La parcimonie: une valeur d'avenir ?

Rémi Gribonval
INRIA Rennes - Bretagne Atlantique

Merci

- **Mentors :**
  - S. Mallat, E. Bacry, R. De Vore, F. Bimbot, ...
- **Doctorants :**
  - L. Benaroya, S. Lesage, A. Ozerov, S. Arberet, B. Mailhé, P. Sudhakar, N. Duong, N. Ito, A. Benichoux, A. Bourrier, ...
- **Postdocs :**
  - N. Bertin, V. Emiya, S. Nam, N. Stefanakis, ...
- **Collègues et co-auteurs :**
  - M. Nielsen, B. Torrésani, L. Daudet, P. Vandergheynst, E. Vincent, C. Fevotte, M. Davies, M. Plumbley, M. Elad, K. Schnass, A. Cohen, F. Bach, R. Jenatton, ...
- **Et aussi bien sûr :**
  - S. Lemaile, H. Van Herbruggen, J.-P. Banatre, P. Gelin, Ch. Le Tonquèze, R. Ronchaud, N. Lacaux, C. Bachelet, ...

L’équipe METISS

Blaise Pascal

- La Pascaline

- Des outils mathématiques:
  - Géométrie
  - Combinatoire
  - Probabilités
  - Calcul automatisé

- Un pari métaphysique ...

Copie d'une peinture de François II Quesnel gravée par Gérard Edelinck en 1691. Source: Wikipedia.

350 ans plus tard ...

- La Pascaline

- Des outils mathématiques:
  - Géométrie
  - Combinatoire
  - Probabilités
  - Calcul automatisé
- Un pari métaphysique ...

- La démixeuse

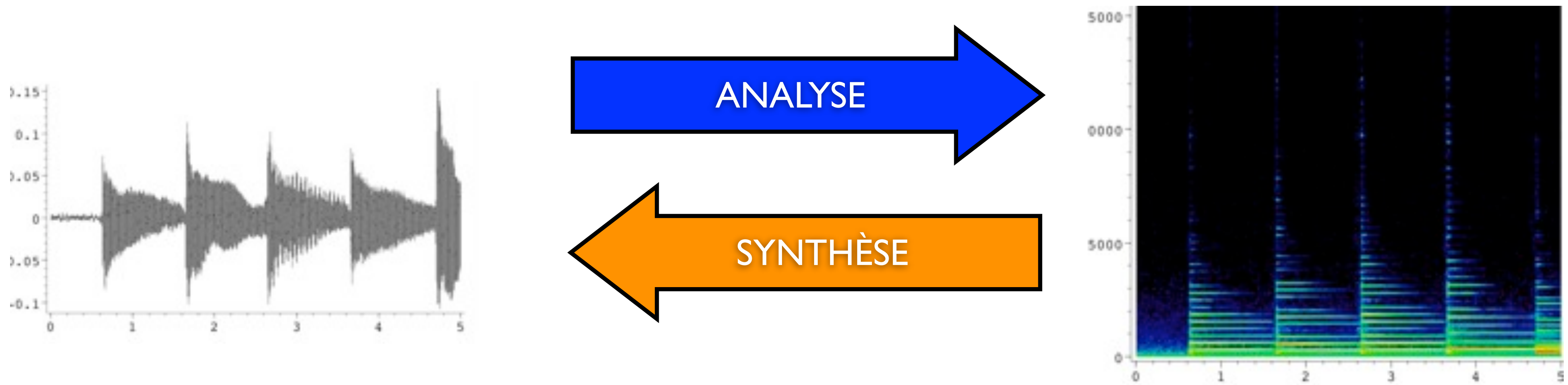
- Des outils mathématiques:
  - Géométrie
  - Combinatoire
  - Probabilités
  - Informatique
- Un «miracle»: la **parcimonie**

Données numériques: le défi du volume

- **Fait** : données numériques = grands volumes
  - ✓ 1 seconde d'audio stéréo, qualité CD = 1,4 Mbit
  - ✓ 1 image non compressée, 10 Mpixels = 240 Mbit
- **Besoins** : représentations «concises»
  - ✓ stockage & transmission (capacité & débits) ...
  - ✓ manipulation & traitement (complexité algorithmique)

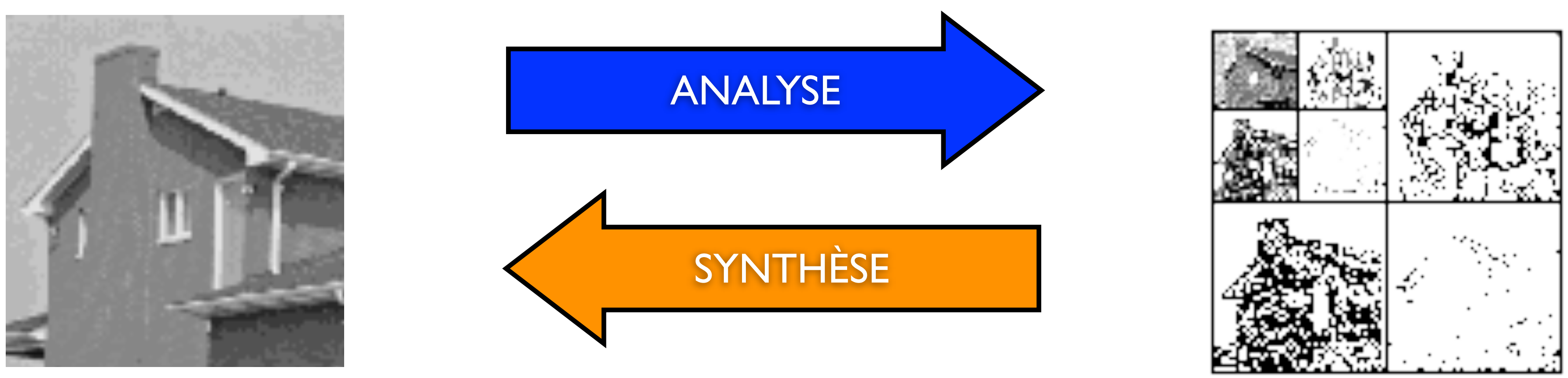
Représentations parcimonieuses

- Audio : représentations temps-fréquence (MP3)



Noir
= zéro

- Images : transformée en ondelettes (JPEG2000)



Blanc
= zéro

Expression mathématique

- Signal / image = vecteur en grande dimension

$$\mathbf{y} \in \mathbb{R}^N$$

- **Modèle** = combinaison linéaire de vecteurs de base (ex: *atomes temps-fréquence, ondelettes*)

$$\mathbf{y} \approx \sum_k x_k \varphi_k = \mathbf{\Phi} x$$

atomes

dictionnaire

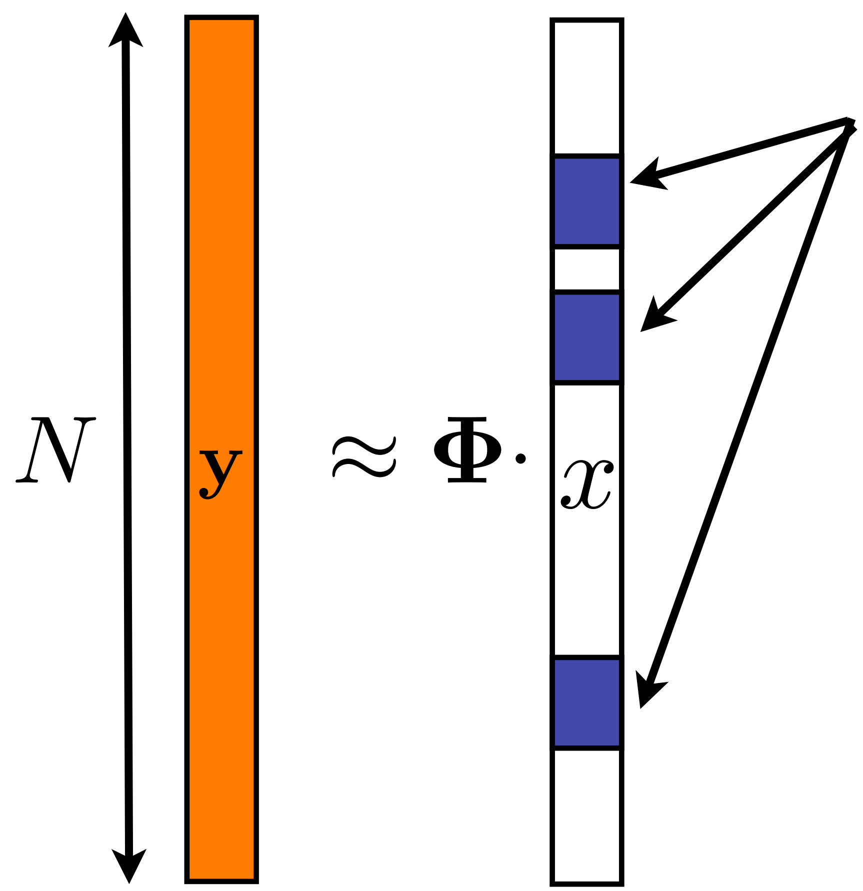
[Mallat & Zhang1993]

- **Parcimonie** = petite (pseudo)-norme L0

$$\|x\|_0 = \sum_k |x_k|^0 = \mathrm{card}\{k, x_k \neq 0\}$$

Parcimonie & compression

- Vecteur plein
- Vecteur parcimonieux



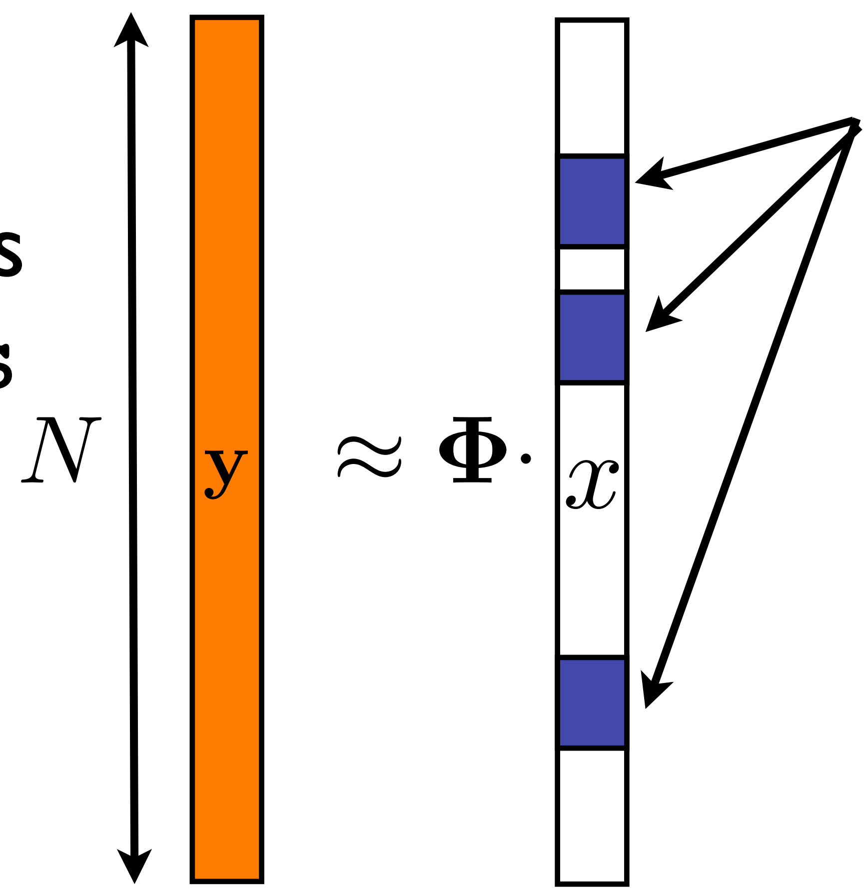
$k \ll N$ coefficients non nuls
= k `floats`

Parcimonieux = économique

Parcimonie & compression

- Vecteur plein
- Vecteur parcimonieux

N coefficients
= N `floats`



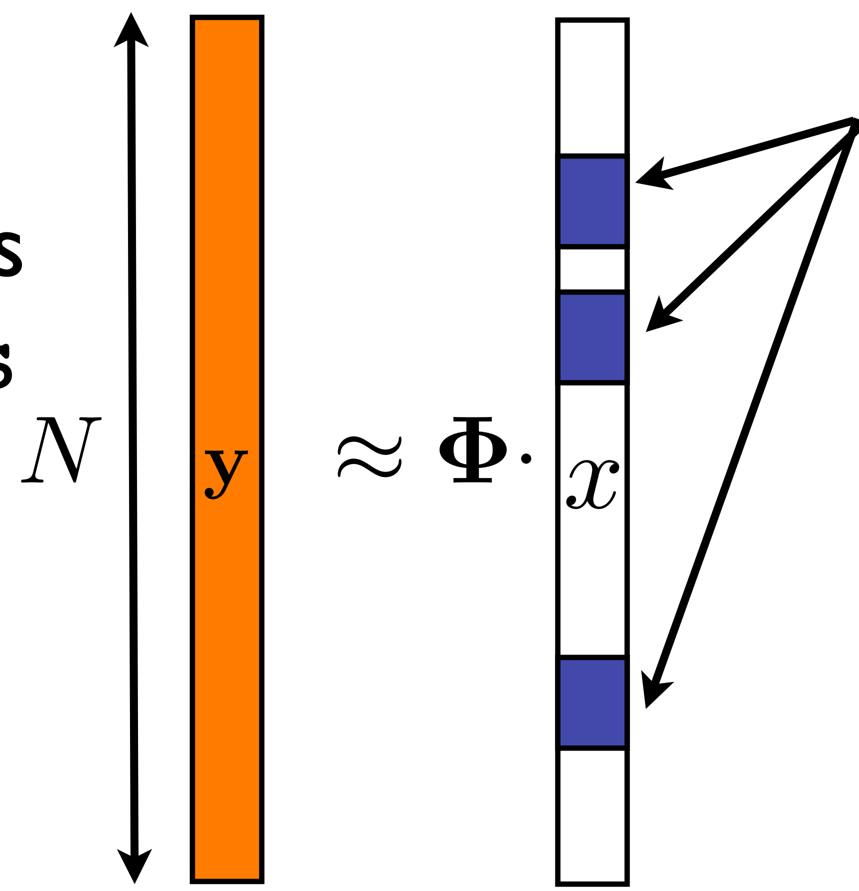
$k \ll N$ coefficients non nuls
= k `floats`

