

# Méthodes parcimonieuses : modèles, algorithmes de calcul, de la théorie mathématique vers l'échantillonnage compressif

Sébastien Bourguignon

Sebastien.Bourguignon@ec-nantes.fr

Webinaire ECND PDL, 31 mai 2021



■ **LS2N = UMR CNRS / Université de Nantes / Centrale Nantes / IMTA**

~ 500 membres (moitié permanents)

- Conception et Conduite de Systèmes
- Robotique, Procédés, Calcul
- Sciences des Données et de la Décision
- Science du Logiciel et des Systèmes Distribués
- **Signaux, Images, Ergonomie et Langues**
  - ▶ Data User Knowledge
  - ▶ Image Perception Interaction
  - ▶ **Signal, Image et Son**
  - ▶ Traitement Automatique du Langage Naturel

■ **SIMS = 6 enseignants-chercheurs (5 CN + 1 UN + 3 CNRS)**

+ 12 doctorants en cours

■ **Problèmes inverses**

- Restauration et reconstruction de signaux et d'images
- Exploitation de la parcimonie
- Gestion de la grande dimension

■ **Apprentissage statistique**

- Modélisation de signaux audionumériques
- Apprentissage profond faiblement supervisé (biomédical)

■ **Multidisciplinarité**

- Co-conception acquisition / algorithme
- Collaborations avec physiciens, spécialistes, médecins, ...

↪ sciences de la vie et santé, audio/acoustique, matériaux, ...

## ■ **CND ultrasonore, imagerie ultrasonore**

- collaboration LAUM (L. Simon, A. Duclos, J-P Groby)
- collaboration industrielle DB-SAS

[Thèses E. Carcreff (2014), N. Laroche (2021)]

## ■ **Imagerie micro-ondes (tomographie de diffraction)**

- collaboration Poly Montréal (Y. Goussard)

[Thèse C. Friedrich (2016)]

## ■ **Inversion de mesures résistives et capacitives → béton**

- collaboration UGE (G. Villain, S. