶ 
$$\hat{\boldsymbol{\theta}}_{\text{MAP}} = \text{Arg max}_{\boldsymbol{\theta}} \left[ \log p(\boldsymbol{\theta}) + \sum_{i=1}^m \log p_b(y_i - \mathbf{a}_i^\top \boldsymbol{\theta}) \right]$$



# Observation model

Assumed model:

$$\mathbf{y} = \mathbf{A}\boldsymbol{\theta} + \mathbf{b} \quad \text{where:}$$

$$\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} \in \mathbb{R}^m \quad \mathbf{A} = \begin{bmatrix} \mathbf{a}_1^\top \\ \vdots \\ \mathbf{a}_m^\top \end{bmatrix} \in \mathbb{R}^{m \times n} \quad \mathbf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} \in \mathbb{R}^m, \text{ i.i.d.}$$

- ▶ **linear model with independent additive noise** reads also  $(\forall i = 1, \dots, m) y_i = \mathbf{a}_i^\top \boldsymbol{\theta} + b_i$  with  $b_i$  i.i.d. noise with pdf  $b_i \sim p_b(\cdot)$  and  $\mathbf{a}_i$  given.

- ▶ Joint distribution:  $p(\boldsymbol{\theta}, \mathbf{y}) = p(\boldsymbol{\theta}) \prod_{i=1}^m p_b(y_i - \mathbf{a}_i^\top \boldsymbol{\theta})$

- ▶  $\hat{\boldsymbol{\theta}}_{\text{MAP}} = \text{Arg max}_{\boldsymbol{\theta}} \left[ \log p(\boldsymbol{\theta}) + \sum_{i=1}^m \log p_b(y_i - \mathbf{a}_i^\top \boldsymbol{\theta}) \right]$

## Gaussian prior: RIDGE estimator

- Assumption:
- noise  $\mathbf{b} \sim \mathcal{N}(\mathbf{0}, \sigma_b^2 \mathbf{I})$
  - prior  $\boldsymbol{\theta} \sim \mathcal{N}(\mathbf{0}, \sigma_\theta^2 \mathbf{I})$

$$p(\boldsymbol{\theta}, \mathbf{y}) \propto \exp \left[ -\frac{\|\boldsymbol{\theta}\|_2^2}{2\sigma_\theta^2} \right] \exp \left[ -\frac{\|\mathbf{y} - \mathbf{A}\boldsymbol{\theta}\|_2^2}{2\sigma_b^2} \right]$$

$$\hat{\boldsymbol{\theta}}_{\text{MAP}} = \text{Arg min}_{\boldsymbol{\theta}} \left[ \|\mathbf{y} - \mathbf{A}\boldsymbol{\theta}\|_2^2 + \frac{\sigma_b^2}{\sigma_\theta^2} \|\boldsymbol{\theta}\|_2^2 \right]$$

Defining,  $\lambda = \frac{\sigma_b^2}{\sigma_\theta^2}$  above estimator known as «RIDGE»:

$$\hat{\boldsymbol{\theta}}_{\text{ridge}} = \text{Arg min}_{\boldsymbol{\theta}} \left[ \|\mathbf{y} - \mathbf{A}\boldsymbol{\theta}\|_2^2 + \lambda \|\boldsymbol{\theta}\|_2^2 \right]$$

Above cost is the sum of:

- data fidelity term  $\leftrightarrow$  fit to the data
- regularization / penalization term  $\leftrightarrow$  prior distribution / information

See properties of RIDGE compared to least-squares in exercise.

## Double exponential prior: LASSO estimator

Assumption: • noise  $\mathbf{b} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$

• prior  $\boldsymbol{\theta} \sim \prod_{j=1}^n \frac{1}{2\alpha} e^{-\frac{|\theta_j|}{2\alpha}}$  (iid + Laplace)

$$p(\boldsymbol{\theta}, \mathbf{y}) \propto \exp \left[ -\frac{1}{2\alpha} \sum_{j=1}^n |\theta_j| \right] \exp \left[ -\frac{\|\mathbf{y} - \mathbf{A}\boldsymbol{\theta}\|_2^2}{2\sigma_b^2} \right]$$

$$\hat{\boldsymbol{\theta}}_{\text{MAP}} = \text{Arg min}_{\boldsymbol{\theta}} \left[ \|\mathbf{y} - \mathbf{A}\boldsymbol{\theta}\|_2^2 + \frac{\sigma_b^2}{\alpha} \|\boldsymbol{\theta}\|_1 \right]$$

Defining,  $\lambda = \frac{\sigma_b^2}{\alpha}$  above estimator known as «LASSO»:

$$\hat{\boldsymbol{\theta}}_{\text{ridge}} = \text{Arg min}_{\boldsymbol{\theta}} \left[ \|\mathbf{y} - \mathbf{A}\boldsymbol{\theta}\|_2^2 + \lambda \|\boldsymbol{\theta}\|_1 \right]$$

Above cost is the sum of:

- data fidelity term  $\leftrightarrow$  fit to the data
  - regularization / penalization term  $\leftrightarrow$  prior distribution / information
- $\ell_1$ -norm promotes **sparse** solutions

# Uniform prior

- Assumption:
- noise  $\mathbf{b} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$
  - prior  $\boldsymbol{\theta} \sim \prod_{j=1}^n \frac{1}{2\alpha} \mathbb{1}_{[-\alpha, \alpha]}(\theta_j)$  (iid + uniform)

$$p(\boldsymbol{\theta}, \mathbf{y}) \propto \prod_{j=1}^n \mathbb{1}_{[-\alpha, \alpha]}(\theta_j) \exp \left[ -\frac{\|\mathbf{y} - \mathbf{A}\boldsymbol{\theta}\|_2^2}{2\sigma_b^2} \right]$$

$$\hat{\boldsymbol{\theta}}_{\text{MAP}} = \underset{\|\boldsymbol{\theta}\|_\infty \leq \alpha}{\text{Arg min}} \|\mathbf{y} - \mathbf{A}\boldsymbol{\theta}\|_2^2$$

$\hat{\boldsymbol{\theta}}_{\text{MAP}}$  is constrained least-squares

## Perfect linear observations

Assumption:

- **no noise:**  $\mathbf{y} = \mathbf{A}\boldsymbol{\theta}$        $\mathbf{A} \in \mathbb{R}^{m \times n}$
- prior  $\boldsymbol{\theta} \sim p_{\boldsymbol{\theta}}(\boldsymbol{\theta})$

If  $m \geq n$ , more equations than unknowns, at most one solution to  $\mathbf{y} = \mathbf{A}\boldsymbol{\theta}$ .

If  $m < n$ , less equations than unknowns and:

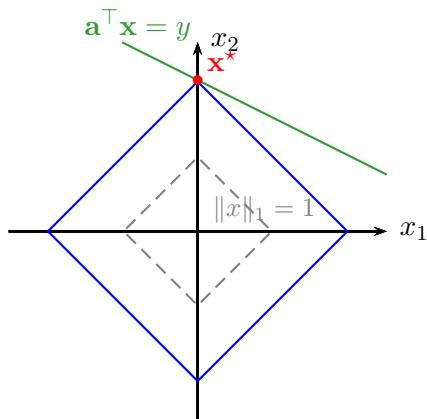
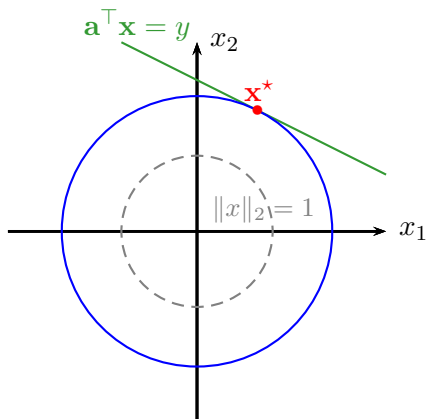
$$\hat{\boldsymbol{\theta}}_{\text{MAP}} = \text{Arg min}_{\boldsymbol{\theta}} \log p_{\boldsymbol{\theta}}(\boldsymbol{\theta}) \quad \text{s.t. } \mathbf{y} = \mathbf{A}\boldsymbol{\theta}$$

- If  $\boldsymbol{\theta} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_d)$ :  $\hat{\boldsymbol{\theta}}_{\text{MAP}} = \text{Arg min}_{\boldsymbol{\theta}} \|\boldsymbol{\theta}\|_2^2$  s.t.  $\mathbf{y} = \mathbf{A}\boldsymbol{\theta}$   
(least  $\ell_2$ -norm solution to linear equation)
- If  $\boldsymbol{\theta} \sim$  double-exponential:  $\hat{\boldsymbol{\theta}}_{\text{MAP}} = \text{Arg min}_{\boldsymbol{\theta}} \|\boldsymbol{\theta}\|_1$  s.t.  $\mathbf{y} = \mathbf{A}\boldsymbol{\theta}$   
(least  $\ell_1$ -norm solution to linear equation: basis pursuit)

Comparison  $\ell_2$  vs  $\ell_1$  norm:  $\ell_1$  promotes sparse solutions

# Comparison between $\ell_2$ and $\ell_1$ norms

$$\begin{aligned} \min. \quad & \|\mathbf{x}\| \\ \text{s.t.} \quad & \mathbf{a}^\top \mathbf{x} = y \end{aligned}$$



## Part III

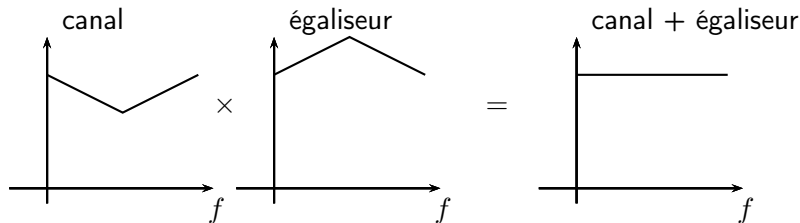
### Wiener filter

# Aspect fréquentiel du problème d'égalisation



- Egaliseur audio:
- (pré-)amplifier des fréquences absorbées/compenser imperfections de la chaîne d'écoute

- En communications numériques:

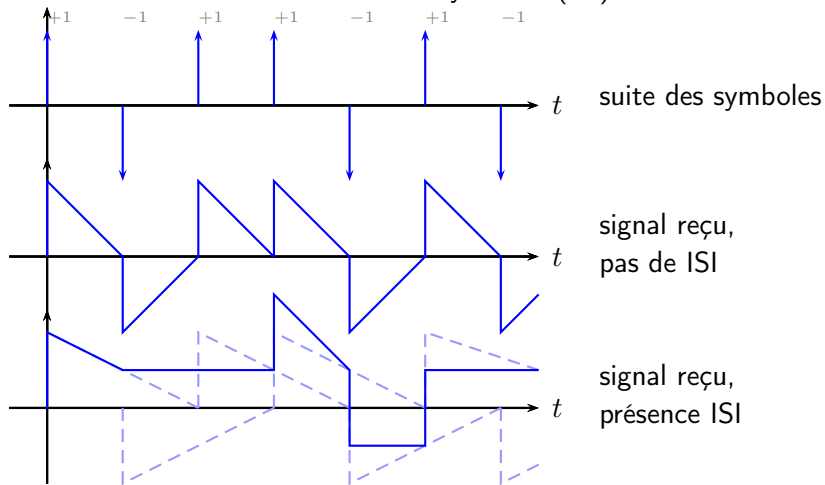


- égalisation  $\leftrightarrow$  compenser les effets du canal de propagation



## Aspect temporel du problème d'égalisation

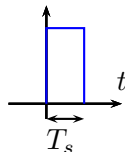
- Phénomène d'interférence entre symboles (ISI)



- Diagramme de l'oeil (simulation)

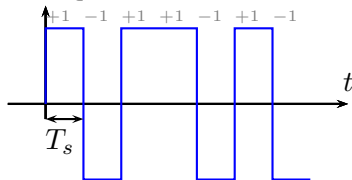
## Signal analogique en bande de base

- symboles à transmettre:  $(a_k)_{k \in \mathbb{Z}}$  à la période symbole  $T_s$
- impulsion/filtre de mise en forme:  $\Pi(t)$



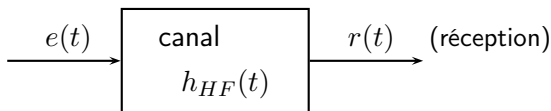
- signal en bande de base

$$d(t) = \sum_{k \in \mathbb{Z}} a_k \Pi(t - kT_s) = \left( \sum_k a_k \delta(t - kT_s) \right) \star \Pi(t)$$



- Rq: souvent,  $a_k \in \mathbb{C}$  (transmission de plusieurs bits sur un symbole) et  $d(t) \in \mathbb{C}$

# Modulation, enveloppe complexe

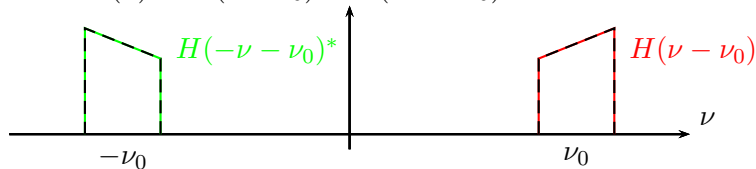


- $d(t) = p(t) + iq(t)$ : enveloppe complexe du signal émis
- signal émis: bande étroite autour de  $\nu_0$  (modulation)

$$e(t) = \Re\{d(t)e^{i2\pi\nu_0 t}\} = p(t) \cos(2\pi\nu_0 t) - q(t) \sin(2\pi\nu_0 t)$$

- filtre canal  $h_{HF}(t)$  réel et opère sur bande de  $e(t)$

$$\rightsquigarrow H_{HF}(\nu) = H(\nu - \nu_0) + H(-\nu - \nu_0)^*$$



## Enveloppe complexe, filtre passe-bas équivalent

En émission:

$$e(t) = \Re\{d(t)e^{i2\pi\nu_0 t}\} = \frac{1}{2} \left( d(t)e^{i2\pi\nu_0 t} + d(t)^* e^{-i2\pi\nu_0 t} \right)$$

$$E(\nu) = \frac{1}{2} (D(\nu - \nu_0) + D(-\nu - \nu_0)^*)$$

En réception:

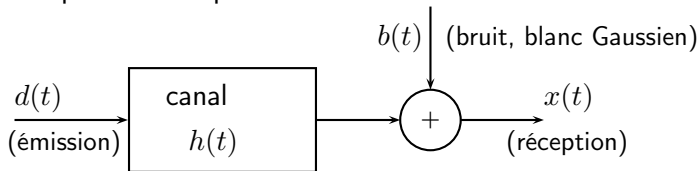
$$\begin{aligned} R(\nu) &= H_{HF}(\nu)E(\nu) \\ &= \frac{1}{2} \left( H(\nu - \nu_0) + H(-\nu - \nu_0)^* \right) \left( D(\nu - \nu_0) + D(-\nu - \nu_0)^* \right) \\ &= \frac{1}{2} \left( \underbrace{H(\nu - \nu_0)D(\nu - \nu_0)}_{X(\nu - \nu_0)} + \underbrace{H(-\nu - \nu_0)^* D(-\nu - \nu_0)^*}_{X(-\nu - \nu_0)^*} \right) \end{aligned}$$

$$r(t) = \Re\{x(t)e^{i2\pi\nu_0 t}\} \quad \text{où: } x(t) \xrightarrow{\text{TF}} X(\nu) = H(\nu)D(\nu)$$

$\Rightarrow$  enveloppes complexes  $d(t)$ ,  $x(t)$  et filtre passe-bas équivalent  $H(\nu)$  au lieu de  $e(t)$ ,  $r(t)$  et  $H_{HF}(\nu)$ .

## Modèle équivalent en bande de base (1/2)

- Signaux émis  $e(t)$  et reçu  $r(t)$  représentés par leurs enveloppes complexes  $d(t)$  et  $x(t)$ .
- Canal «haute-fréquence»  $H_{HF}$  représenté par son passe-bas équivalent  $H(\nu)$
- Modèle avec prise en compte d'un bruit additif:

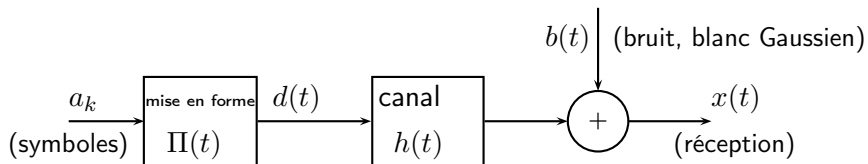


$$x(t) = h(t) \star d(t) + b(t)$$

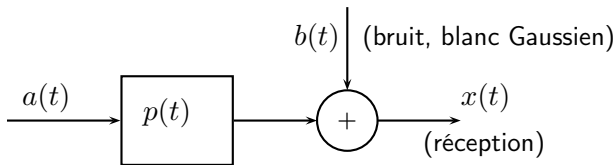
$$X(\nu) = H(\nu)X(\nu) + B(\nu)$$

avec mise en forme:  $d(t) = \sum_k a_k \delta(t - kT_s) \star \Pi(t)$

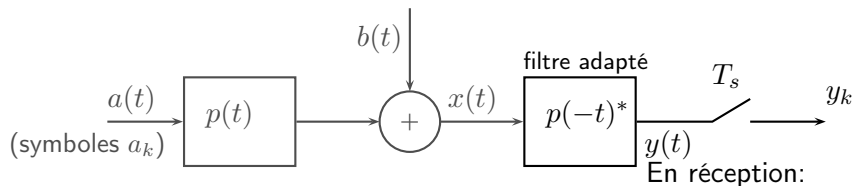
## Modèle équivalent en bande de base (2/2)



- Avec  $p(t) = h(t) \star \Pi(t)$  et  $a(t) = \sum_k a_k \delta(t - kT_s)$ :



# Filterage adapté, échantillonnage



$$x(t) = \sum_{k \in \mathbb{Z}} a_k p(t - kT_s) + b(t)$$

- Réponse impulsionnelle filtre adapté:  $p(-t)^* y(t) = x(t) \star p(-t)^*$
- Échantillonnage au rythme symbole  $T_s$

$$y_n = y(nT_s) = \int x(u) p(u - nT_s)^* du$$

⇒ les  $(y_n)_{n \in \mathbb{Z}}$  constituent un résumé exhaustif de  $(x(t))_{t \in \mathbb{R}}$ .

## Modèle temps discret

- Mise en forme, canal + bruit:

$$x(t) = \sum_{k \in \mathbb{Z}} a_k p(t - kT_s) + b(t) \quad \text{où: } p(t) = h(t) \star \Pi(t)$$

- Après filtre adapté

$$y(t) = x(t) \star p(-t)^*$$

$$y(t) = \sum_{k \in \mathbb{Z}} a_k g(t - kT_s) + \underbrace{w(t)}_{\text{bruit coloré}}$$

$$\text{avec: } g(t) = p(t) \star p(-t)^* \quad w(t) = p(-t)^* \star b(t)$$

- Echantillonnage  $y_n = y(nT_s)$  et en posant  $g_n = g(nT_s)$ ,  
 $w_n = w(nT_s)$ :

$$y_n = \sum_k a_k g_{n-k} + w_n$$



## Critère de Nyquist

- Avec  $g_n = g(nT_s)$ :

$$y_n = \sum_{k \in \mathbb{Z}} a_k g_{n-k} + \underbrace{w_n}_{\text{bruit coloré}}$$

- Pour éviter interférence entre symboles (ISI), il faut et il suffit que (en notant  $g(t) \xrightarrow{\text{TF}} G(\nu)$ ):

$$\begin{cases} g_0 = 1 \\ g_k = 0, k \neq 0 \end{cases} \Leftrightarrow \underbrace{\sum_k g_k e^{-i2\pi\nu T_s}}_{\frac{1}{T_s} \sum_k G(\nu + \frac{k}{T_s})} = 1 \quad \forall \nu \in \left[ \frac{-1}{2T_s}, \frac{1}{2T_s} \right]$$

$$\Rightarrow \text{critère de Nyquist: } \frac{1}{T_s} \sum_k G(\nu + \frac{k}{T_s}) = 1$$

## Propriété et factorisation spectrale

- $g(t) = p(t) \star p(-t)^*$  (fonction autocorrélation déterministe)
- si  $p(t)$  s'étale sur  $L$  symboles,  $g(t)$  sur  $2L$  symboles
- $g_n = g_{-n}^*$  et  $G(z) = \sum_n g_n z^{-n}$  vérifie  $G(z) = \left(G(1/z^*)\right)^*$
- si  $\xi_0$  est racine de  $G(z)$ , alors  $1/\xi_0^*$  aussi
- ⇒ il existe une factorisation spectrale

$$G(z) = F(z)F(1/z^*)^* \quad \text{avec } F(z) = \alpha \prod_{n=1}^L (1 - \xi_n z^{-1})$$

(correspond à la factorisation d'une densité spectrale).

- Rq: en prenant dans  $F(z)$  les  $L$  zéros à l'intérieur du cercle unité, on obtient  $F(z)$  à minimum de phase.

## Blanchiment

- On a le modèle discret:

$$y_n = \sum_{k \in \mathbb{Z}} a_k g_{n-k} + \underbrace{w_n}_{\text{bruit coloré}}$$

- si  $g_k \xrightarrow{\text{Tz}} G(z) = F(z)F(1/z^*)^*$ , alors le filtre  $B(z) = \frac{1}{F(1/z^*)^*}$  permet de blanchir le bruit.

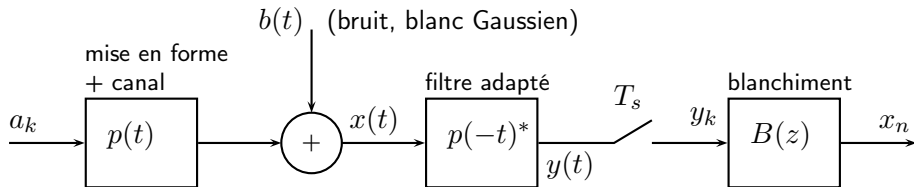
Preuve:

$$\begin{aligned} \mathbb{E}\{w_k w_l^*\} &= \mathbb{E}\left\{ \iint b(t)p(t - kT_s)^* b(\tau)p(\tau - lT_s) dt d\tau \right\} \\ &= 2N_0 \int p(t - kT_s)^* p(\tau - lT_s) dt = 2N_0 \int p(t)p(t - (k-l)T_s)^* dt \\ &= 2N_0 [p(t) \star p(-t)^*]_{(k-l)T_s} = 2N_0 g((k-l)T_s) = 2N_0 g_{k-l} \end{aligned}$$

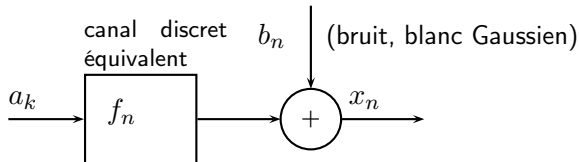
- Modèle après blanchiment:

$$x_n = \sum_k a_k f_{n-k} + b_n \quad \text{avec } b_n: \text{ bruit blanc}$$

## Résumé de la chaîne de traitement



Modèle équivalent:



$$x_n = \sum_k a_k f_{n-k} + \underbrace{b_n}_{\text{bruit blanc}}$$

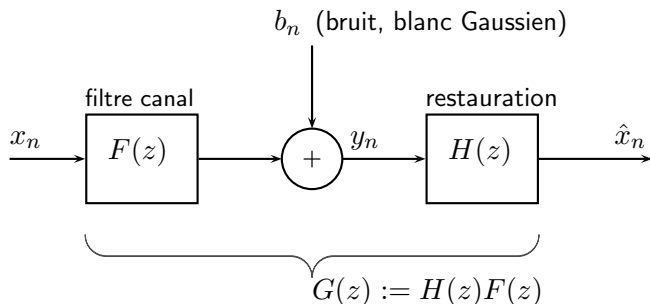
## Insuffisances du modèle

- canal inconnu: filtrage adapté difficile, blanchiment délicat
  - filtre non causaux et RII  $\rightarrow$  tronquer RI et introduire retard
  - choix des instants d'échantillonnage  $\rightarrow$  problème de synchronisation
- $\rightarrow$
- ▶ égalisation fractionnée (FSE: fractionally spaced equalizer)
  - ▶ voir cours plus avancés ...

# Critères d'égalisation et stratégies

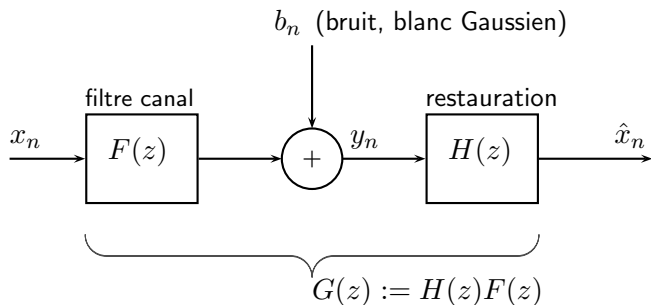
- Diagramme de l'oeil : qualitatif, empirique
- Probabilité d'erreur sur un symbole (BER: bit error rate)
  - ▶ critère le plus pertinent. Souvent, tracé de BER fonction du SNR
  - ▶ difficulté conception
  - ▶ égaliseur Optimum MLSE (maximum likelihood sequence estimation)
- Erreur quadratique moyenne (MSE: mean square error)
  - ▶  $MSE = \mathbb{E}\{|\hat{a}_n - a_n|^2\}$
  - ▶ pas de lien simple entre MSE et BER (mais en général bon BER si MSE faible)
  - ▶ égaliseur MMSE (minimum mean square error)
- Taux d'interférence entre symboles (ISI: inter-symbol interference)
  - ▶  $ISI = \frac{\sum_{k \neq 0} |g_k|^2}{|g_0|^2}$
  - ▶ le plus simple, utilisé quand autres critères non disponibles
  - ▶ égaliseur ZF (zero-forcing)

## Modèle



- Signal non observé à reconstruire:  $(x_n)_{n \in \mathbb{Z}}$
- Observations:  $(y_n)_{n \in \mathbb{Z}}$
- Filtre dégradation:  $F(z)$
- Filtre restauration:  $H(z)$
- Chaîne globale:  $G(z) = H(z)F(z)$
- Signal estimé après restauration:  $(\hat{x}_n)_{n \in \mathbb{Z}}$

## Modèle



- Signal non observé à reconstruire:  $(x_n)_{n \in \mathbb{Z}}$
- Observations:  $y_n = \sum_{k \in \mathbb{Z}} f_k x_{n-k} + b_n = f_n \star x_n + b_n$
- Signal estimé:  $\hat{x}_n = \sum_{k \in \mathbb{Z}} h_k y_{n-k} = h_n \star y_n$
- Chaîne globale:  $g_n = h_n \star f_n$



# Expression matricielle du modèle

## Canal RIF

- Canal supposé RIF  $(f_0, \dots, f_{K-1})$

- Signal reçu:  $y_n = \sum_{k=0}^{K-1} f_k x_{n-k} + b_n$

- Forme matricielle (avec ici  $\mathbf{A} \in \mathbb{C}^{(N+1) \times (N+K)}$ ):

$$\mathbf{A} = \begin{bmatrix} f_{K-1} & \dots & & f_0 \\ & f_{K-1} & \dots & f_0 \\ & & \ddots & \\ & & & f_{K-1} & \dots & f_0 \end{bmatrix} \quad \mathbf{x} = \begin{bmatrix} x_{-K+1} \\ \vdots \\ x_N \end{bmatrix}$$

$$\mathbf{y} = \begin{bmatrix} y_0 \\ \vdots \\ y_N \end{bmatrix}$$

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$$

# Filtre inverse

(égalité ZF: Zero Forcing)

- Bruit négligé:

$$\hat{x}_n = h_n \star y_n \quad y_n = f_n \star x_n \text{ et donc puisque } g_n = h_n \star f_n:$$

$$\hat{x}_n = g_n \star x_n = \sum_{k \in \mathbb{Z}} g_k x_{n-k} = g_0 x_n + \underbrace{\sum_{k \neq 0} g_k x_{n-k}}_{\text{interf. entre échantillons}}$$

- Objectif: annuler l'interférence ISI =  $\frac{\sum_{n \neq 0} |g_n|^2}{|g_0|^2}$
- Solution:  $g_0 = 1$  et  $g_k = 0, k \neq 0$  et donc  $G(z) = 1$

⇒ le filtre inverse:

$$H_{\text{inv}}(z) = \frac{1}{F(z)}$$

annule les interférences entre échantillons successifs de  $x_n$ .

## Filtre inverse en présence de bruit

- En présence de bruit:

$$\hat{x}_n = h_n \star y_n = h_n \star (f_n \star x_n + b_n)$$

$$\hat{x}_n = g_n \star x_n + h_n \star b_n = \underbrace{g_0 x_n + \sum_{k \neq 0} g_k x_{n-k}}_{=x_n \text{ si filtre inverse}} + \underbrace{h_n \star b_n}_{\text{bruit par le filtre inverse}}$$

- Filtre canal sélectif en fréquence:  $F(e^{i\omega}) \xrightarrow{\omega \rightarrow \omega_0} 0$ . Alors:

$$H_{\text{inv}}(e^{i\omega}) = \frac{1}{F(e^{i\omega})} \xrightarrow{\omega \rightarrow \omega_0} \infty$$

⇒ Amplification possible du bruit, éventuellement incontrôlée!

# Filtre de Wiener

(égalité MMSE: Minimum Mean Square Error)

- Estimation  $n^{\text{ème}}$  symbole:  $\hat{x}_n = h_n \star y_n = \sum_k h_k y_{n-k}$
- Erreur d'estimation:  $\epsilon_n = x_n - \hat{x}_n$
- Objectif: minimiser erreur quadratique moyenne

$$\min. \text{MSE} := \mathbb{E}\{|\epsilon_n|^2\}$$

→ projection orthogonale sur  $\overline{\text{span}}\{y_{n-p}\}_{p \in \mathbb{Z}}$ .

- Condition d'orthogonalité  $\forall p, \epsilon_n \perp y_{n-p}$  donne équations vérifiées par la réponse impulsionnelle  $(h_k)_{k \in \mathbb{Z}}$ :

$$\forall p \in \mathbb{Z} \quad \gamma_{xy}(p) = \sum_{k \in \mathbb{Z}} h_k \gamma_y(p - k)$$

## Expressions du filtre de Wiener (1/2)

- Réponse en fréquence:

$$H_{\text{Wiener}}(e^{i\omega}) = \frac{\Gamma_{xy}(e^{i\omega})}{\Gamma_y(e^{i\omega})}$$

Avec:

$\Gamma_y(e^{i\omega})$ : densité spectrale de puissance de  $y_n$

$\Gamma_{xy}(e^{i\omega})$ : densité inter-spectrale de puissance de  $x_n$  et  $y_n$

- Transformée en  $z$ :

$$H_{\text{Wiener}}(z) = \frac{\Gamma_{xy}(z)}{\Gamma_y(z)}$$

## Expressions du filtre de Wiener (2/2)

- Hypothèses:

- ▶ signal  $(x_n)_{n \in \mathbb{Z}}$  i.i.d. centré,  $\mathbb{E}\{|x_n|^2\} = \sigma_x^2$
- ▶ bruit additif blanc Gaussien  $(b_n)_{n \in \mathbb{Z}}$  indépendant, puissance  $\sigma_b^2$  (centré)
- ▶  $y_n = \sum_k f_k x_{n-k} + b_n$  et  $f_k \xrightarrow{\text{Tz}} F(z)$ .

- Réponse en fréquence:

$$H_{\text{Wiener}}(e^{i\omega}) = \frac{F(e^{i\omega})^*}{|F(e^{i\omega})|^2 + \sigma_b^2/\sigma_x^2}$$

- Transformée en  $z$ :

$$H_{\text{Wiener}}(z) = \frac{F(1/z^*)^*}{F(z)F(1/z^*)^* + \sigma_b^2/\sigma_x^2}$$

# Comportement du filtre de Wiener

$$H_{\text{Wiener}}(e^{i\omega}) = \frac{F(e^{i\omega})^*}{|F(e^{i\omega})|^2 + \sigma_b^2}$$

- Bruit faible  $\frac{\sigma_b}{\sigma_x} \rightarrow 0$ :  $H_{\text{Wiener}}(e^{i\omega}) = H_{\text{inv}}(e^{i\omega}) =$  filtre inverse.
- Bruit fort  $\frac{\sigma_b}{\sigma_x} \rightarrow \infty$ :  $H_{\text{Wiener}}(e^{i\omega}) = 0$  (signal noyé sous le bruit, reconstruction impossible)
- Comportement intermédiaire entre les cas ci-dessus en fonction rapport signal/bruit  $\frac{\sigma_b^2}{\sigma_x^2}$  et amplification/atténuation du filtre canal  $|F(e^{i\omega})|^2$  à chaque fréquence  $\omega$ .

# Illustration

- Les trois canaux test A, B, C (Proakis).
- Simulation du filtre inverse (égaliseur ZF).
- Simulation du filtre de Wiener (égaliseur MMSE).



## Filtre de Wiener RIF à horizon fini (1/2)

- Réponse impulsionnelle finie:  $h_k = 0$  pour  $k \notin \{0, \dots, L-1\}$

- Filtrage:  $\hat{x}_n = \sum_{k=0}^{L-1} h_k y_{n-k}$

- min.  $\mathbb{E}\{|x_n - \hat{x}_n|^2\}$  donne:

$$\forall p \in \{0, \dots, L-1\} \quad \mathbb{E}\{(x_n - \hat{x}_n)y_{n-p}^*\} = 0$$

càd avec  $\gamma_{xy}(p) := \mathbb{E}\{x_n y_{n-p}^*\}$  et  $\gamma_y(p) := \mathbb{E}\{y_n y_{n-p}^*\}$ :

$$\forall p \in \{0, \dots, L-1\} \quad \gamma_{xy}(p) = \sum_{k=0}^{L-1} h_k \gamma_y(p-k)$$

## Filtre de Wiener RIF à horizon fini (2/2)

- Réponse impulsionnelle  $(h_0, \dots, h_{L-1})$  donnée par:

$$\forall p \in \{0, \dots, L-1\} \quad \gamma_{xy}(p) = \sum_{k=0}^{L-1} h_k \gamma_y(p-k)$$

- Forme matricielle:

$$\underbrace{\begin{bmatrix} \gamma_{xy}(0) \\ \vdots \\ \vdots \\ \vdots \\ \gamma_{xy}(L-1) \end{bmatrix}}_{\mathbf{R}_{xy}} = \underbrace{\begin{bmatrix} \gamma_y(0) & \gamma_y(-1) & \dots & \dots & \gamma_y(-L+1) \\ \gamma_y(1) & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \gamma_y(-1) \\ \gamma_y(L-1) & \dots & \dots & \gamma_y(1) & \gamma_y(0) \end{bmatrix}}_{\mathbf{R}_{yy}} \underbrace{\begin{bmatrix} h_0 \\ \vdots \\ \vdots \\ \vdots \\ h_{L-1} \end{bmatrix}}_{\mathbf{h}}$$

- $\mathbf{h}_{\text{Wiener}} = \mathbf{R}_{yy}^{-1} \mathbf{R}_{xy}$

# Expression matricielle du modèle

## Canal RIF

- Canal supposé RIF ( $f_0, \dots, f_{K-1}$ )
- Signal reçu:  $y_n = \sum_{k=0}^{K-1} f_k x_{n-k} + b_n$
- Forme matricielle (avec ici  $\mathbf{A} \in \mathbb{C}^{(N+1) \times (N+K)}$ ):

$$\mathbf{A} = \begin{bmatrix} f_{K-1} & \dots & & f_0 & & & \\ & f_{K-1} & \dots & f_0 & & & \\ & & \ddots & & \ddots & & \\ & & & f_{K-1} & \dots & f_0 & \end{bmatrix} \quad \mathbf{x} = \begin{bmatrix} x_{-K+1} \\ \vdots \\ x_N \end{bmatrix}$$

$$\mathbf{y} = \begin{bmatrix} y_0 \\ \vdots \\ y_N \end{bmatrix} \quad \mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$$

## Expression matricielle filtre de Wiener

- Forme matricielle:  $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$ 
  - ▶  $\mathbf{A}$ : matrice donnée (ici: Toeplitz, taille  $\mathbb{C}^{n \times p}$ )
  - ▶  $\mathbf{x}$ : suite d'échantillons i.i.d.,  $\mathbb{E}\{x_i\} = 0$ ,  $\mathbb{E}\{|x_i|^2\} = \sigma_x^2$
  - ▶  $\mathbf{b}$ : bruit i.i.d. indépendant de  $\mathbf{a}$ ,  $\mathbb{E}\{b_i\} = 0$ ,  $\mathbb{E}\{|b_i|^2\} = \sigma_b^2$
- estimation:  $\hat{\mathbf{x}} = \mathbf{H}_{\text{Wiener}}\mathbf{y}$  avec:

$$\mathbf{H}_{\text{Wiener}} = \text{Arg min}_{\mathbf{H}} \mathbb{E}\{\|\mathbf{H}\mathbf{y} - \mathbf{x}\|^2\}$$

- Après développement MSE, en posant  $\lambda = \frac{\sigma_b^2}{\sigma_x^2}$ :

$$\mathbf{H}_{\text{Wiener}} = \text{Arg min}_{\mathbf{H}} \text{Tr} [(\mathbf{H}\mathbf{A} - \mathbf{I}\mathbf{d})(\mathbf{H}\mathbf{A} - \mathbf{I}\mathbf{d})^{\text{H}} + \lambda\mathbf{H}\mathbf{H}^{\text{H}}]$$

- Annulation gradient ou projection orthogonale sur  $\text{span } \mathbf{y}$  donnent:

$$\mathbf{H}_{\text{Wiener}} = \mathbf{A}^{\text{H}}(\mathbf{A}\mathbf{A}^{\text{H}} + \lambda\mathbf{I}\mathbf{d})^{-1}$$

# Filtre de Wiener

## Calcul (1/2)

→  $\forall p, \epsilon_n \perp y_{n-p}$  d'où:

$$\begin{aligned} 0 &= \mathbb{E}\{\epsilon_n y_{n-p}^*\} = \mathbb{E}\left\{\left(x_n - \sum_k h_k y_{n-k}\right) y_{n-p}^*\right\} \\ &= \underbrace{\mathbb{E}\{x_n y_{n-p}^*\}}_{\gamma_{xy}(p)} - \sum_k h_k \underbrace{\mathbb{E}\{y_{n-k} y_{n-p}^*\}}_{\gamma_y(p-k)} \end{aligned}$$

$$\forall p \in \mathbb{Z} \quad \gamma_{xy}(p) = \sum_{k \in \mathbb{Z}} h_k \gamma_y(p-k) \text{ et après T.z:}$$

$$\Gamma_{xy}(z) = H(z) \Gamma_y(z) \text{ d'où: } H_{\text{Wiener}}(z) = \frac{\Gamma_{xy}(z)}{\Gamma_y(z)}$$

## Filtre de Wiener

## Calcul (2/2)

$$\gamma_{xy}(p) = \mathbb{E}\{x_n y_{n-p}^*\} = \mathbb{E}\left\{x_n \left(\sum_k x_k f_{n-p-k} + b_{n-p}\right)^*\right\} = \sigma_x^2 f_{-p}^*$$

$$\text{d'où } \Gamma_{xy}(z) = \sigma_x^2 F(1/z^*)^*$$

$$\gamma_y(p) = \mathbb{E}\{y_n y_{n-p}^*\} = \dots = \sigma_x^2 f_n \star f_{-n}^* + \sigma_b^2 \delta_p$$

$$\text{d'où } \Gamma_y(z) = \sigma_x^2 F(z)F(1/z^*)^* + \sigma_b^2$$

$$H_{\text{Wiener}}(z) = \frac{F(1/z^*)^*}{F(z)F(1/z^*)^* + \sigma_b^2/\sigma_x^2}$$

$$H_{\text{Wiener}}(e^{i\omega}) = \frac{F(e^{i\omega})^*}{|F(e^{i\omega})|^2 + \sigma_b^2/\sigma_x^2}$$

## Détails calcul (1/3: MSE)

Modèle d'observation  $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$

( $\mathbf{A} \in \mathbb{C}^{n \times p}$  donnée, ici Toeplitz)

- $\mathbf{x}$ : suite d'échantillons i.i.d.,  $\mathbb{E}\{x_i\} = 0$ ,  $\mathbb{E}\{|x_i|^2\} = \sigma_x^2$
- $\mathbf{b}$ : bruit i.i.d. indépendant de  $\mathbf{x}$ ,  $\mathbb{E}\{b_i\} = 0$ ,  $\mathbb{E}\{|b_i|^2\} = \sigma_b^2$

$$\begin{aligned} \mathbb{E}\{\|\mathbf{H}\mathbf{y} - \mathbf{x}\|^2\} &= \mathbb{E}\{\|(\mathbf{H}\mathbf{A} - \mathbf{Id})\mathbf{x} + \mathbf{H}\mathbf{b}\|^2\} \\ &= \mathbb{E}\{\text{Tr} [(\mathbf{H}\mathbf{A} - \mathbf{Id})\mathbf{x} + \mathbf{H}\mathbf{b})(\mathbf{H}\mathbf{A} - \mathbf{Id})\mathbf{x} + \mathbf{H}\mathbf{b})^H]\} \\ &= p\sigma_x^2 \text{Tr} [(\mathbf{H}\mathbf{A} - \mathbf{Id})(\mathbf{H}\mathbf{A} - \mathbf{Id})^H] + p\sigma_b^2 \text{Tr} [\mathbf{H}\mathbf{H}^H] \end{aligned}$$

et donc avec  $\lambda = \frac{\sigma_b^2}{\sigma_x^2}$ :

$$\begin{aligned} \mathbf{H}_{\text{Wiener}} &= \text{Arg min}_{\mathbf{H}} \text{Tr} [(\mathbf{H}\mathbf{A} - \mathbf{Id})(\mathbf{H}\mathbf{A} - \mathbf{Id})^H + \lambda\mathbf{H}\mathbf{H}^H] \\ &= \text{Arg min}_{\mathbf{H}} \mathbb{E}\{\|\mathbf{H}\mathbf{y} - \mathbf{x}\|^2\} \end{aligned}$$

## Détails calcul (2/3: gradient)

Critère MSE:  $\mathcal{J}(\mathbf{H}) = \text{Tr} \left[ (\mathbf{H}\mathbf{A} - \mathbf{Id})(\mathbf{H}\mathbf{A} - \mathbf{Id})^H + \lambda\mathbf{H}\mathbf{H}^H \right].$

$$\begin{aligned} \mathcal{J}(\mathbf{H} + \mathbf{E}) &= \text{Tr} \left[ (\mathbf{H}\mathbf{A} - \mathbf{Id} + \mathbf{E}\mathbf{A})(\mathbf{H}\mathbf{A} - \mathbf{Id} + \mathbf{E}\mathbf{A})^H \right. \\ &\quad \left. + \lambda(\mathbf{H} + \mathbf{E})(\mathbf{H} + \mathbf{E})^H \right] \\ &= \mathcal{J}(\mathbf{H}) + \text{Tr} \left[ (\mathbf{H}\mathbf{A} - \mathbf{Id})\mathbf{A}^H\mathbf{E}^H + \mathbf{E}\mathbf{A}(\mathbf{H}\mathbf{A} - \mathbf{Id})^H \right. \\ &\quad \left. + \lambda(\mathbf{H}\mathbf{E}^H + \mathbf{E}\mathbf{H}^H) \right] + o(\mathbf{E}) \end{aligned}$$

d'où la dérivée:

$$\frac{\partial \mathcal{J}}{\partial \overline{\mathbf{H}}} = (\mathbf{H}\mathbf{A} - \mathbf{Id})\mathbf{A}^H + \lambda\mathbf{H}.$$

Annulation de la dérivée pour  $\mathbf{H}(\mathbf{A}\mathbf{A}^H + \lambda\mathbf{I}) = \mathbf{A}^H$  et donc:

$$\mathbf{H}_{\text{Wiener}} = \text{Arg min}_{\mathbf{H}} \mathcal{J}(\mathbf{H}) = \mathbf{A}^H(\mathbf{A}\mathbf{A}^H + \lambda\mathbf{Id})^{-1}$$



## Détails calcul (3/3: projection orthogonale)

- $\hat{\mathbf{x}} = \mathbf{H}_{\text{Wiener}}\mathbf{y}$  est tel que  $\hat{\mathbf{x}} - \mathbf{x} \perp \mathbf{y}$  càd:

$$\mathbb{E}\{(\hat{\mathbf{x}} - \mathbf{x})\mathbf{y}^H\} = \mathbb{E}\{(\mathbf{H}_{\text{Wiener}}\mathbf{y} - \mathbf{x})\mathbf{y}^H\} = \mathbf{0} \quad \text{d'où:}$$

$$\mathbf{H}_{\text{Wiener}}\mathbb{E}\{\mathbf{y}\mathbf{y}^H\} = \mathbb{E}\{\mathbf{x}\mathbf{y}\} \text{ donc:}$$

- En utilisant les hypothèses du modèle  $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$ :

$$\mathbb{E}\{\mathbf{y}\mathbf{y}^H\} = \mathbb{E}\{(\mathbf{A}\mathbf{x} + \mathbf{b})(\mathbf{A}\mathbf{x} + \mathbf{b})^H\} = \sigma_x^2\mathbf{A}\mathbf{A}^H + \sigma_b^2\mathbf{Id}$$

$$\mathbb{E}\{\mathbf{x}\mathbf{y}^H\} = \mathbb{E}\{\mathbf{x}(\mathbf{A}\mathbf{x} + \mathbf{b})^H\} = \sigma_x^2\mathbf{A}^H$$

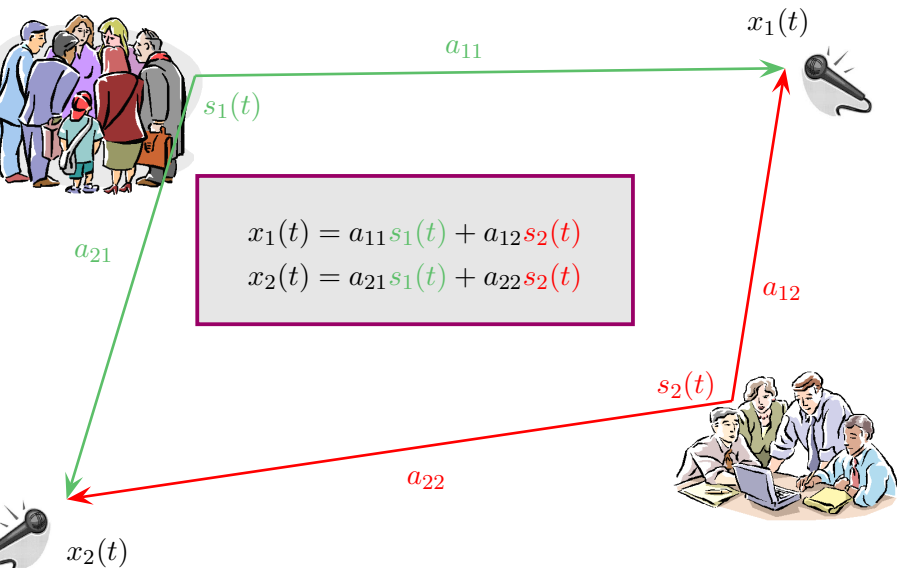
- Finalement avec  $\lambda = \frac{\sigma_b^2}{\sigma_x^2}$ :

$$\mathbf{H}_{\text{Wiener}} = \mathbb{E}\{\mathbf{x}\mathbf{y}\} \mathbb{E}\{\mathbf{y}\mathbf{y}^H\}^{-1} = \mathbf{A}^H(\mathbf{A}\mathbf{A}^H + \lambda\mathbf{Id})^{-1}$$

## Part IV

# Independent Component Analysis

## «Cocktail party» problem



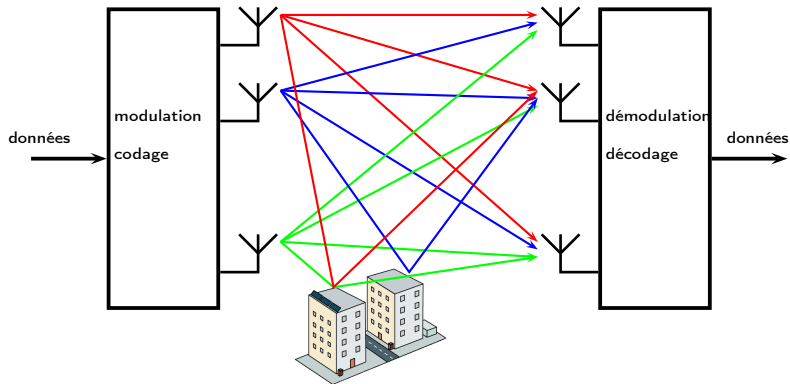
# Mélange linéaire instantané

- $\mathbf{s}(t) = \begin{pmatrix} s_1(t) \\ s_2(t) \end{pmatrix}$ : vecteur des **sources**
- $\mathbf{x}(t) = \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix}$ : vecteur des **observations**
- $\mathbf{A}$ : **matrice de mélange** de taille  $2 \times 2$
- lien sources - observations:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t)$$

- ↪ Plus généralement dans ce cours:  
nombre de sources = nombre d'observations =  $N \geq 2$

- Transmission hertzienne «classique»: 1 émetteur - 1 récepteur
- Cas MIMO: multi input - multi output



- $\Rightarrow$  augmenter la fiabilité et le débit pour les réseaux locaux sans fils, téléphonie mobile, ...

## Autres applications possibles

- Parole, audio, son
- **Image** : décomposition d'une image, watermarking
- **Biomédical** :
  - ▶ magnéto- et électroencéphalographie et/ou IRMf (reconstruction des sources neuronales)
  - ▶ électrocardiogramme du fœtus
- Télécommunications
- Cosmologie: extraction du CMB (Cosmic Microwave Background)
- Sismique, géophysique
- *etc.*

# Cadre aveugle

- Modèle instantané considéré:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) = \begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix} s_1(t) + \begin{pmatrix} a_{12} \\ a_{22} \end{pmatrix} s_2(t)$$

- But : reconstruire les sources  $s_1(t), s_2(t)$  à partir des observations  $x_1(t), x_2(t)$ , c'ad trouver  $\mathbf{B}$  tel que:

$$\mathbf{y}(t) = \mathbf{B}\mathbf{x}(t) \quad \text{et:} \quad \mathbf{y}(t) \leftrightarrow \mathbf{s}(t)$$

- Approches possibles:
  - ▶ Connaissances sur le mélange (modèle physique de propagation, ...) : problème de reconstruction inverse.
  - ▶ **Pas de connaissances sur  $\mathbf{A}$**  : cadre **aveugle**.  $\rightsquigarrow$  nécessite des hypothèses sur les sources.

## Séparation de signaux sonores

- **Sources** : 3 **signaux sonores** provenant d'**enregistrement CD** échantillonnage à 44100Hz, 221088 échantillons (durée approx. 5s).
- **Mélange** par la matrice (tirée aléatoirement):

$$\begin{pmatrix} -0.2507 & 0.1714 & 0.9637 \\ -0.9653 & -0.6833 & -0.0305 \\ 0.0726 & 0.7098 & 0.2652 \end{pmatrix}$$

- Algorithme utilisé: CoM2 (voir [5] et page web correspondante)
- ↪ Objectifs du cours: présenter quelques principes de base de la séparation aveugle



# Est-ce possible en aveugle?

- **Indéterminations** :

- ▶ **ordre des sources**  
→ reconstitution à une permutation près
- ▶ **facteur multiplicatif** pour chaque source

$$x_j(t) = \sum_{i=1}^N a_{ij} s_i(t) \quad x_j(t) = \sum_{i=1}^N \frac{a_{ij}}{\lambda_i} \lambda_i s_i(t)$$

→ normalisation :  $\langle s_1(t)^2 \rangle = \langle s_2(t)^2 \rangle = 1$

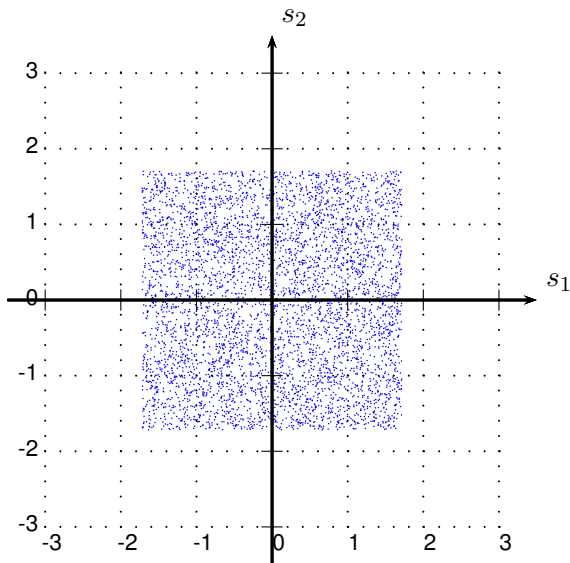
(  $\langle \cdot \rangle$  = valeur moyenne)

↔ dans  $\mathbb{R}$  (resp.  $\mathbb{C}$ ), indétermination de signe (resp. rotation complexe).

- Hypothèse forte, mais souvent plausible:  
**indépendance statistique des sources.**

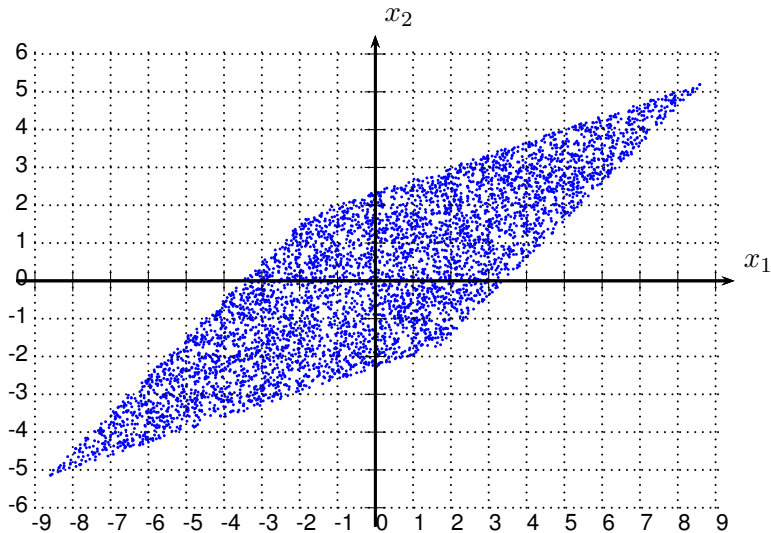
# Sources indépendantes

Deux sources  $s_1$  et  $s_2$  distribuées uniformément et indépendantes:



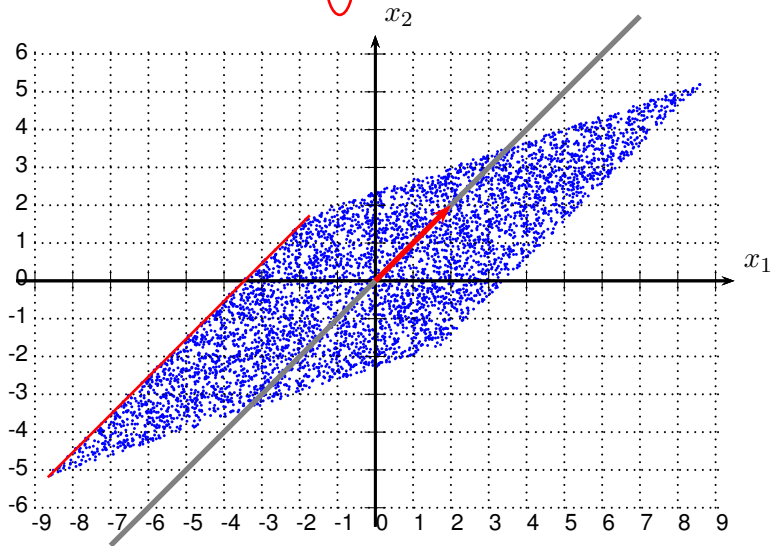
## Illustration du cas aveugle

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) \quad \mathbf{A} = \begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix}$$



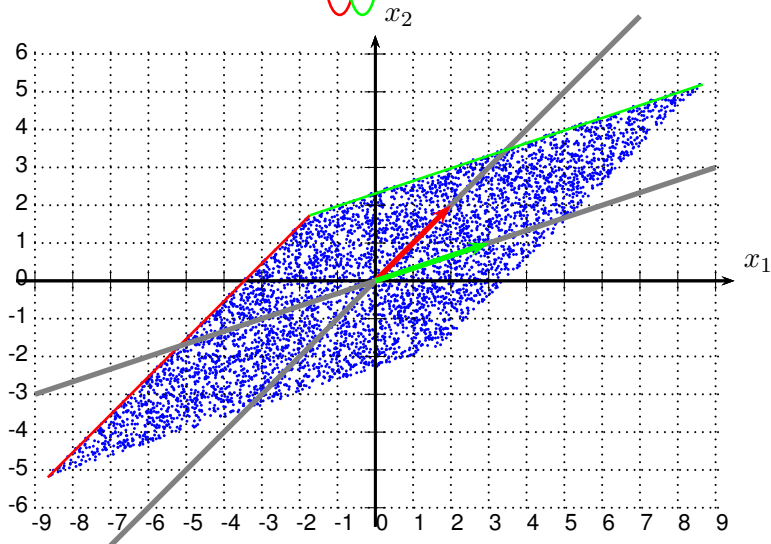
## Illustration du cas aveugle

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) \quad \mathbf{A} = \begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix}$$



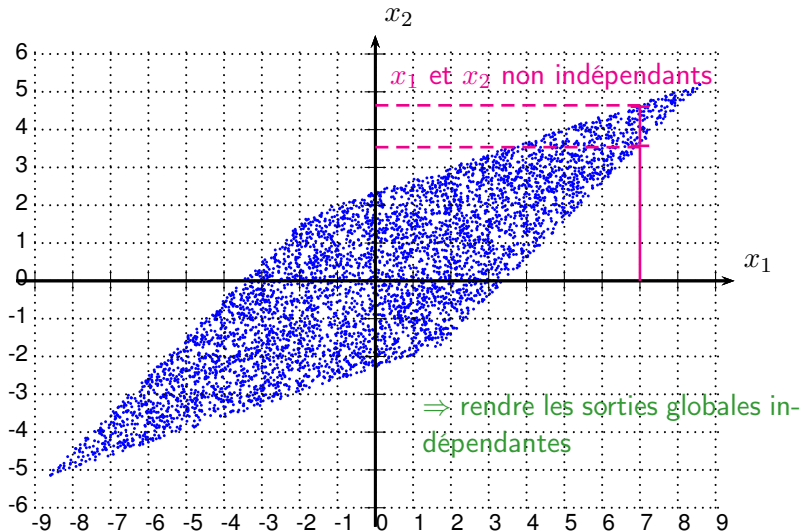
## Illustration du cas aveugle

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) \quad \mathbf{A} = \begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix}$$



# Utiliser l'indépendance

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) \quad \mathbf{A} = \begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix}$$



# Séparation aveugle de sources / ICA

- Détermination des sources  $s_i(t)$  à partir des  $x_j(t)$ :

$$x_j(t) = \sum_{i=1}^N a_{ij} s_i(t) \quad / \quad \mathbf{x}(t) = \mathbf{A}\mathbf{s}(t)$$

⇒ **Séparation aveugle de sources**

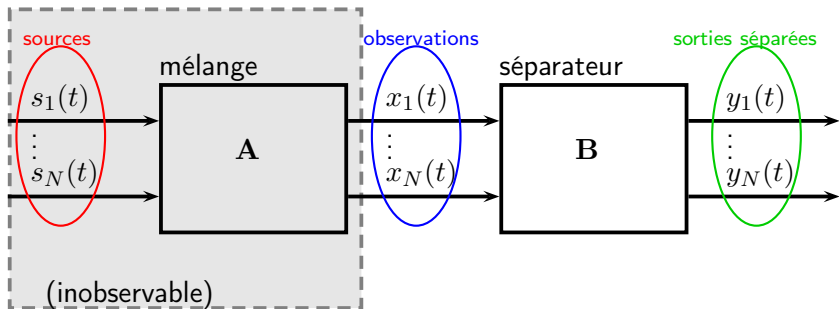
(Blind Source Separation, **BSS**)

- Détermination des composantes indépendantes:

$$\mathbf{x}(t) = \sum_{i=1}^N \underbrace{\begin{pmatrix} a_{1i} \\ \vdots \\ a_{Ni} \end{pmatrix}}_{\text{composante indépendante}} s_i(t) = \mathbf{A}\mathbf{s}(t)$$

⇒ **ICA: Independent Component Analysis**

# Notations



- Equations correspondantes:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) \quad \mathbf{y}(t) = \mathbf{B}\mathbf{x}(t)$$

- En posant  $\mathbf{G} := \mathbf{B}\mathbf{A}$ :

$$\mathbf{y}(t) = \mathbf{G}\mathbf{s}(t)$$



## Cadre considéré

- Sources temporellement **indépendantes et identiquement distribuées (i.i.d.)**

$$\mathbf{x} = \mathbf{A}\mathbf{s} \quad \mathbf{y} = \mathbf{B}\mathbf{x}$$

avec  $\mathbf{s}$ ,  $\mathbf{x}$ : vecteurs aléatoires taille  $N$ .

↪ Dans ce cadre, la **structure temporelle des signaux n'est pas prise en compte.**

- Sources centrées et de variance unité:

$$\forall i \in \{1, \dots, N\} \quad \mathbb{E}\{s_i\} = 0 \quad \mathbb{E}\{|s_i|^2\} = 1$$

# Indépendance de variables aléatoires (1/2)

- $s_1$  et  $s_2$  sont **indépendantes** lorsque:

$$f_{s_1, s_2}(u, v) = f_{s_1}(u)f_{s_2}(v)$$

$$\text{où: } f_{s_1}(u) = \int f_{s_1, s_2}(u, v) dv \text{ et: } f_{s_2}(v) = \int f_{s_1, s_2}(u, v) du$$

⇒ Pour toutes fonctions  $h_1$  et  $h_2$ :

$$\mathbb{E}\{h_1(s_1)h_2(s_2)\} = \mathbb{E}\{h_1(s_1)\}\mathbb{E}\{h_2(s_2)\}$$

## Indépendance de variables aléatoires (2/2)

- Les composantes de  $\mathbf{s} = (s_1, \dots, s_N)$  sont (**mutuellement indépendantes**) lorsque:

$$f_{\mathbf{s}}(s_1, \dots, s_N) = \prod_{k=1}^N f_{s_k}(s_k)$$

⇒ Pour toutes fonctions  $h_1, \dots, h_N$ :

$$\mathbb{E}\{h_1(s_1) \dots h_N(s_N)\} = \mathbb{E}\{h_1(s_1)\} \dots \mathbb{E}\{h_N(s_N)\}$$

- Attention à ne pas confondre avec l'**indépendance par paires**:
  - ▶ Si  $(s_1, \dots, s_N)$  sont indépendantes mutuellement, elles le sont deux par deux.
  - ▶ L'indépendance par paires n'entraîne pas (en général) l'indépendance mutuelle.

## Décorrélation de variables aléatoires

- Deux variables aléatoires  $s_1$  et  $s_2$  sont **décorrélées** lorsque :

$$\mathbb{E}\{s_1 s_2\} = \mathbb{E}\{s_1\}\mathbb{E}\{s_2\}$$

- **variables aléatoires indépendantes**  $\Rightarrow$  **décorrélées**

**MAIS** réciproque fausse.

$\rightsquigarrow$  La décorrélation est un (premier) pas vers l'indépendance

- Pour  $\mathbf{s} = (s_1, \dots, s_N)^\top$ , on définit sa matrice de covariance

$$\mathbf{R}_s := \mathbb{E}\{\mathbf{s}\mathbf{s}^\top\} \quad \text{càd: } (\mathbf{R}_s)_{ij} = \mathbb{E}\{s_i s_j\}$$

Le vecteur est décorrélé lorsque  $\mathbf{R}_s$  est diagonal.

Rq: vecteur décorrélé  $\Leftrightarrow$  toutes les paires sont décorrélées

# Blanchiment

Rappel: modèle  $\mathbf{x} = \mathbf{A}\mathbf{s}$  / objectif: trouver  $\mathbf{B}$  telle que  $\mathbf{B}\mathbf{A} = \mathbf{Id}$ .

**Blanchir** les données: trouver une transformation linéaire qui rende les nouvelles données **décorrélées**

- Sorties décorréées notées  $\mathbf{z} := \mathbf{W}\mathbf{x}$
- Contrainte de décorrélation :

$$\mathbb{E}\{\mathbf{z}\mathbf{z}^T\} = \mathbf{Id}_N$$

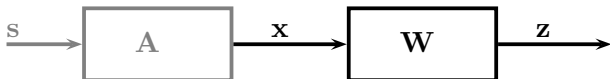
↪ par symétrie,  $N(N + 1)/2$  équations

→ **insuffisant** pour la détermination des  $N^2$  paramètres de  $\mathbf{A}$  (ou  $\mathbf{B}$ ).

- Rq: mise en œuvre du blanchiment aisée (SVD)

## Conséquence(s) du blanchiment

- $\mathbf{z} := \mathbf{W}\mathbf{x} = \mathbf{W}\mathbf{A}\mathbf{s}$  avec:  $\mathbb{E}\{\mathbf{z}\mathbf{z}^T\} = \mathbf{Id}_N \rightarrow$  insuffisant

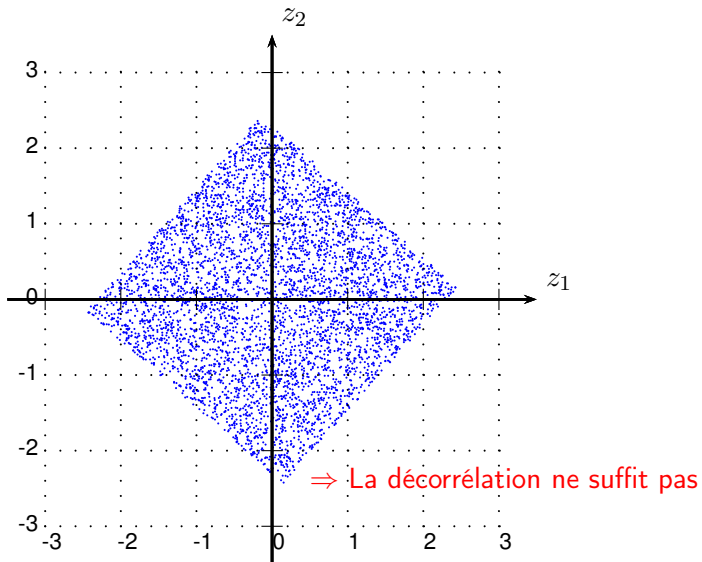


- En posant  $\mathbf{C} = \mathbf{W}\mathbf{A}$ , nouveau mélange  $\mathbf{z} = \mathbf{C}\mathbf{s}$  qui vérifie:

$$\mathbf{Id}_N = \mathbb{E}\{\mathbf{z}\mathbf{z}^T\} = \mathbf{C}\mathbb{E}\{\mathbf{s}\mathbf{s}^T\}\mathbf{C}^T = \mathbf{C}\mathbf{C}^T$$

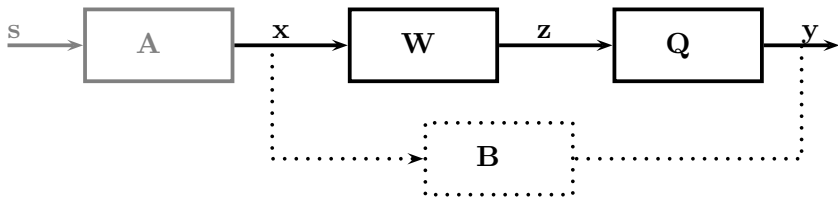
- ↪ matrice mélange équivalente  $\mathbf{C}$  orthogonale ( $\mathbf{C}\mathbf{C}^T = \mathbf{Id}$ )
- ↪ transformation isométrique

# Illustration du blanchiment



## Conséquence(s) du blanchiment

- En posant  $C = WA$ , nouveau mélange  $z = Cs$  avec  $C^T C = \text{Id}$ .



- Inverse du mélange blanchi:  $Q \leftrightarrow C^{-1}$   
 $Q$  cherché est  $C^{-1} = C^T$ , donc orthogonale
- ↪ paramétrisation de  $Q$  par des angles (Jacobi)
- Matrice de séparation a été décomposée en  $B = QW$



## Statistiques d'ordre supérieur

Rappel:  $s_1, \dots, s_N$  sont **indépendantes** lorsque pour toutes fonctions  $h_1, \dots, h_N$ :

$$\mathbb{E}\{h_1(s_1) \dots h_N(s_N)\} = \mathbb{E}\{h_1(s_1)\} \dots \mathbb{E}\{h_N(s_N)\}$$

- Décorrélation/blanchiment: **(insuffisant)**

$$\mathbb{E}\{z_i z_j\} = \mathbb{E}\{z_i\} \mathbb{E}\{z_j\} \text{ pour toutes les paires } i, j$$

→ Introduire des produits de  $k > 2$  termes:  $\mathbb{E}\{z_{i_1} \dots z_{i_k}\}$

Prendre des fonctions  $h_i$  polynomiales (monômes degré  $> 2$ )

↪ **Statistiques d'ordre supérieur (à deux)**

# Moments

- Variable aléatoire  $r$ 
  - Moment** d'ordre  $k$  de  $r$ :  $\mathbb{E}\{r^k\}$
- Vecteur aléatoire  $\mathbf{r} = (r_1, \dots, r_N)^\top$ :
  - Moment (croisé)** d'ordre  $k$ :  $\mathbb{E}\{r_{i_1} \dots r_{i_k}\}$
- Moment d'ordre supérieur (à deux):  $k > 2$
- manipulation peu pratique (pas de relation pour le moment d'une somme, en cas indépendance, ...)
- Rq: les moments apparaissent dans le développement de la première fonction caractéristique

$$\varphi(u) = \mathbb{E}\{e^{iur}\} \quad (\text{cas scalaire})$$

$$\varphi(\mathbf{u}) = \mathbb{E}\{e^{i\mathbf{u}^\top \mathbf{r}}\} \quad (\text{cas vectoriel})$$

## Cumulants

**Cumulants** d'ordre  $k$  : s'expriment polynomialement en fonction moments d'ordre  $\leq k$  (et réciproquement).

- Pour des variables centrées ( $\mathbb{E}\{r_i\} = 0$ ):

$$\text{Cum}\{r_1, r_2\} = \mathbb{E}\{r_1 r_2\}$$

$$\text{Cum}\{r_1, r_2, r_3\} = \mathbb{E}\{r_1 r_2 r_3\}$$

$$\begin{aligned} \text{Cum}\{r_1, r_2, r_3, r_4\} &= \mathbb{E}\{r_1 r_2 r_3 r_4\} - \mathbb{E}\{r_1 r_2\}\mathbb{E}\{r_3 r_4\} \\ &\quad - \mathbb{E}\{r_1 r_3\}\mathbb{E}\{r_2 r_4\} - \mathbb{E}\{r_1 r_4\}\mathbb{E}\{r_2 r_3\} \end{aligned}$$

- Cumulant d'ordre quatre (**kurtosis**) d'une variable  $r$  centrée:

$$\text{Cum}\{r, r, r, r\} = \mathbb{E}\{r^4\} - 3(\mathbb{E}\{r^2\})^2 =: \kappa(r)$$

- Rq: les cumulants apparaissent dans le développement de la deuxième fonction caractéristique

$$\psi(u) = \log \mathbb{E}\{e^{iur}\} \quad (\text{cas scalaire})$$

$$\psi(\mathbf{u}) = \log \mathbb{E}\{e^{i\mathbf{u}^\top \mathbf{r}}\} \quad (\text{cas vectoriel})$$

## Propriétés des cumulants (1/3)

(i) **symétrie**: pour toute permutation des indices:

$$\text{Cum} \{r_1, \dots, r_k\} = \text{Cum} \{r_{i_1}, \dots, r_{i_k}\}$$

(ii) **multilinéarité** (par rapport à chacun des arguments):

$$\begin{aligned}\text{Cum} \{a + b, r_2, \dots, r_k\} &= \text{Cum} \{a, r_2, \dots, r_k\} + \text{Cum} \{b, r_2, \dots, r_k\} \\ \text{Cum} \{\lambda a, r_2, \dots, r_k\} &= \lambda \text{Cum} \{a, r_2, \dots, r_k\}\end{aligned}$$

---

Vérification aisée dans le cas des expressions du transparent 99.

## Propriétés des cumulants (2/3: indépendance)

- (iii) s'il existe un sous-ensemble de variables  $(a_i)_{i=1,\dots,m}$  **indépendantes** des autres  $(b_j)_{j=1,\dots,n}$ :

$$\text{Cum} \{a_1, \dots, a_m, b_1, \dots, b_n\} = 0$$

- (iii') Si  $(a_i)_{i=1,\dots,k}$  **indépendantes** de  $(b_j)_{j=1,\dots,k}$ ,

$$\text{Cum} \{a_1 + b_1, \dots, a_k + b_k\} = \text{Cum} \{a_1, \dots, a_k\} + \text{Cum} \{b_1, \dots, b_k\}$$

**ATTENTION:** l'indépendance par paires ne suffit pas!

---

Vérification aisée dans le cas des expressions du transparent 99.

- (iii)  $\Leftrightarrow$  (iii') évident avec la multi-linéarité.

## Propriétés des cumulants (3/3: cas gaussien)

(iv) pour un vecteur  $(a_1, \dots, a_k)$  **gaussien**:

$$\text{Cum} \{a_1, \dots, a_k\} = 0 \quad (k > 2)$$

Pour  $r$  variable aléatoire gaussienne, le kurtosis est nul:

$$\kappa(r) = 0$$

- CONSEQUENCE(s) pour un vecteur gaussien:
  - ▶ statistiques d'ordre deux définissent sa loi,
  - ▶ décorrélation (par paires) équivaut à indépendance (mutuelle),
  - ▶ le blanchiment exploite toute l'information.

# Sources non gaussiennes?

- Variable aléatoire gaussienne:
  - ▶ définie par ses deux premiers moments
  - ▶ décorrélation  $\Leftrightarrow$  indépendance

$\Rightarrow$  Toute l'information statistique prise en compte lors du blanchiment.

- **Hypothèses** fondamentales de l'analyse en composantes indépendantes

(ICA) :

H.1 Les sources sont mutuellement **indépendantes**

H.2 Les sources ont une distribution de probabilité **non gaussienne** (sauf éventuellement une seule d'entre elles).

# Identifiabilité

Justification théorique (théorème de Darmois-Skitovitch, 1953):

Soit  $\mathbf{s} = (s_1, \dots, s_N)$  un vecteur aléatoire.

Si

- composantes de  $\mathbf{s}$  indépendantes,
- au plus l'une d'entre elles est gaussienne,

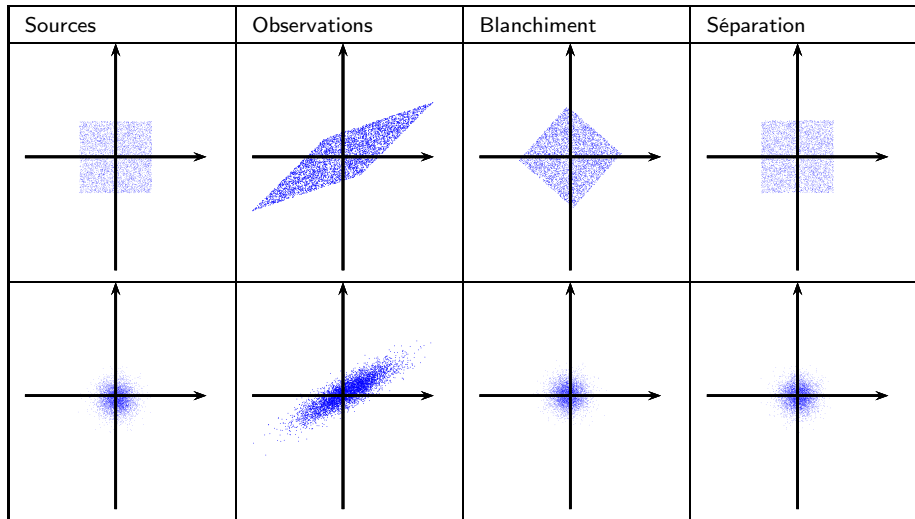
alors:

- composantes de  $\mathbf{y} = \mathbf{G}\mathbf{s}$  indépendantes  $\Leftrightarrow \mathbf{G} = \mathbf{PD}$ .  
( $\mathbf{P}$ : permutation,  $\mathbf{D}$ : matrice diagonale)

$\rightsquigarrow$  «**non-gaussianité**» et **indépendance** sont les deux hypothèses centrales pour identifier un mélange de sources (cas i.i.d.)



# Distributions uniforme et gaussienne



# ICA et PCA

## Analyse en composantes principales (PCA)

- Si données gaussiennes, ICA n'apporte rien de plus que PCA.
- PCA unique car condition de décorrélation couplée à contrainte  $\mathbf{A}\mathbf{A}^T = \mathbf{Id}$ .
- PCA recherche des composantes par ordre décroissant d'énergie.
- PCA plus modeste, mais toujours possible.
- Algorithme PCA universel et simple, pas ICA.

# Fonction de contraste

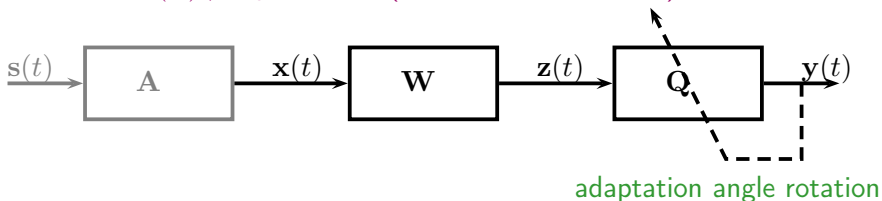
Fonction de **contraste**  $\phi$ : critère dont la maximisation permet de séparer les sources.

- Différents cadres de validité:
  - ▶ séparation de *toutes* les sources
  - ▶ extraction d'*une* source
  - ▶ avec/sans blanchiment
- Definition (adaptée au 1er cadre): Avec  $\mathbf{s}$ : vecteur de sources indépendantes,  $\phi$  est une fonction de contraste si et seulement si:
  - (i)  $\forall \mathbf{A} \quad \phi(\mathbf{A}\mathbf{s}) \leq \phi(\mathbf{s})$
  - (ii) Si  $\mathbf{A} = \mathbf{P}\mathbf{D}$ , alors  $\phi(\mathbf{A}\mathbf{s}) = \phi(\mathbf{s})$
  - (iii) Si  $\phi(\mathbf{A}\mathbf{s}) = \phi(\mathbf{s})$ , alors  $\mathbf{A}$  s'écrit  $\mathbf{A} = \mathbf{P}\mathbf{D}$ .( $\mathbf{P}$ : permutation,  $\mathbf{D}$ : matrice diagonale)

## Somme des kurtosis

- Cadre et hypothèses:

- ▶ blanchiment
- ▶ sources indépendantes
- ▶  $\kappa(s_i) \neq 0$  pour tout  $i$  (sauf éventuellement une).



- **Contraste:**  $\max. \mathcal{J}(\mathbf{y}) = \sum_{i=1}^N |\kappa(y_i)|$

Rq:

- ▶  $\kappa(s_i) \neq 0 \Rightarrow s_i$  non gaussienne
- ▶  $\kappa(\cdot) \leftrightarrow$  écart à la «gaussiannité»
- ▶  $\max. \sum_{i=1}^N (\kappa(y_i))^2$  est aussi un contraste.

## Justification contraste p.108 (1/3: exprimer les cumulants)

- Matrice globale sources/sorties séparées notée  $\mathbf{G} = [G_{ij}]_{ij}$ .

$$y_i = \sum_{j=1}^N G_{ij} s_j$$

- 

$$\kappa(y_i) = \text{Cum} \{y_i, y_i, y_i, y_i\}$$

par multilinéarité:

$$\kappa(y_i) = \sum_{j_1=1}^N \sum_{j_2=1}^N \sum_{j_3=1}^N \sum_{j_4=1}^N G_{ij_1} G_{ij_2} G_{ij_3} G_{ij_4} \text{Cum} \{s_{j_1}, s_{j_2}, s_{j_3}, s_{j_4}\}$$

par indépendance des sources:

$$\kappa(y_i) = \sum_{j=1}^N G_{ij}^4 \kappa(s_j)$$

## Justification contraste p.108 (2/3: majorer le critère)

- Blanchiment:  $\mathbf{y} = \mathbf{G}\mathbf{s}$  avec  $\mathbf{G}$  orthogonale ( $\mathbf{G}^\top \mathbf{G} = \mathbf{Id}$ )
- 

$$\begin{aligned}
 \mathcal{J}(\mathbf{y}) &= \sum_{i=1}^N |\kappa(y_i)| = \sum_{i=1}^N \left| \sum_{j=1}^N G_{ij}^4 \kappa(s_j) \right| \\
 &\leq \sum_{i=1}^N \sum_{j=1}^N G_{ij}^4 |\kappa(s_j)| \quad \text{par convexité} \\
 &\leq \sum_{i=1}^N \sum_{j=1}^N G_{ij}^2 |\kappa(s_j)| \quad \text{car } |G_{ij}| \leq 1 \\
 &\leq \sum_{j=1}^N |\kappa(s_j)| \quad \text{car } \sum_{i=1}^N G_{ij}^2 = 1
 \end{aligned}$$

## Justification contraste p.108 (3/3: si majorant atteint)

- En cas d'égalité précédemment, puisque  $\kappa(s_j) \neq 0$ :

$$\sum_{i=1}^N \sum_{j=1}^N (G_{ij}^2 - G_{ij}^4) |\kappa(s_j)| = 0 \text{ entraîne } G_{ij} \in \{0, -1, +1\}$$

- $\mathbf{G}$  étant orthogonale par blanchiment, c'est une matrice «triviale» (permutation+changement de signe)

# Mise en œuvre

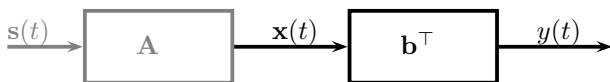
- estimation cumulants: à partir **estimateurs empiriques** des moments
- paramétrisation de  $\mathbf{Q}$  par des angles de **Jacobi**
- solution explicite de la maximisation uni-dimensionnelle et balayage cyclique des angles de Jacobi (voir algos CoM2 et CoM1).
- autres algorithmes disponibles: JADE, FastICA



## Kurtosis (1/2)

- Cadre:

- ▶ extraction d'une source sans blanchiment
- ▶ sources indépendantes
- ▶  $\kappa(s_i) \neq 0$  pour tout  $i$  dont extraction souhaitée
- ▶  $\mathbf{b}^\top$ : filtre ligne extracteur, sortie scalaire:  $y = \sum_{i=1}^N b_i x_i$



- Contraste:

$$\max. \mathcal{J}(y) = |\kappa(y)| \text{ sous contrainte } \mathbb{E}\{|y|^2\} = 1$$

Rq: forme normalisée pour implantation pratique:

$$\max. \left| \frac{\kappa(y)}{\mathbb{E}\{|y|^2\}^2} \right| \text{ sous contrainte } \mathbb{E}\{|y|^2\} = 1$$

## Justification du contraste p. 113

Avec  $\mathbf{g}^\top := \mathbf{b}^\top \mathbf{A}$ , on a  $y = \mathbf{b}^\top \mathbf{x} = \mathbf{b}^\top \mathbf{A} \mathbf{s}$ .

$$\begin{aligned}
 y &= \mathbf{g}^\top \mathbf{s} = \sum_{i=1}^N g_i s_i \\
 \kappa(y) &= \text{Cum} \left\{ \mathbf{g}^\top \mathbf{x}, \mathbf{g}^\top \mathbf{x}, \mathbf{g}^\top \mathbf{x}, \mathbf{g}^\top \mathbf{x} \right\} \\
 &= \sum_{i=1}^N g_i^4 \kappa(s_i) \quad (\text{multilinéarité+indépendance})
 \end{aligned}$$

Sous la contrainte  $\|\mathbf{g}\|^2 = \sum_{i=1}^N g_i^2 = 1 \leftrightarrow \mathbb{E}\{|y|^2\} = 1$ :

$$|\kappa(y)| \leq \sum_{i=1}^N g_i^4 |\kappa(s_i)| \leq \sum_{i=1}^N g_i^2 |\kappa(s_i)| \leq \max_{i=1}^N |\kappa(s_i)|$$

Maximum atteint si et seulement si  $\mathbf{g}$  est séparent.

## Kurtosis (2/2)

- Contraste kurtosis au carré:

$$\max. \mathcal{J}(y) = (\kappa(y))^2 \text{ sous contrainte } \mathbb{E}\{|y|^2\} = 1$$

- Contraste kurtosis en valeur absolue (équivalent):

$$\max. |\kappa(y)| \text{ sous contrainte } \mathbb{E}\{|y|^2\} = 1$$

- Si  $\kappa(s_i) < 0$  pour tout  $i$  dont extraction souhaitée:

$$\min. \kappa(y) \text{ sous contrainte } \mathbb{E}\{|y|^2\} = 1$$

- Rq: formes normalisées (sous contrainte  $\mathbb{E}\{|y|^2\} = 1$ ):

$$\max. \left( \frac{\kappa(y)}{\mathbb{E}\{|y|^2\}^2} \right)^2 \quad \left| \quad \max. \left| \frac{\kappa(y)}{\mathbb{E}\{|y|^2\}^2} \right| \quad \left| \quad \min. \frac{\kappa(y)}{\mathbb{E}\{|y|^2\}^2} \right.$$

## Module constant (CMA)

- Valeurs dans  $\mathbb{C}$ , signaux centrés et circulaires:

$$\kappa(y) = \text{Cum}\{y, y, y^*, y^*\} = \mathbb{E}\{|y|^4\} - \underbrace{|\mathbb{E}\{y^2\}|^2}_{=0 \text{ (circularité)}} - 2\mathbb{E}\{|y|^2\}^2$$

- Contraste si  $\kappa(s_i) < 0$ :

$$\min. \kappa(y) = \mathbb{E}\{|y|^4\} - 2$$

- Algorithme **module constant** (CMA):

$$\min. \mathbb{E}\{(|y|^2 - \alpha)^2\} = \mathbb{E}\{|y|^4\} - 2\alpha\mathbb{E}\{|y|^2\} + \alpha^2$$

↪ identiques avec la contrainte  $\alpha = \mathbb{E}\{|y|^2\} = 1$ .

# Mise en œuvre

- estimation cumulants: à partir **estimateurs empiriques** des moments
- algorithme:
  - ▶ gradient / gradient projeté
  - ▶ forme normalisée des critères
  - ▶ pas optimal à chaque optimisation unidimensionnelle: racine d'un polynôme
- extraction de plusieurs sources par déflation

## Déflation

- Mélange  $\mathbf{x} = \mathbf{A}\mathbf{s}$ .

Une source connue ou précédemment extraite:  $y = \alpha s_1$   
(facteur multiplicatif  $\alpha$  près)

- «**Déflation**»: recherche d'un vecteur colonne  $\mathbf{h}$  tel que:

$$\tilde{\mathbf{h}} := \arg \min_{\mathbf{h}} \|\mathbf{x} - \mathbf{h}y\|^2 \text{ et définir: } \tilde{\mathbf{x}} := \mathbf{x} - \tilde{\mathbf{h}}y$$

- Mélange de  $N - 1$  sources après déflation:  $\tilde{\mathbf{x}} = \sum_{i=2}^N \mathbf{a}_i s_i$   
En effet:

$$\|\mathbf{x} - \mathbf{h}y\|^2 = \|(\mathbf{a}_1 - \alpha\mathbf{h})s_1\|^2 + \left\| \sum_{i=2}^N \mathbf{a}_i s_i \right\|^2$$

- Reconstruction de toutes les sources par **extraction/déflation**.

# Maximum de vraisemblance

- Modèle  $\mathbf{x} = \mathbf{A}\mathbf{s}$  paramétré par  $\mathbf{A}$  et la densité de probabilité supposée des sources  $q(\mathbf{s})$ .

⇒ Densité de probabilité de  $\mathbf{x}$ :

$$p(\mathbf{x}|\mathbf{A}, q) = |\det \mathbf{A}|^{-1} q(\mathbf{A}^{-1}\mathbf{x})$$

⇒ Log-vraisemblance normalisée de  $T$  échantillons  $\mathbf{X} = (\mathbf{x}(1), \dots, \mathbf{x}(T))$ :

$$\frac{1}{T} \log p(\mathbf{X}|\mathbf{A}, q) = \frac{1}{T} \sum_{t=1}^T \log q(\mathbf{A}^{-1}\mathbf{x}(t)) - \log |\det \mathbf{A}|$$

## Divergence de Kullback-Leibler

- On vient de voir:

$$\frac{1}{T} \log p(\mathbf{X}|A, q) = \frac{1}{T} \sum_{t=1}^T \log q(\mathbf{A}^{-1} \mathbf{x}(t)) - \log |\det \mathbf{A}|$$

- d'où:

$$\frac{1}{T} \log p(\mathbf{X}|\mathbf{A}, q) \xrightarrow{T \rightarrow \infty} -D(\mathbf{A}^{-1} \mathbf{x}|\mathbf{s}) + \text{cste}$$

⇒ Contraste associé:

$$\phi_{\text{MV}}(\mathbf{y}) = -D(\mathbf{A}^{-1} \mathbf{x}|\mathbf{s})$$

↔ principe «Infomax», connu dans la communauté réseaux de neurones.



## Information mutuelle (1/2)

- $\tilde{\mathbf{y}}$  vecteur aléatoire:
  - composantes indépendantes
  - mêmes marginales que  $\mathbf{y}$
- $I(\mathbf{y}) = D(\mathbf{y}|\tilde{\mathbf{y}}) =$  information mutuelle du vecteur  $\mathbf{y}$ .
- 

$$D(\mathbf{y}|\mathbf{s}) = D(\mathbf{y}|\tilde{\mathbf{y}}) + D(\tilde{\mathbf{y}}|\mathbf{s})$$

## Information mutuelle (2/2)



$$D(\mathbf{y}|\mathbf{s}) = D(\mathbf{y}|\tilde{\mathbf{y}}) + D(\tilde{\mathbf{y}}|\mathbf{s})$$

max vrais = info mut + adéquation données

- Max de vraisemblance optimise la somme:
  - terme d'indépendance
  - terme ajustement à la distribution supposée de  $\mathbf{s}$ .
- ⇒ Contraste associé:  $\phi_{\text{IM}}(\mathbf{y}) = -I(\mathbf{y})$

# Entropie différentielle

- Définition:

$$H(\mathbf{y}) = -\mathbb{E}\{\log p_{\mathbf{y}}(\mathbf{y})\} = -\int p_{\mathbf{y}}(\mathbf{y}) \log p_{\mathbf{y}}(\mathbf{y}) d\mathbf{y}$$

- Lien avec l'information mutuelle:

$$I(\mathbf{y}) = \sum_{i=1}^N H(y_i) - H(\mathbf{y})$$

## Néguentropie (1/2)

- $\mathbf{y}_{\text{gauss}}$  vecteur/variable aléatoire:
  - gaussienne,
  - même matrice de covariance que  $\mathbf{y}$
- Néguentropie:

$$J(\mathbf{y}) = H(\mathbf{y}_{\text{gauss}}) - H(\mathbf{y})$$

- Néguentropie:

$$J(\mathbf{y}) = D(\mathbf{y}|\mathbf{y}_{\text{gauss}})$$

- Conséquence:  $J(\mathbf{y}) \geq 0$  et  $J(\mathbf{y}) = 0$  si et seulement si  $\mathbf{y} = \mathbf{y}_{\text{gauss}}$ .

## Néguentropie (2/2)

- Avec  $\mathbf{y} = \mathbf{G}\mathbf{s}$ :

$$\phi_{\text{IM}}(\mathbf{y}) = -I(\mathbf{y}) = -\sum_{i=1}^N H(y_i) + \log |\det \mathbf{G}| + H(\mathbf{s})$$

- Sous contrainte de blancheur  $|\det \mathbf{G}| = \text{cste}$

⇒ Contraste associé:

$$\phi_{\text{IM}}^o(\mathbf{y}) = -\sum_{i=1}^N H(y_i) = \sum_{i=1}^N J(y_i) + \text{cste}$$

→ maximiser la «non-gaussianité»

## Contrastes

- Regarder l'écart entre les cumulants seulement au lieu des densités de probabilité:

$$\phi_2(\mathbf{y}) \triangleq \sum_{ij} (\kappa_{ij}(\mathbf{y}) - \kappa_{ij}(\mathbf{s}))^2$$

$$\phi_4(\mathbf{y}) \triangleq \sum_{ijkl} (\kappa_{ijkl}(\mathbf{y}) - \kappa_{ijkl}(\mathbf{s}))^2$$

→  $\phi_2$  ne traduit que la décorrélation.

$\phi_4$  est un contraste.

- Approximation du maximum de vraisemblance:

$$\phi_{\text{MV}}(\mathbf{y}) = -D(\mathbf{y}|\mathbf{s}) \approx -\frac{1}{48} (12\phi_2(\mathbf{y}) + \phi_4(\mathbf{y}))$$

## Contrastes, cas «orthogonal»

- blanchiment  $\Rightarrow$  contrainte  $\phi_2(\mathbf{y}) = 0$  imposée
- Dans ce cas,  $\phi_4$  devient:

$$\phi_4^o(\mathbf{y}) = -2 \sum_i k_i \kappa_{iiii}(\mathbf{y}) \quad (\mathcal{C}_4 \text{ dans TP})$$

où  $k_i$  = cumulant supposé des sources dans la modélisation.

- Minimisation de  $\phi_4^o$  par rapport aux  $k_i$  donne:

$$\phi_{\text{ICA}}^o(\mathbf{y}) = - \sum_i \kappa_{iiii}(\mathbf{y})^2 = \sum_{i,j,k,l \neq i,i,i,i} \kappa_{ijkl}(\mathbf{y})^2 + \text{cste}$$

( $\mathcal{C}_3$  dans TP)

# Conclusion

- Mélange linéaire instantané, cadre aveugle
- Hypothèses: sources indépendantes, non gaussiennes
- Approches:
  - ▶ globales, avec blanchiment
  - ▶ extraction d'une source + déflation (sans blanchiment)
- Fonctions de contraste
  - ▶ statistiques d'ordre supérieur
  - ▶ kurtosis
- Points non abordés:
  - ▶ prise en compte du bruit
  - ▶ algorithmes
  - ▶ signaux dans  $\mathbb{C}$  (avec éventuelle circularité)
  - ▶ approche maximum de vraisemblance, théorie de l'information
  - ▶ problèmes liés: modèles linéaires convolutifs, égalisation



# Bibliographie



<http://tsi.enst.fr/~cardoso/stuff.html>.



<http://www.cis.hut.fi/projects/ica/>.



<http://www.cnl.salk.edu/~tewon/ICA/display.htm>.



J.-F. Cardoso.

Blind signal separation: statistical principles.

*Proceedings of the IEEE, special issue on blind identification and estimation*, 90(8):2009–2026, October 1998.



P. Comon.

Independent component analysis, a new concept.

*Signal Processing*, 36(3):287–314, April 1994.



A. Hyvärinen and E. Oja.

Independent component analysis: Algorithms and applications.

*Neural Networks*, 13(4-5):411–430, 2000.