Parcimonieux = économique

Parcimonie & compression

- Vecteur plein

N coefficients
= N `floats`



- Vecteur parcimonieux

$k \ll N$ coefficients non nuls
= k `floats`

+ k positions parmi N

$$= \log_2 \binom{N}{k} \approx k \log_2 \frac{N}{k} \text{ bits}$$

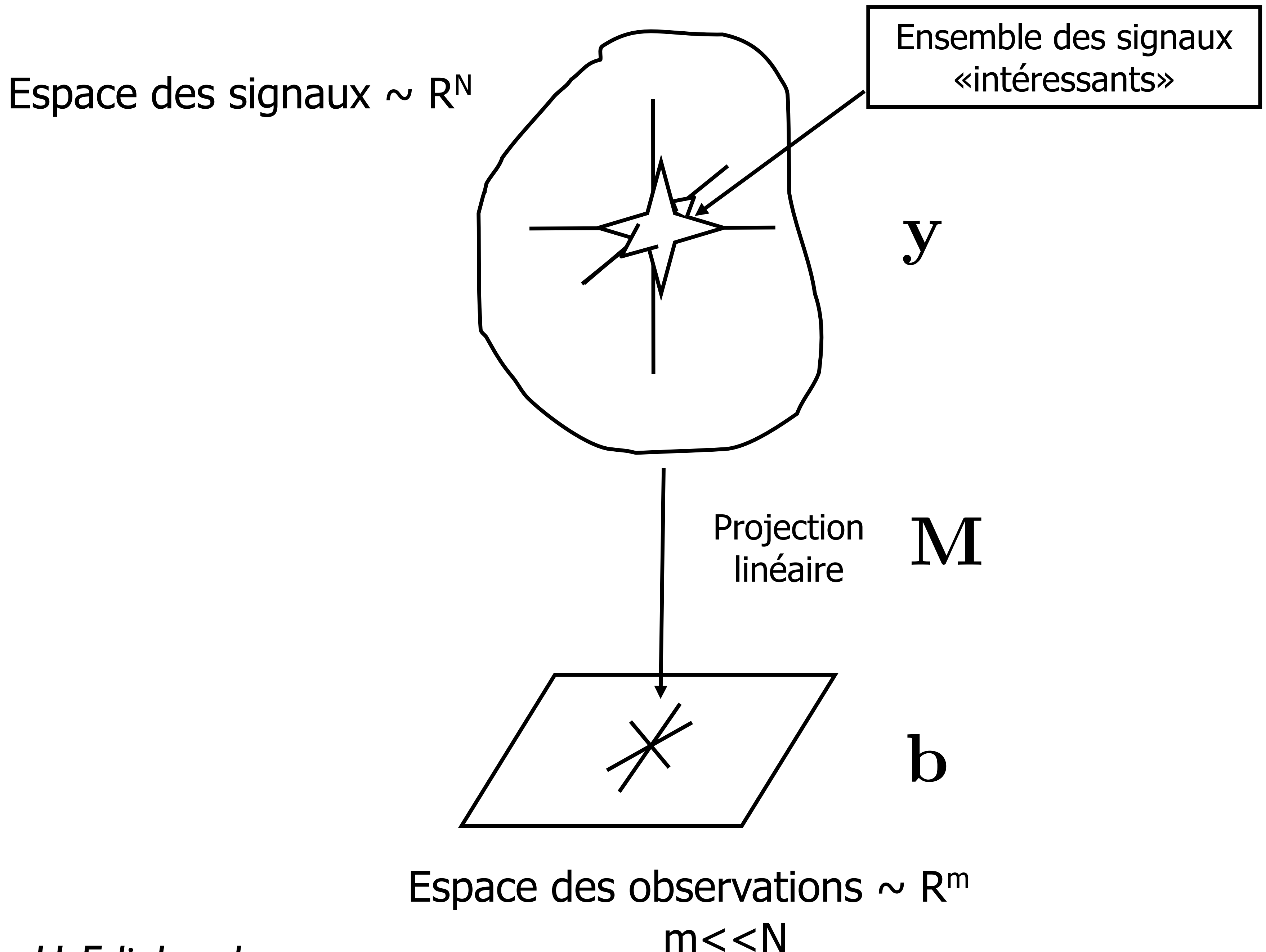
Parcimonieux = économique

Récapitulatif historique

- Années 1990: parcimonie pour la compression
  - au-delà de Fourier : **ondelettes** et approximation non-linéaire
  - concept de **dictionnaire** redondant
  - **algorithmes heuristiques** (*Matching Pursuit*, L1/LASSO/*Basis Pursuit*)
- 2000-2005: parcimonie pour les problèmes inverses
  - **identifiabilité** des représentations parcimonieuse
  - algorithmes d'identification **à performance garantie**
- 2005-
  - **nouveaux algorithmes** (seuillage itératif, LARS/Homotopie)
  - nouveau paradigme d'application : ***Compressed Sensing***
  - rôle des **matrices aléatoires** et de la concentration de la mesure
  - **apprentissage** de dictionnaires

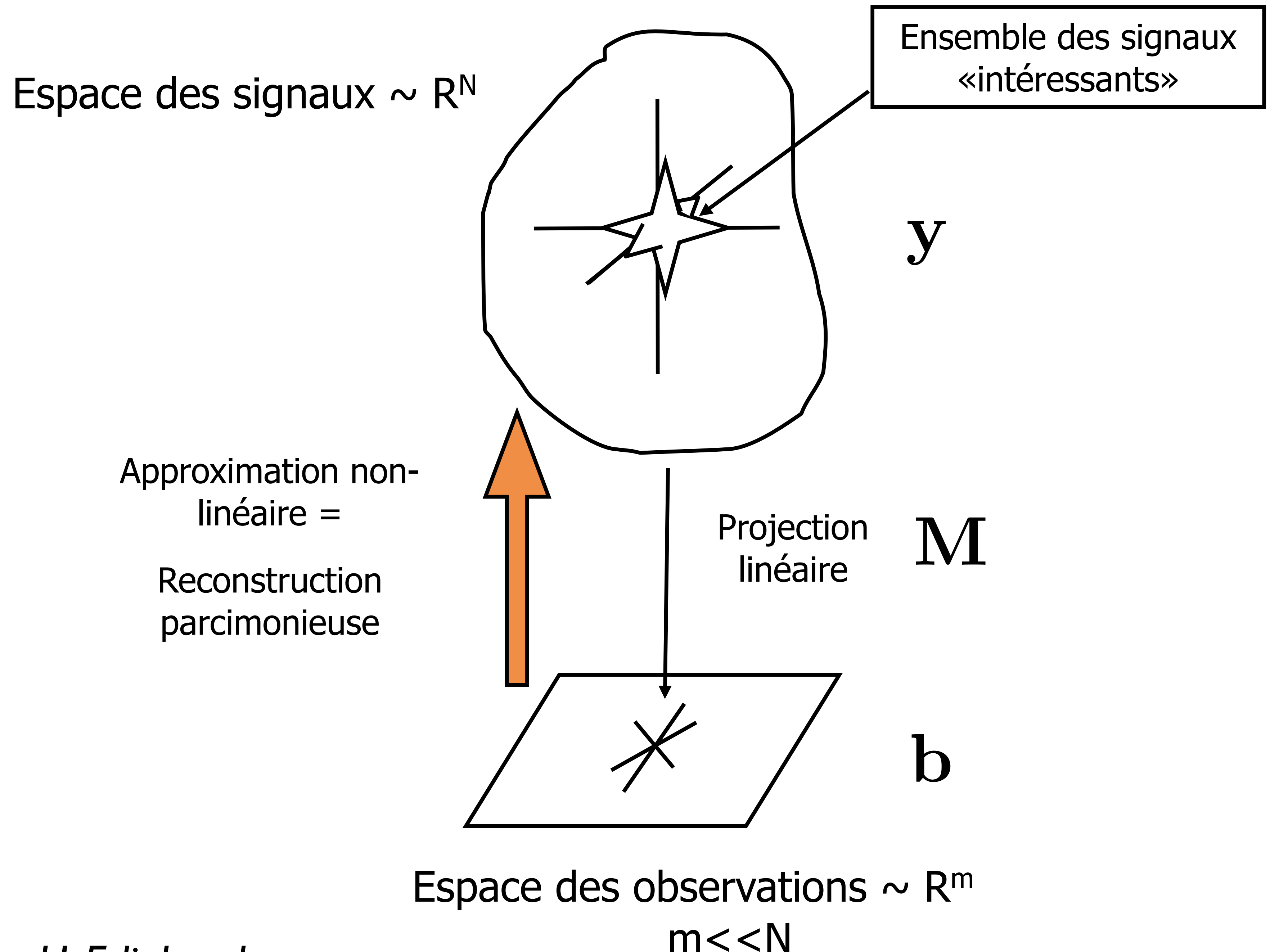
Parcimonie & problèmes inverses

Problèmes inverses: un point de vue géométrique



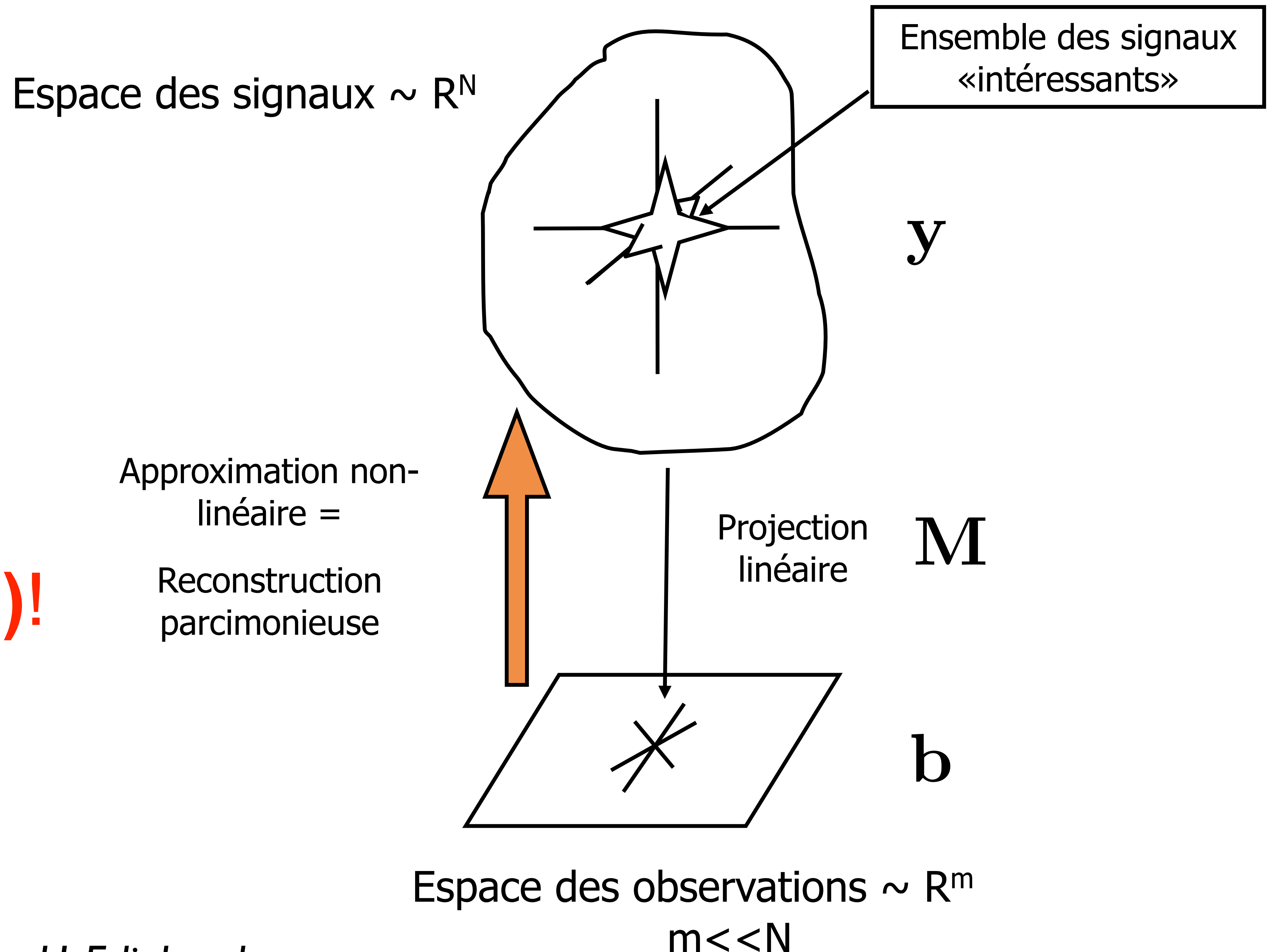
Merci à M. Davies, U. Edinburgh

Problèmes inverses: un point de vue géométrique



Merci à M. Davies, U. Edinburgh

Problèmes inverses: un point de vue géométrique



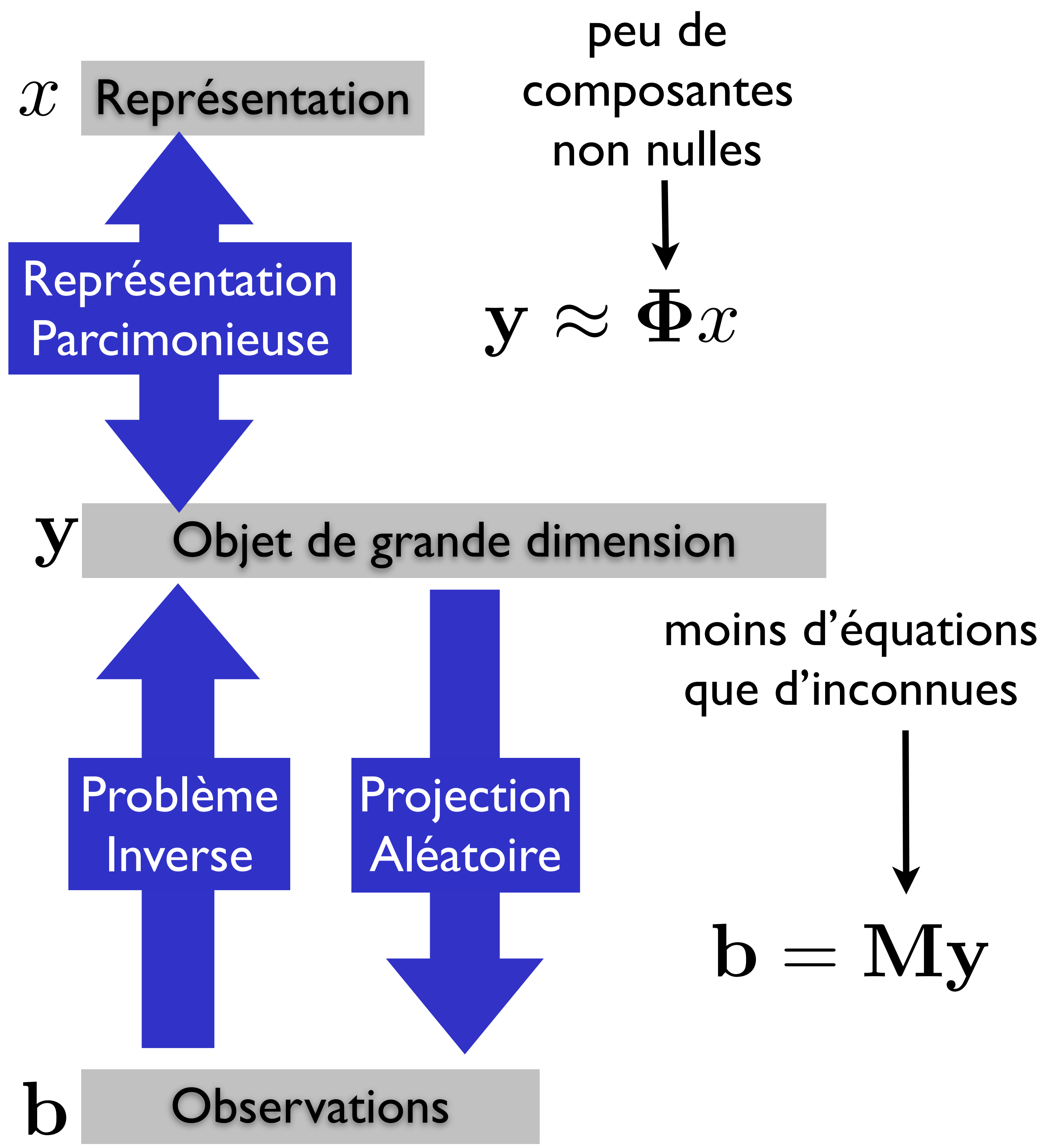
Miracle?

Théorème(s)!

Merci à M. Davies, U. Edinburgh

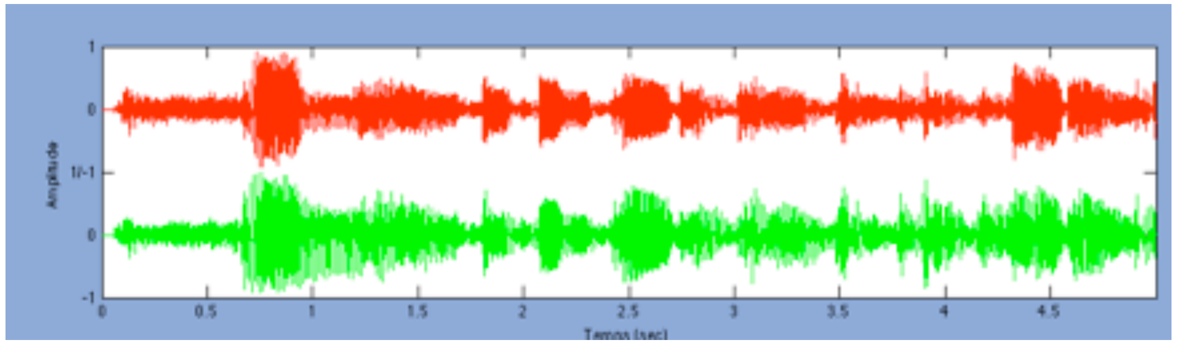
Parcimonie & problèmes inverses

- **Parcimonie** :
  approcher *objets de grande dimension* avec peu de paramètres
  - ✦ ex: signal, image
- **Problème inverse** :
  reconstruire précisément un objet *à partir d'observations incomplètes*
  - ✦ ex: tomographie
- **Projection aléatoire :** réduire *volontairement* la dimension, avec reconstruction stable
  - ✓ garanties théoriques
  - ✓ algorithmes de complexité bornée
    - ✦ ex: acquisition compressée,
    - ✦ **reconnaissance compressée ?**



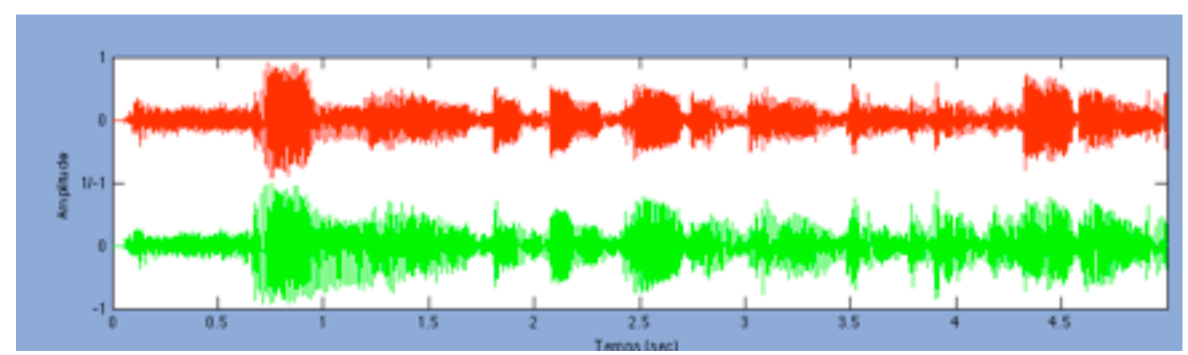
Exemple : séparation de sources audio

- « Softly as in a morning sunrise »



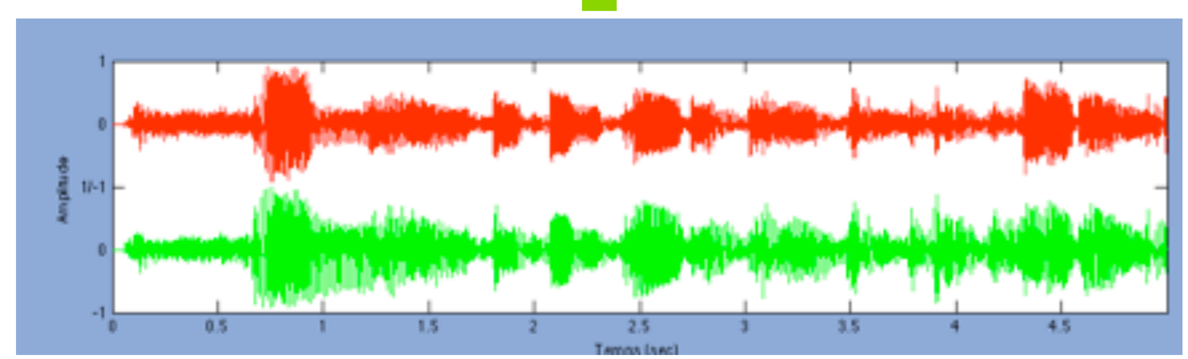
Exemple : séparation de sources audio

- « Softly as in a morning sunrise »



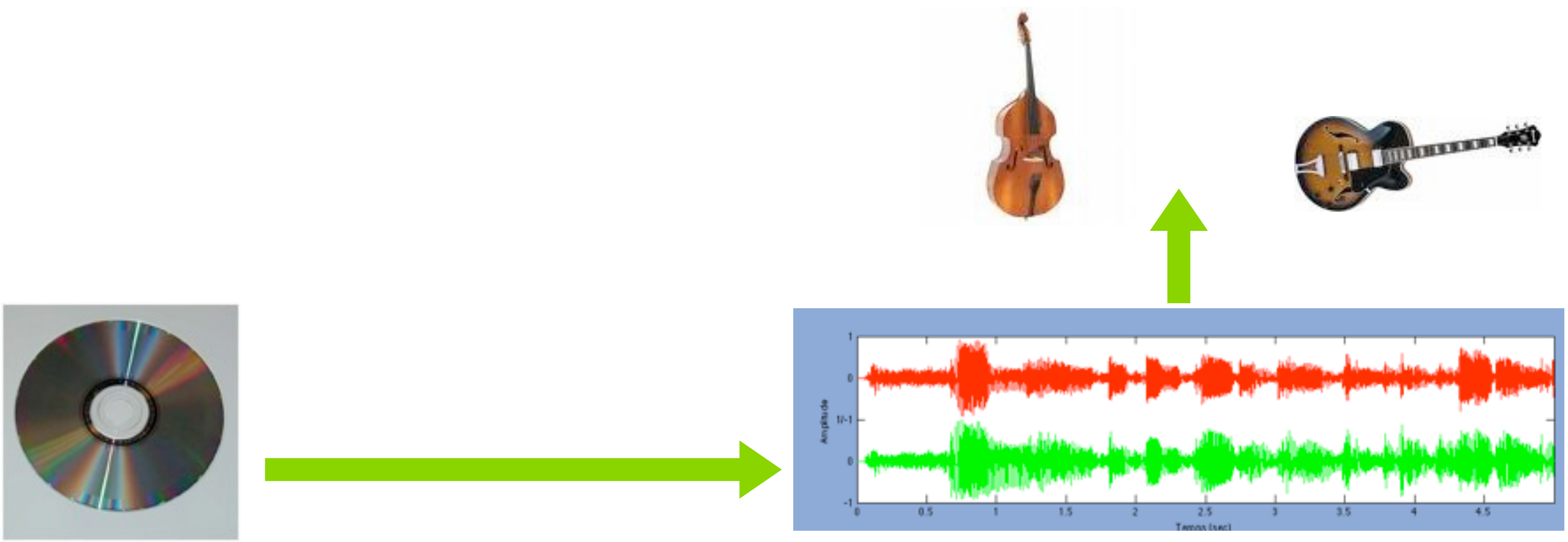
Exemple : séparation de sources audio

- « Softly as in a morning sunrise »



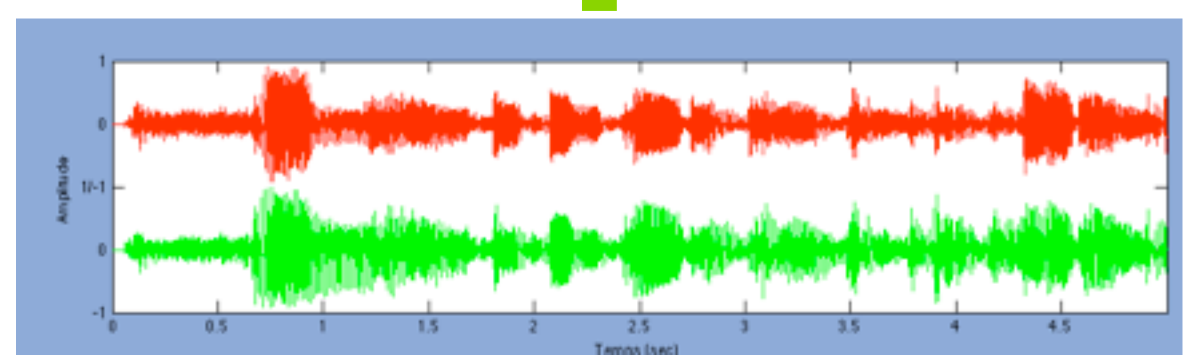
Exemple : séparation de sources audio

- « Softly as in a morning sunrise »



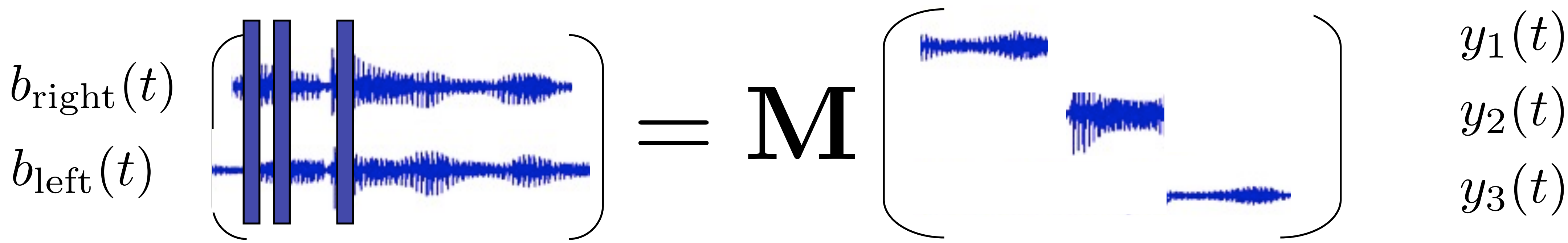
Exemple : séparation de sources audio

- « Softly as in a morning sunrise »

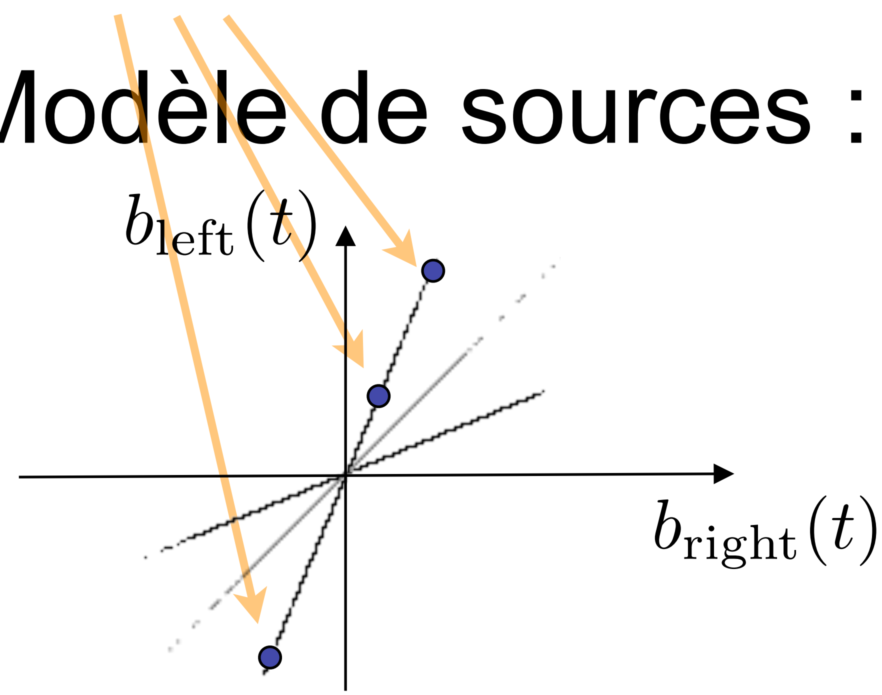


Séparation «aveugle» de sources

- Modèle de mélange: linéaire instantané



- Modèle de sources : supports temporels disjoints …

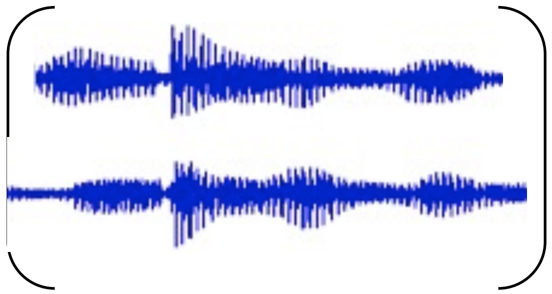


... «clustering» pour :
1- identifier colonnes de la matrice de mélange
2- reconstruire les sources

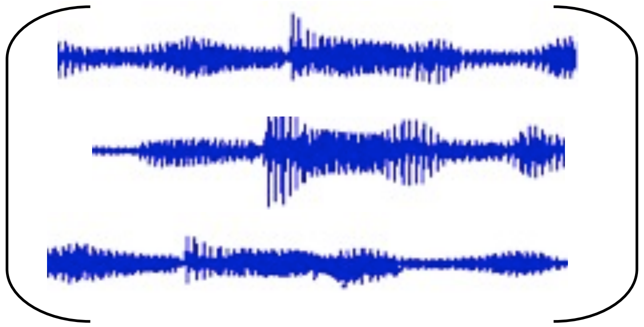
Séparation «aveugle» de sources

- Modèle de mélange: linéaire instantané

$b_{\mathrm{right}}(t)$
$b_{\mathrm{left}}(t)$

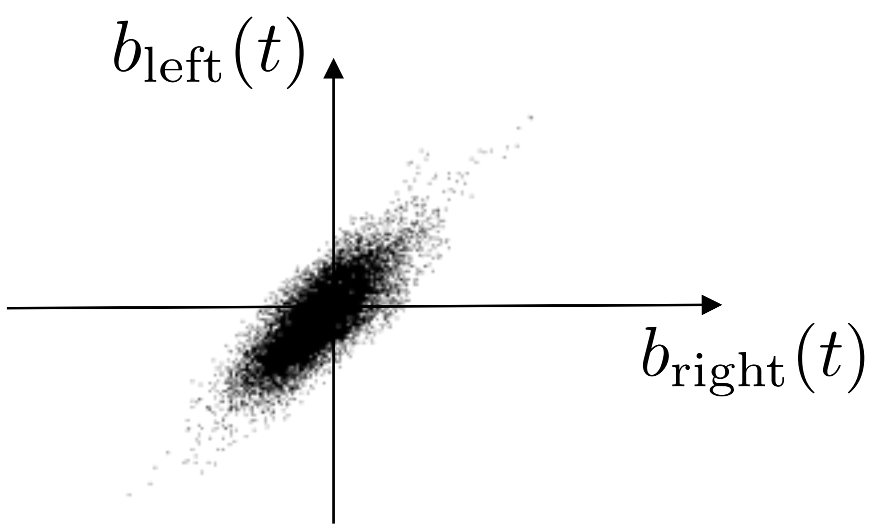


$= \mathbf{M}$



$y_1(t)$
$y_2(t)$
$y_3(t)$

- En pratique ...

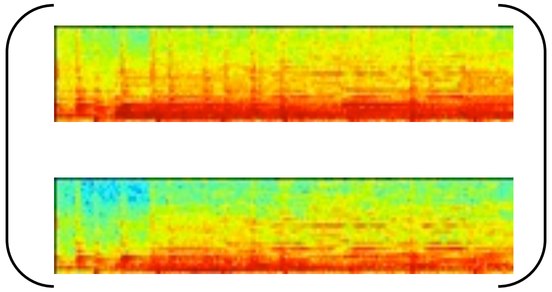


Masquage temps-fréquence

- Modèle de mélange en temps-fréquence

$B_{\text{right}}(\tau, f)$

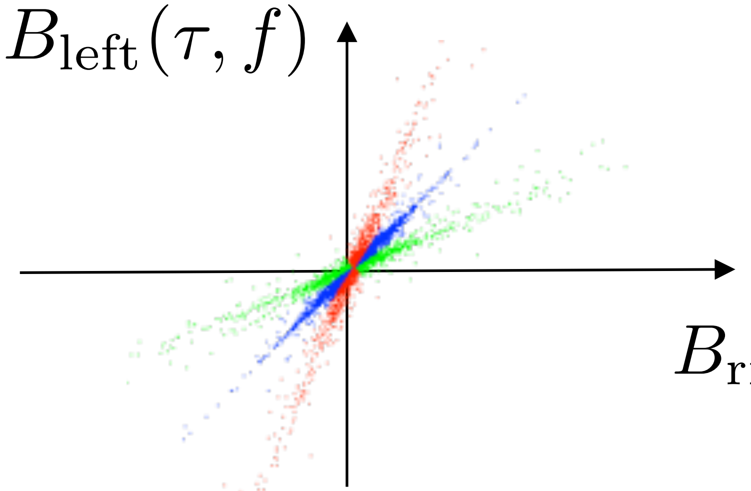
$B_{\text{left}}(\tau, f)$



$$= \mathbf{M}\,\mathbf{Y}(\tau, f)$$

- et "miraculeusement" ...

$B_{\text{left}}(\tau, f)$



$B_{\text{right}}(\tau, f)$

... les représentations temps-fréquence de signaux audio sont (souvent) ***quasi-disjointes****.*

[Jourjine, Rickard & Yilmaz 2000]
[Zibulevsky & Pearlmutter 2001]
[Gribonval 2002]

Fondations mathématiques

- Problème 1990-2000 :
  - ✓ Moins d'équations que d'inconnues = mal posé ...

$$\mathbf{A}x_0 = \mathbf{A}x_1 \not\Rightarrow x_0 = x_1$$

  - ✓ Algorithmes heuristiques de reconstruction de x_0

[Mallat & Zhang 1993] [Gorodnistky & Rao 1997] [Chen Donoho & Saunders 1999]

- Cap franchi en 2001-2006 :
  - ✓ Unicité de la solution parcimonieuse = bien posé!
    - ✦ si x_0, x_1 sont "suffisament parcimonieux",
    - ✦ alors $\mathbf{A}x_0 = \mathbf{A}x_1 \Rightarrow x_0 = x_1$
  - ✓ Garanties de succès des algorithmes heuristiques

Reconstruction parcimonieuse: garanties théoriques

- **Théorème** : soit $\mathbf{y} = \mathbf{A}x_0$

  ✓ si $\|x_0\|_0 \leq k_0(\mathbf{A})$ alors $x_0 = x_0^\star$

  ✓ si $\|x_0\|_0 \leq k_p(\mathbf{A})$ alors $x_0 = x_p^\star$

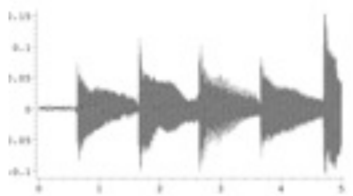
avec $x_p^\star = \arg \min_{\mathbf{A}x=\mathbf{y}} \|x\|_p$

- [Donoho & Huo 01] : *paires de bases, notion de cohérence, p=1*
- [Gribonval & Nielsen 2003] [Donoho & Elad 2003] : *dictionnaires, cohérence*
- [Tropp 2004] : *Orthonormal Matching Pursuit, cohérence cumulative*
- [Candès, Romberg & Tao 2004] : *dictionnaires aléatoires, isométrie restreintes*
- [Gribonval & Nielsen 2007] : *résultats d'interpolation 0<p<1*

La parcimonie se cache-t-elle partout ?

Déluge de données

- Parcimonie : historiquement pour signaux et images
  - ✓ verrou = **passage à l'échelle** algorithmique



Signaux

Images

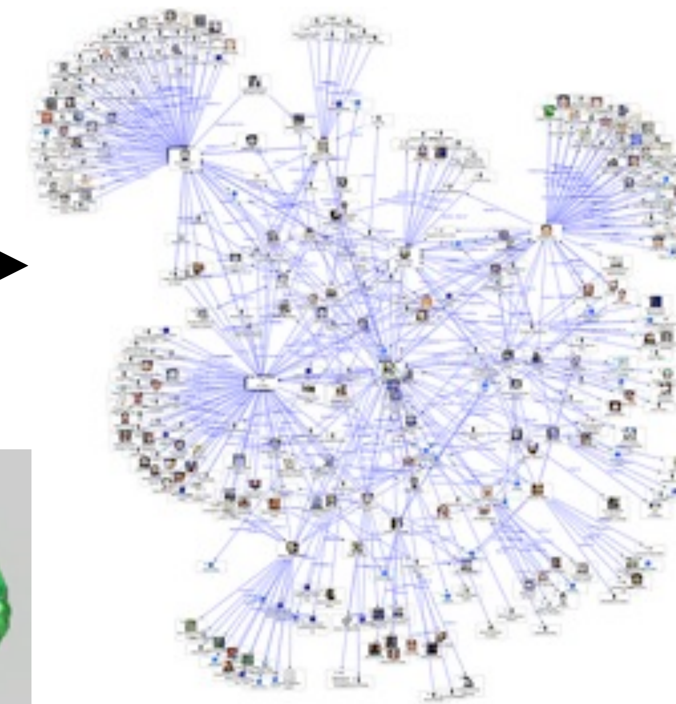
Déluge de données + Jungle

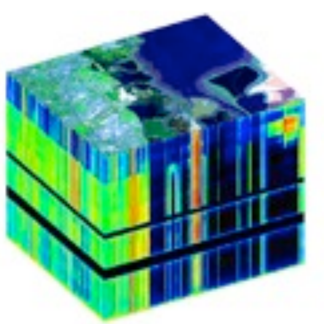
- Parcimonie : historiquement pour signaux et images
  - ✓ verrou = **passage à l’échelle** algorithmique

Signaux

Images

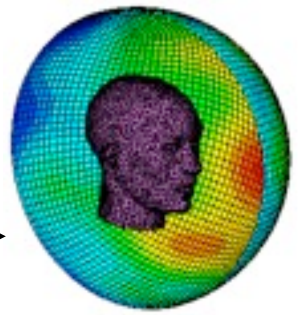
Données sur des graphes
Réseaux sociaux
Connectivité du cerveau

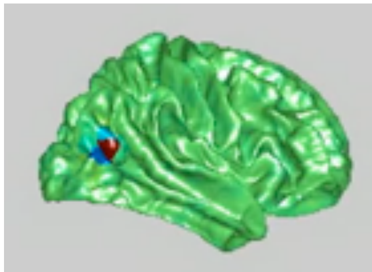




Hyperspectral
Imagerie satellite

Géométrie sphérique
Cosmologie, HRTF (son 3D)





Valeurs vectorielles
Tenseur de diffusion

- Nouvelles données “exotiques” ou composites
  - ✓ verrou = **apprentissage de dictionnaire**

Défis scientifiques d’aujourd’hui ...

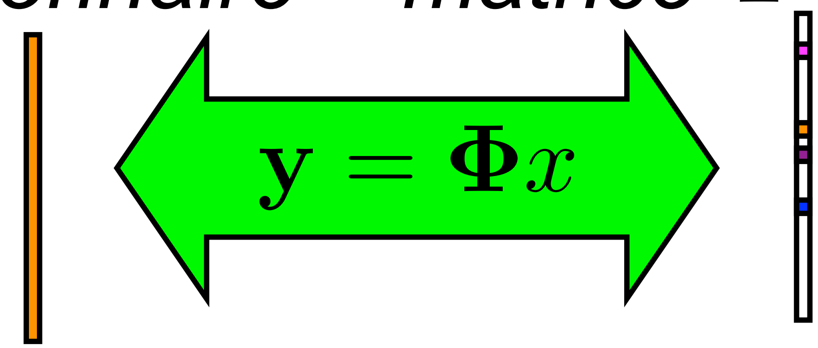
- *Objectifs fondamentaux*
  1. Comprendre le **couplage** entre la parcimonie et les **phénomènes physiques** sous-jacents
  2. Exploiter les **structures** et **invariances** sous-jacentes (graphes, variétés géométriques,...)
  3. Construire des **formalismes génériques** pour modéliser des **données composites** (multicanal, multimodal, ...)
- *Un défi clé :*
  - ✓ **Adapter les modèles** aux données par **apprentissage** à partir de corpus

Découvrir la parcimonie cachée ...

avec K. Schnass, F. Bach, R. Jenatton

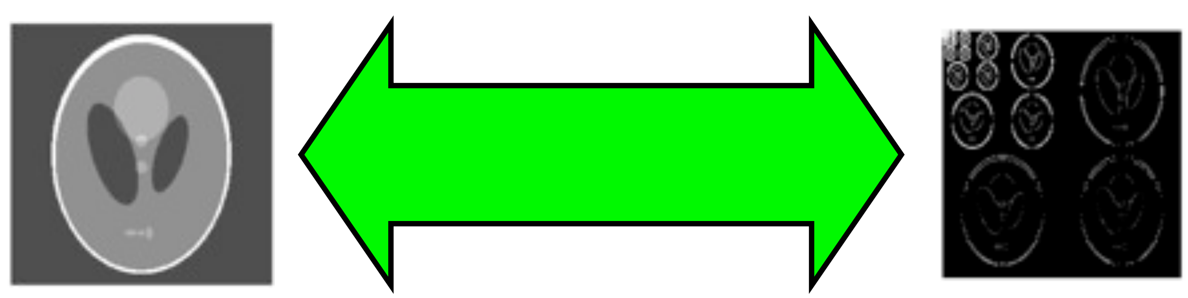
Choix du dictionnaire : De l'analyse harmonique à l'apprentissage

- Notion de *dictionnaire* = *matrice* $\mathbf{\Phi}$



- Propriété souhaitée : **bien approcher les données «intéressantes» avec peu d'*atomes***
- Historiquement : analyse harmonique + choix par un expert parmi une «bibliothèque» de dictionnaires (Fourier, ondelettes, ...)

ex:



Apprentissage de dictionnaire pour les représentations parcimonieuses

- Modèle parcimonieux = choix d'un dictionnaire

Base d'images d'entraînement

$$\mathbf{y}_n = \mathbf{\Phi} x_n, \; 1 \leq n \leq N$$

Apprentissage de dictionnaire pour les représentations parcimonieuses

- Modèle parcimonieux = choix d’un dictionnaire

Base d’images d’entraînement

Exemples d’entraînement

$$\mathbf{y}_n = \mathbf{\Phi} x_n, \; 1 \leq n \leq N$$

Apprentissage de dictionnaire pour les représentations parcimonieuses

- Modèle parcimonieux = choix d'un dictionnaire

Base d'images d'entraînement

Exemples d'entraînement

$$\mathbf{y}_n = \Phi x_n, \quad 1 \leq n \leq N$$

Dictionnaire inconnu	Coefficients parcimonieux inconnus

Apprentissage de dictionnaire pour les représentations parcimonieuses

- Modèle parcimonieux = choix d'un dictionnaire

Base d'images d'entraînement

Exemples d'entraînement

$$\mathbf{y}_n = \mathbf{\Phi} x_n, \quad 1 \leq n \leq N$$

Dictionnaire inconnu

Coefficients parcimonieux inconnus

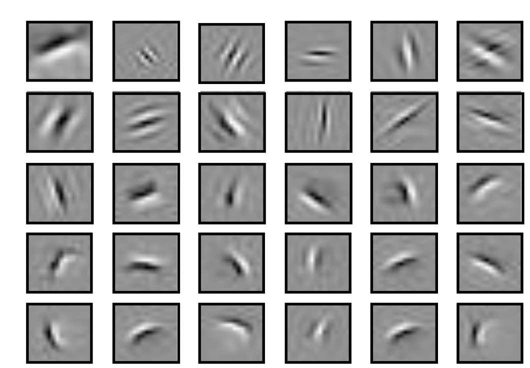
apprentissage

$\hat{\mathbf{\Phi}}$ = atomes ~ contours

[Olshausen & Field 96, Aharon et al 06, Mairal et al 09, ...]

= translations de motifs de base

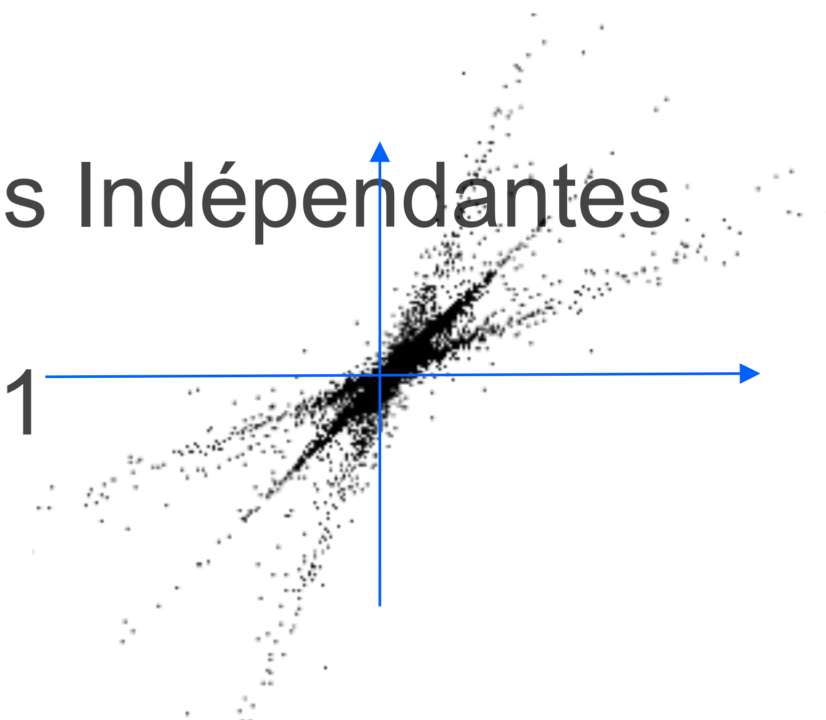
[Blumensath 05, Jost et al 05, ...]



Apprentissage de dictionnaire

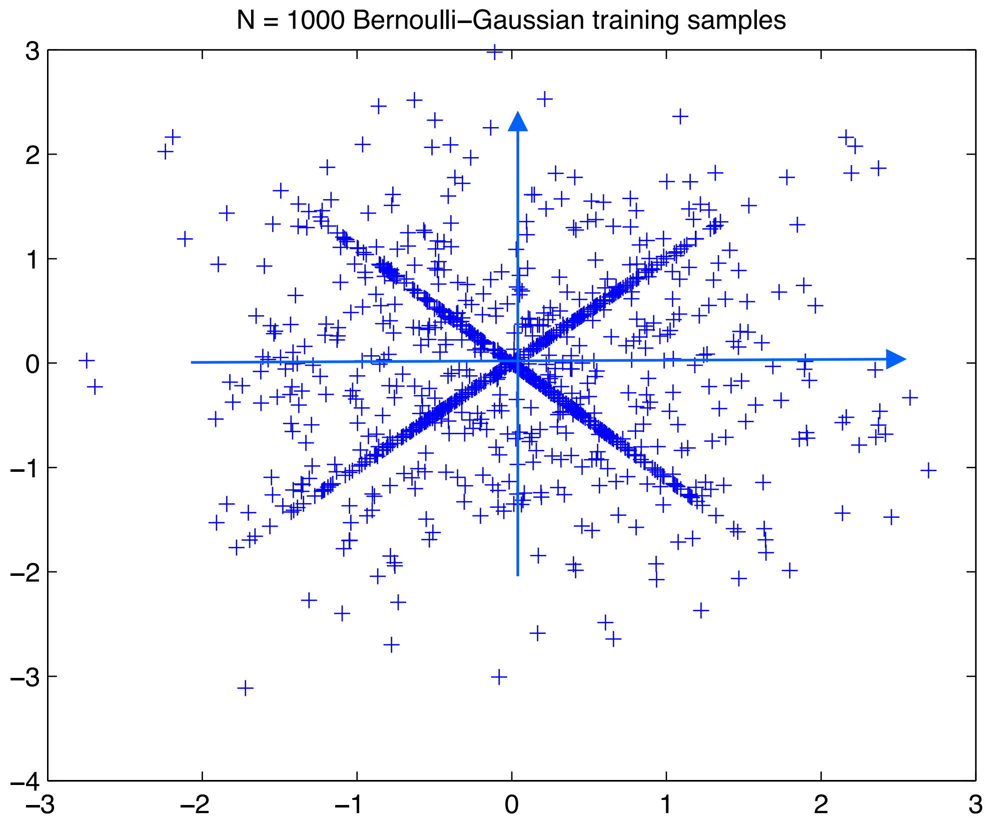
- Formalisation du problème :
  - ✓ N données observées $\mathbf{y}_n = \mathbf{\Phi} x_n \in \mathbb{R}^d \longrightarrow \mathbf{Y} = \mathbf{\Phi} X$
  - ✓ Dictionnaire commun $\mathbf{\Phi}$ inconnu,
  - ✓ Coefficient X inconnus + hypothèse de parcimonie
  - ✓ Objectif: identifier $\mathbf{\Phi}$
- Approches
  - ✓ Etat de l'art : Analyse en Composantes Indépendantes = cadre asymptotique
  - ✓ Approche considérée : minimisation L1

$$\min_{\mathbf{\Phi}, X | \mathbf{Y} = \mathbf{\Phi} X} \|X\|_1$$



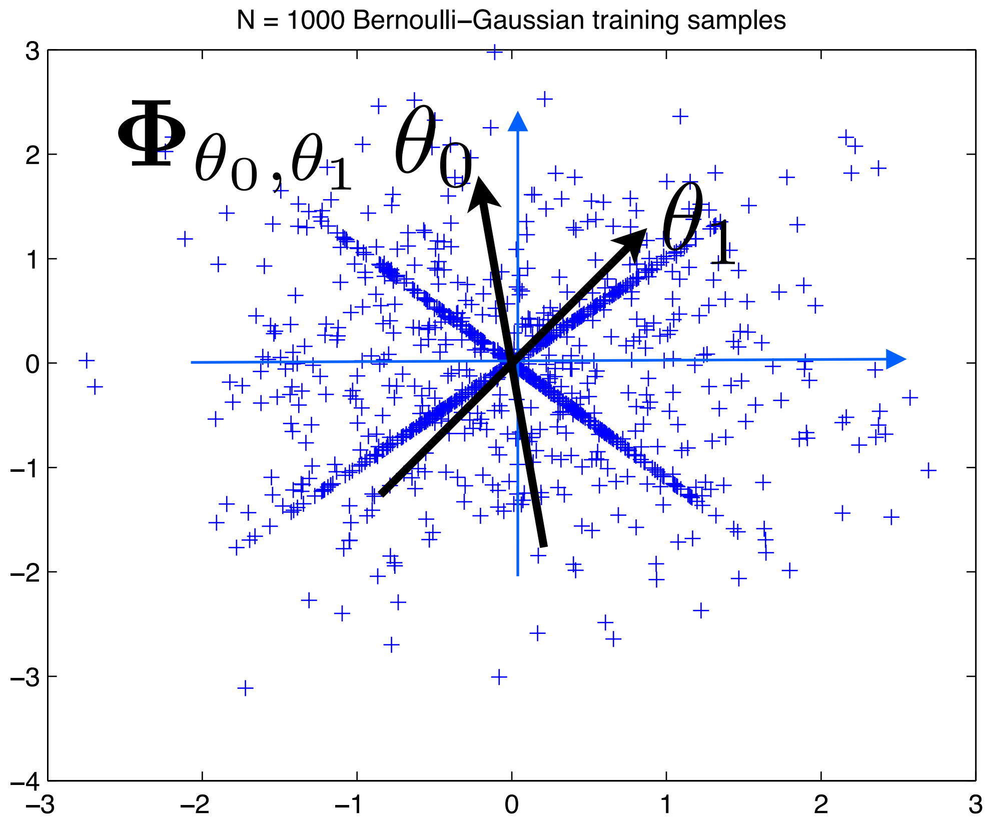
Exemple jouet en 2D

$$\mathbf{Y} = \mathbf{\Phi}_0 X_0$$



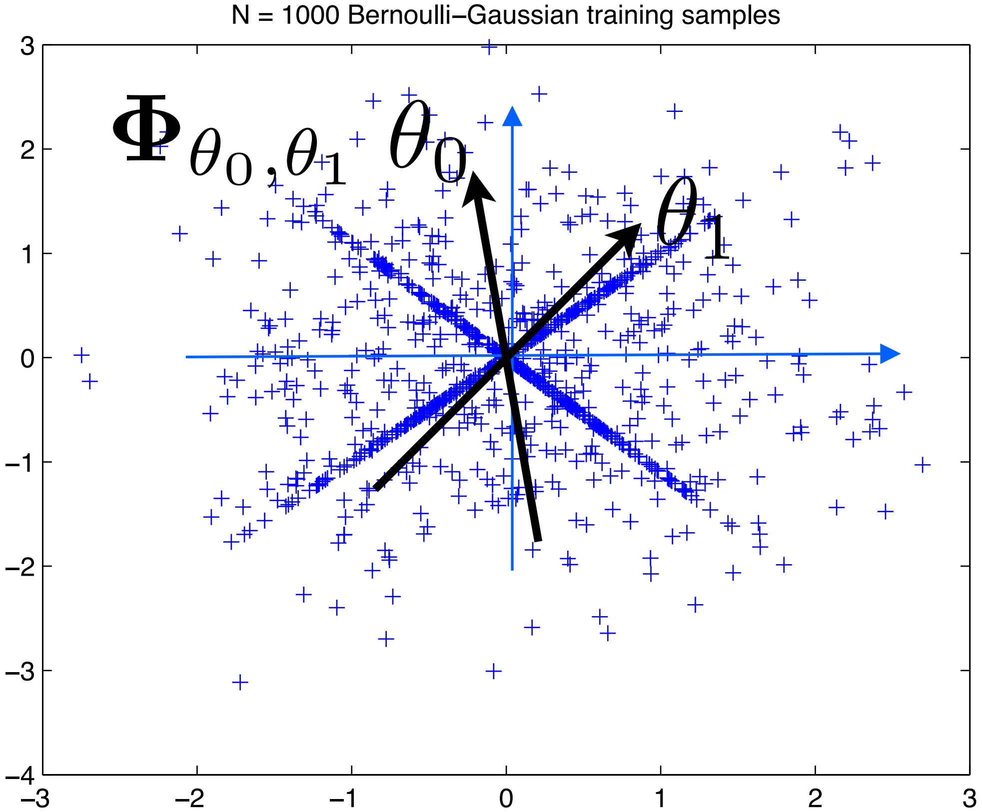
Exemple jouet en 2D

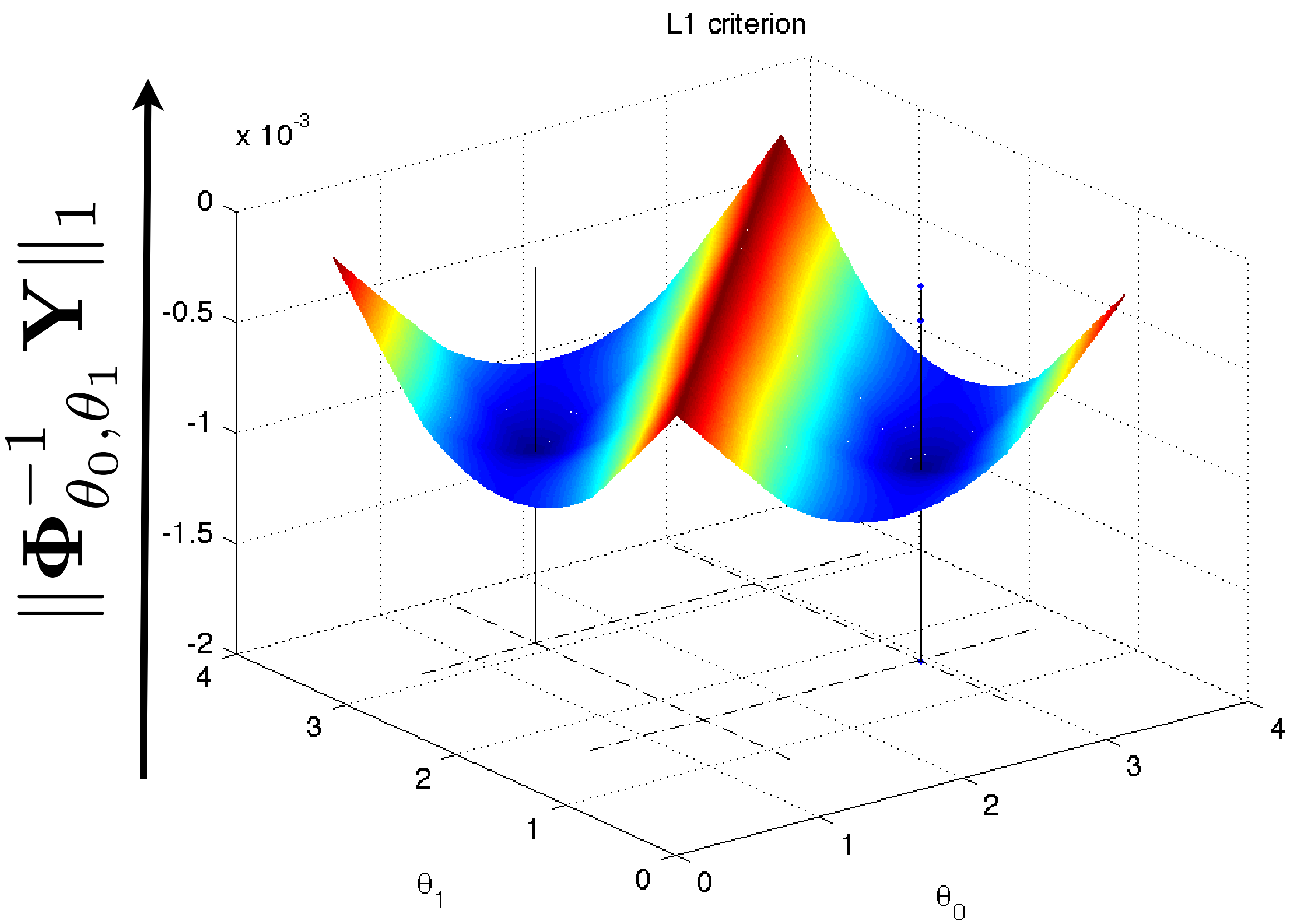
$$\mathbf{Y} = \mathbf{\Phi}_0 X_0$$



Exemple jouet en 2D

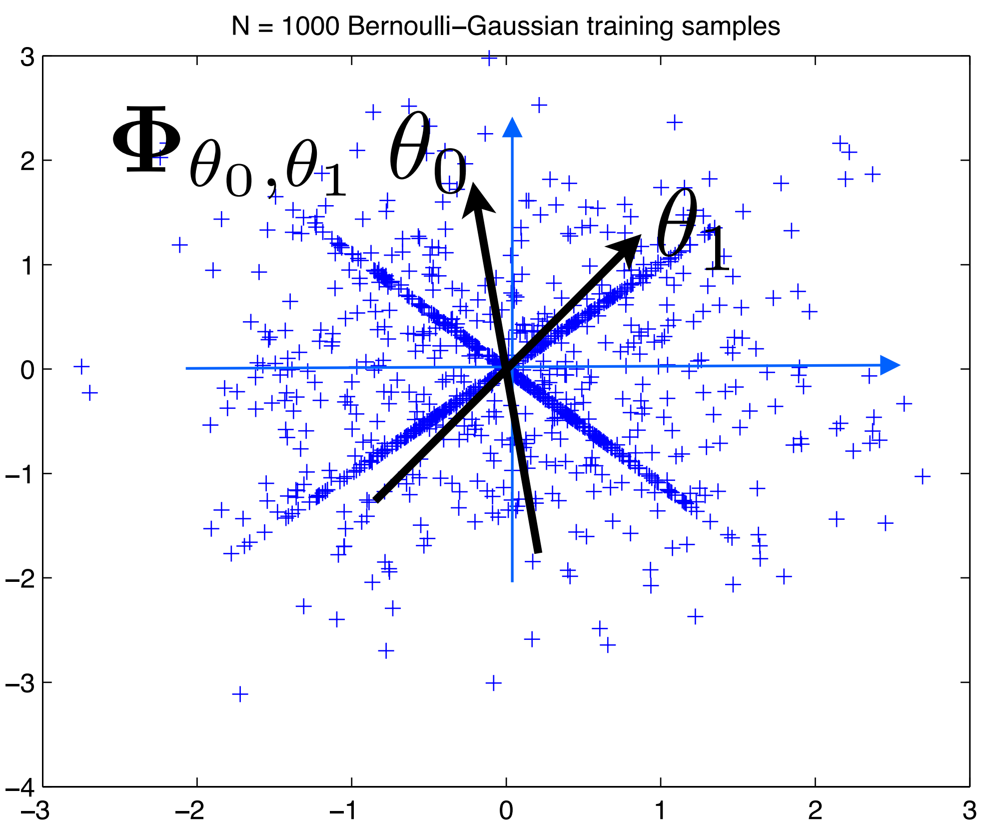
$$\mathbf{Y} = \mathbf{\Phi}_0 X_0$$

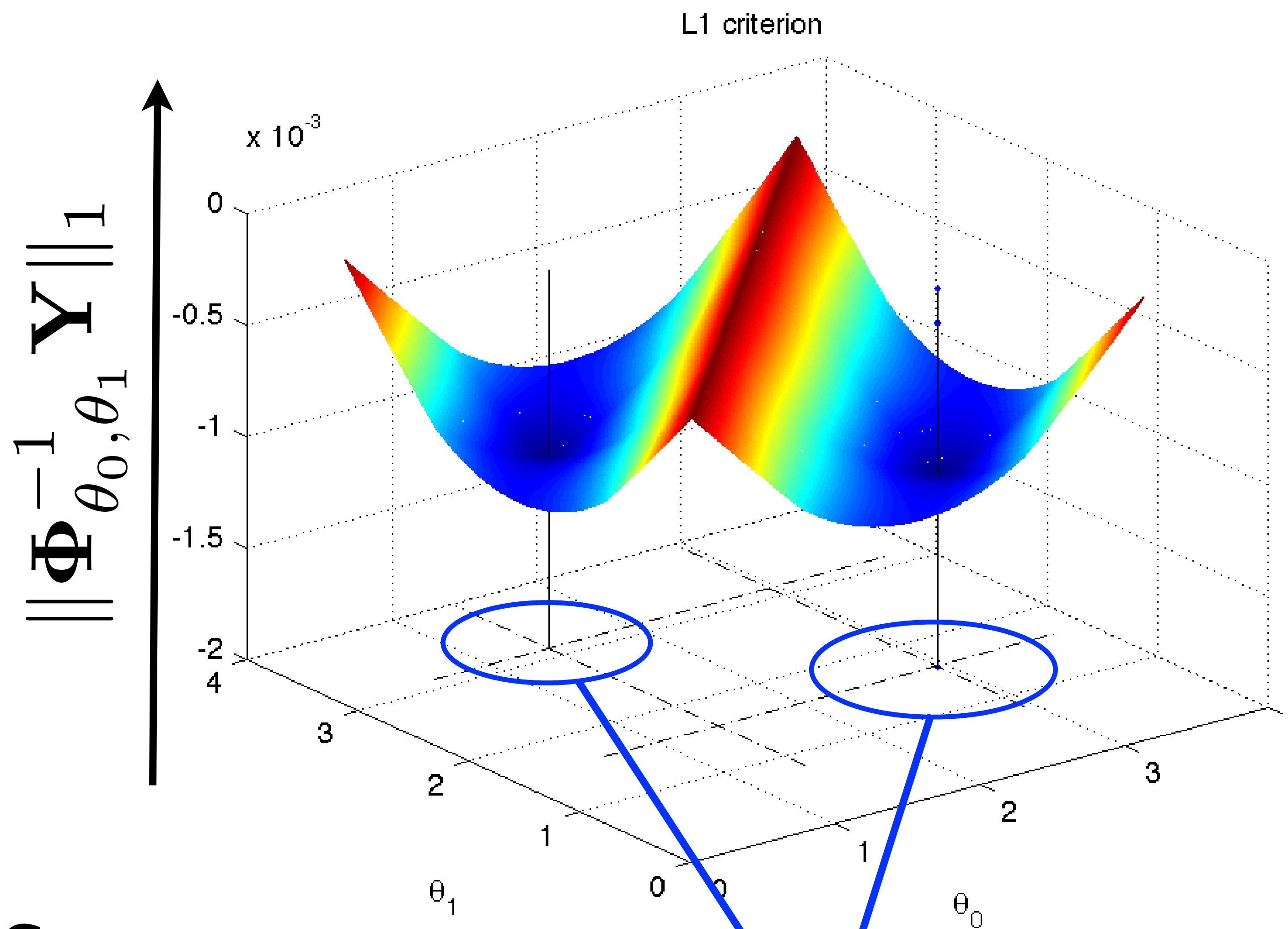




Exemple jouet en 2D

$$\mathbf{Y} = \mathbf{\Phi}_0 X_0$$



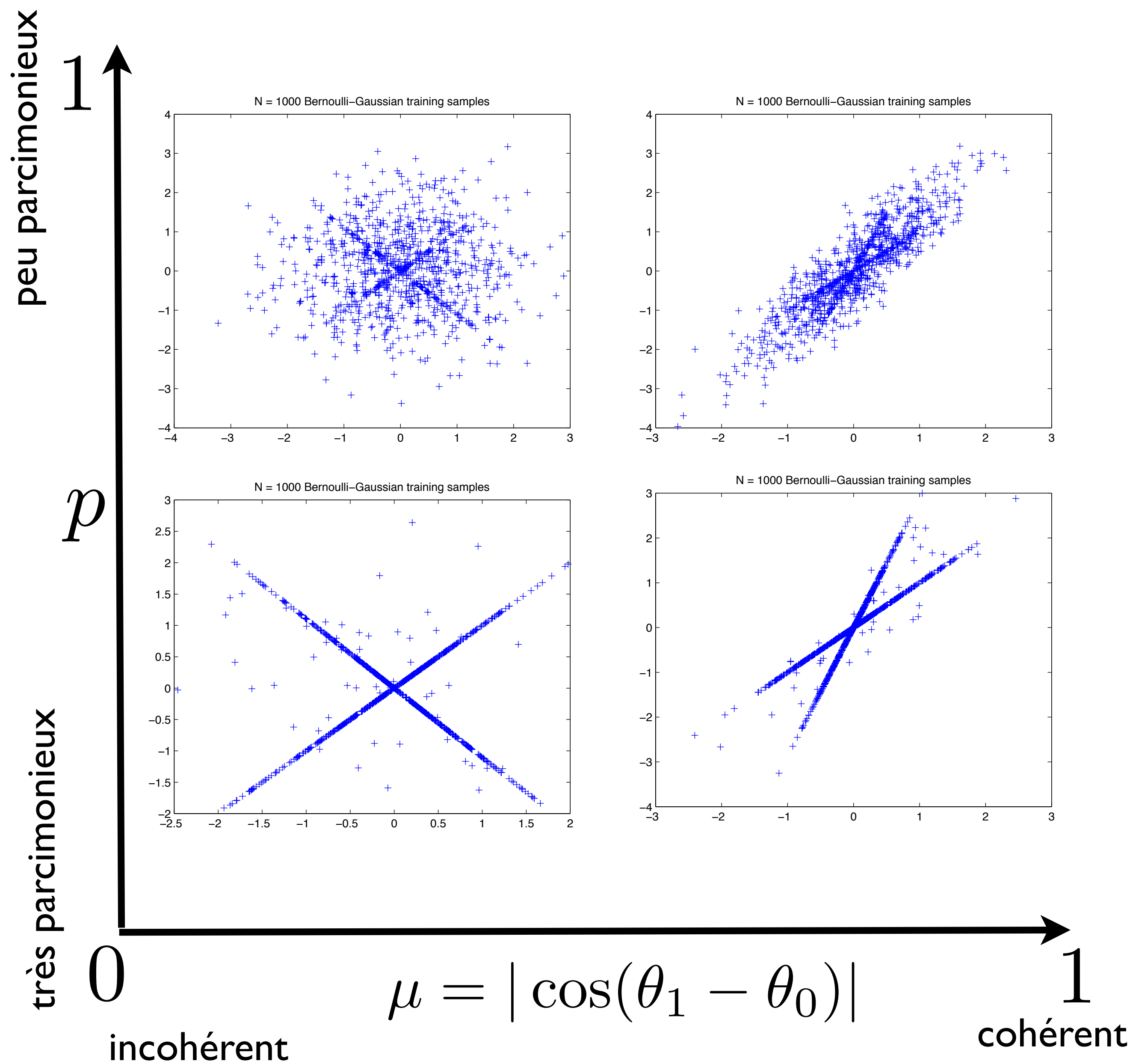


Observations empiriques

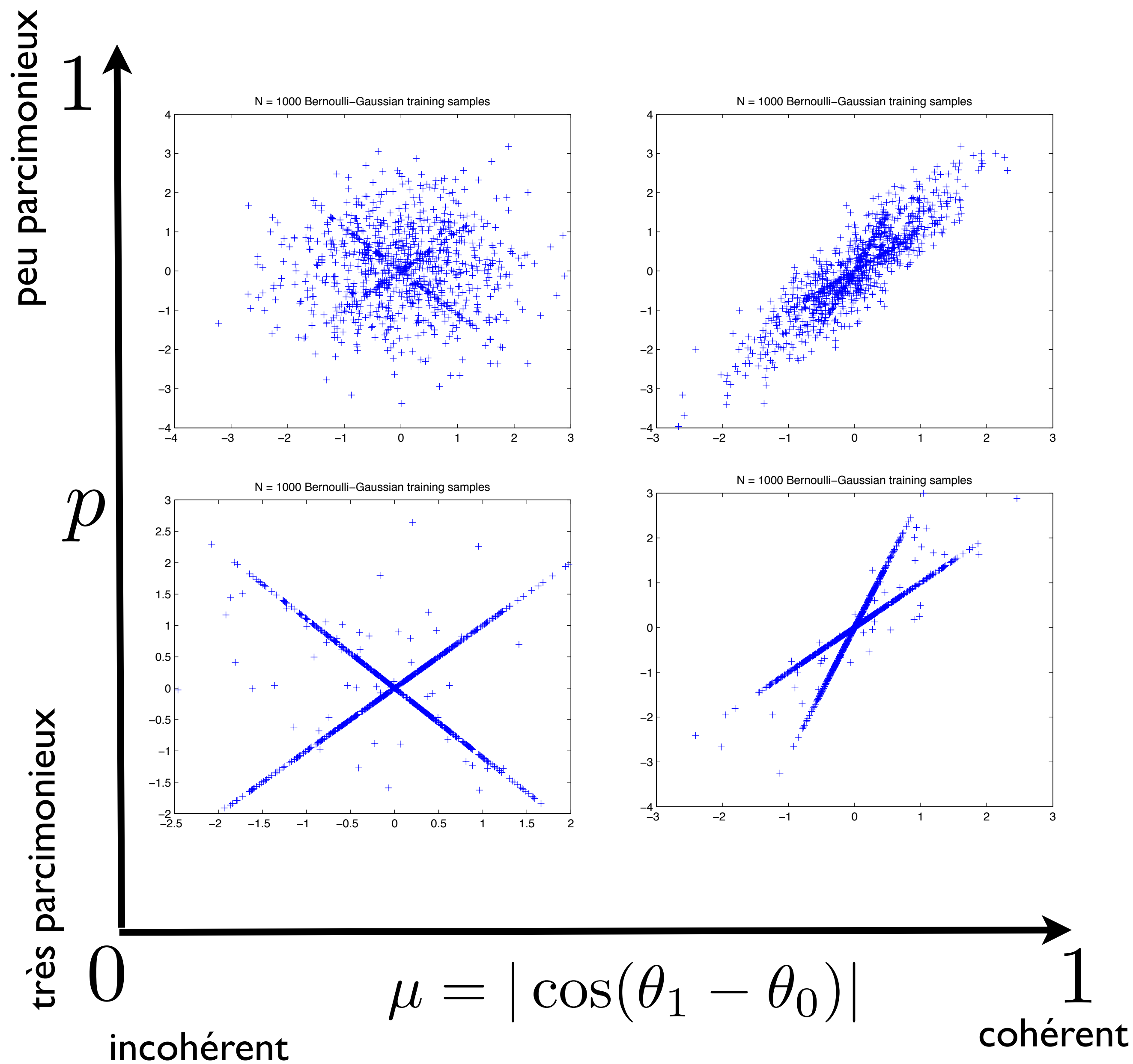
a) Minima globaux correspondent à «vrai» dictionnaire

b) Pas d'autre minimum local.

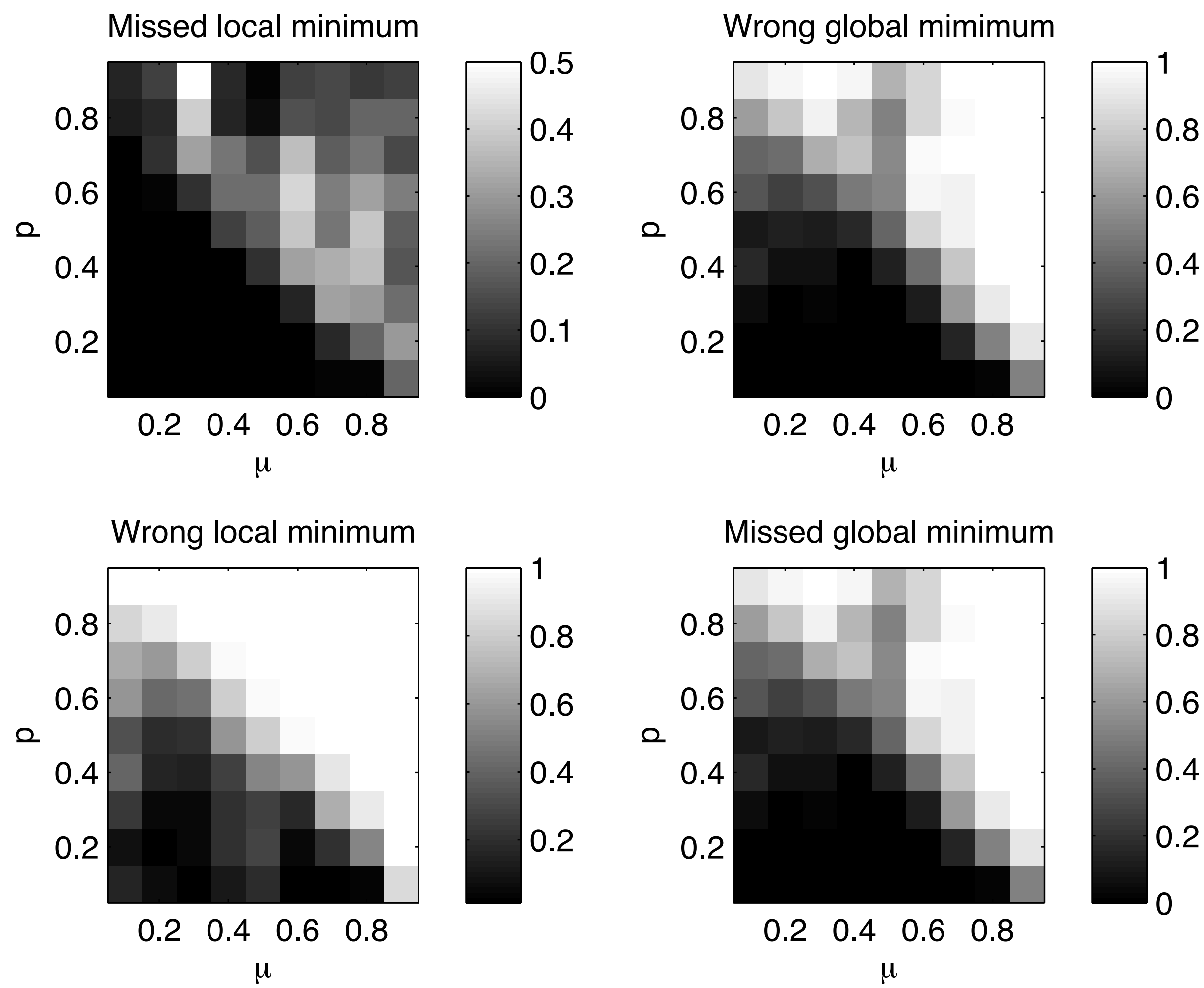
Parcimonie & cohérence



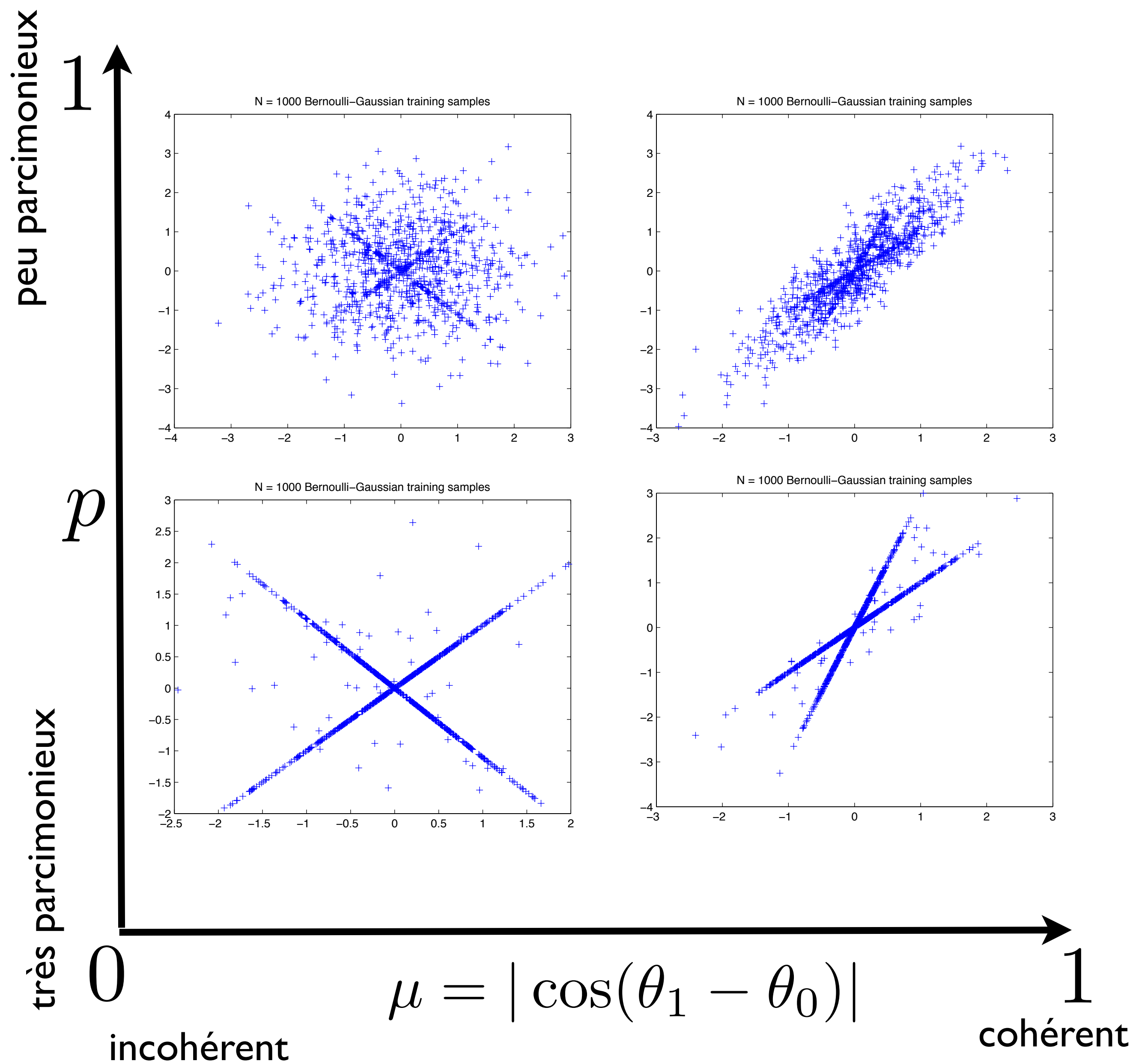
Parcimonie & cohérence



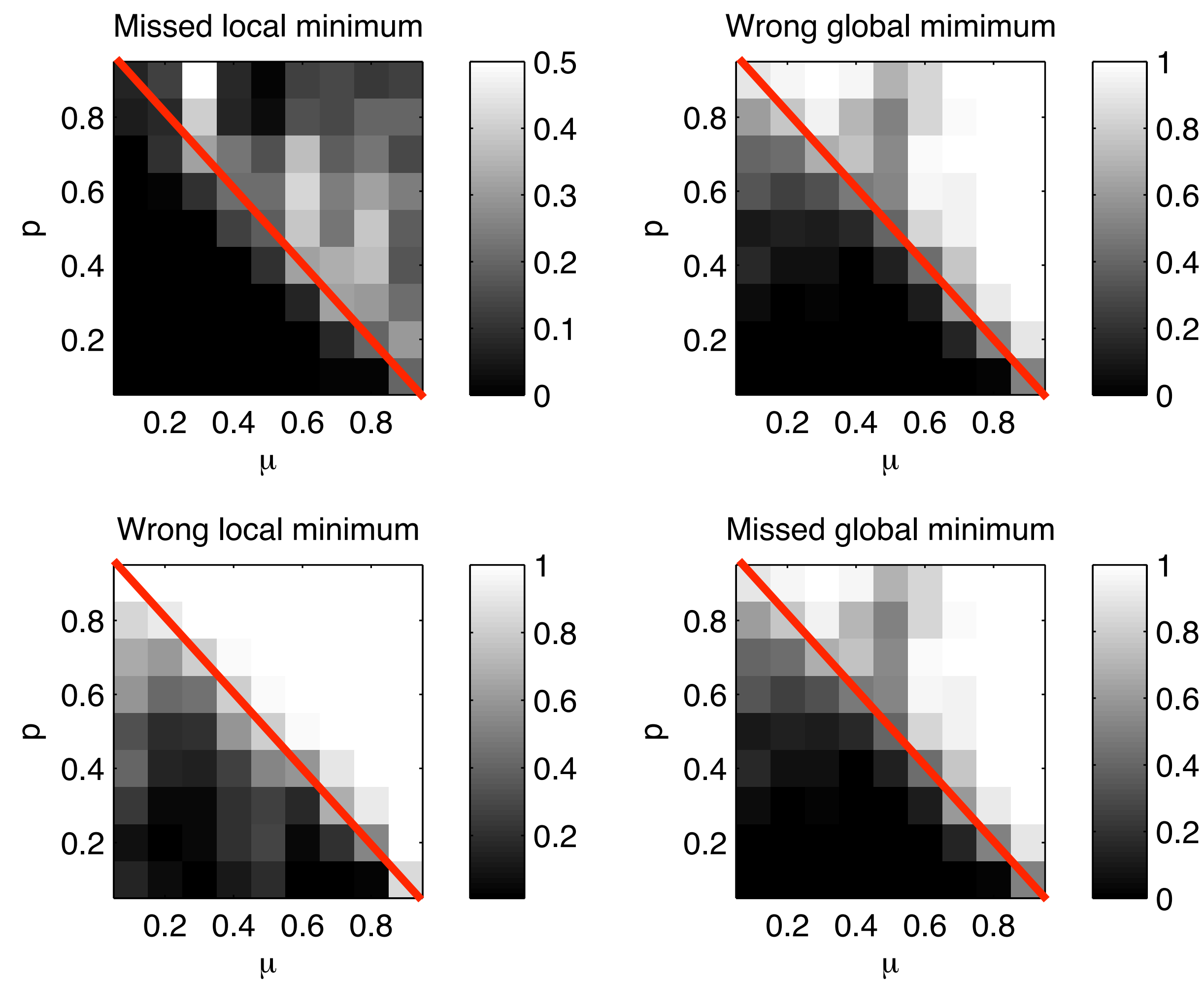
Probabilité empirique de mauvais minima



Parcimonie & cohérence



Probabilité empirique de mauvais minima



Loi empirique: identification de Φ si:

a) $\mu < 1 - p$

b) N assez grand

(assez d'exemples d'entraînement)

Garantes théoriques récentes

- [Gribonval & Schnass 2010]:
  - Conditions d'identifiabilité pour **A inversible**

$$\min_{\mathbf{\Phi},X|\mathbf{Y}=\mathbf{\Phi}X} \|X\|_1$$

  - Contrôle sur le nombre suffisant *N* d'exemples d'apprentissage pour l'**identifiabilité avec forte probabilité** (modèle Bernoulli-Gaussien sur X) en dimension *d*

$$N \geq C d \log d$$

- *[Bach,Jenatton, Gribonval] (work in progress)*
  - *Dictionnaire **A redondant***
  - *Robustesse au **bruit***
  - *Robustesse aux **données aberrantes***

$$\min_{\mathbf{\Phi},X} \frac{1}{2}\|\mathbf{Y}-\mathbf{\Phi}X\|_F^2 + \lambda\|X\|_1$$

La parcimonie dans tous ses états ...

De l'acquisition à la reconnaissance

- **Exploiter la parcimonie**

✓ *Analogique* : objet = donnée	✓ *«Sémantique»*: objet = fonction

$$\mathbf{y} \approx \sum_{\textbf{peu de } k} x_k \varphi_k = \mathbf{\Phi} x$$

- ✓ Modèle de donnée = atomes d'un *dictionnaire*
- ✓ *Compression, représentation, ...*

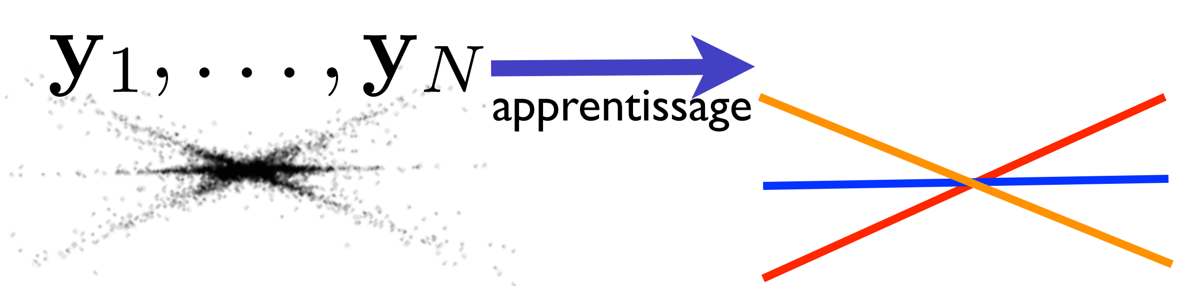
$$f(\mathbf{y}) \approx \sum_{\textbf{peu de } n} \alpha_n K(\mathbf{y}, \mathbf{y}_n)$$

- ✓ Modèle de fonction = *noyau*
- ✓ *Classification, régression, ...*

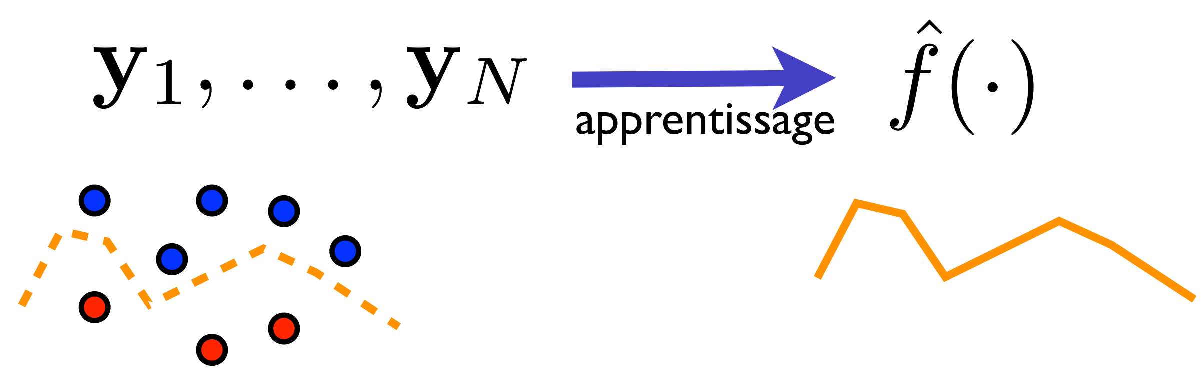
De l'acquisition à la reconnaissance

- **Apprendre**

✓ *Analogique* : objet = donnée ... ✓ *«Sémantique»*: objet = fonction



✓ *Apprentissage de dictionnaire :*
inférer modèle à partir d'exemples



✓ *Machine Learning*:
inférer fonction à partir d'exemples

De l'acquisition à la reconnaissance

- **Apprendre**

✓ *Analogique* : objet = donnée … ✓ *«Sémantique»*: objet = fonction

$$\mathbf{y}_1, \ldots, \mathbf{y}_N \xrightarrow{\text{apprentissage}} \hat{\Phi}$$

$$\mathbf{y}_1, \ldots, \mathbf{y}_N \xrightarrow{\text{apprentissage}} \hat{f}(\cdot)$$

✓ *Apprentissage de dictionnaire :* inférer modèle à partir d'exemples

✓ *Machine Learning*: inférer fonction à partir d'exemples

Et demain ?

- **Cadre unificateur pour traitement des données**
  - Traitement du signal
  - Apprentissage (*Machine Learning*)
- **Prochaines percées attendues**
  - De l'acquisition compressée à l'apprentissage compressé
  - Modèles parcimonieux au delà du concept de dictionnaire
- **Quelques applications d'aujourd'hui et de demain**
  - Inpainting / super-résolution image/vidéo/audio
  - Compression vidéo distribuée
  - Imagerie astronomique (interférométrie)
  - Imagerie médicale (CT & IRM) à faible exposition
  - Enregistrement audio à haute résolution spatiale
  - Capteurs-compresseurs basse consommation
  - Imagerie dynamique haute-résolution du cerveau
  - ...

Questions ?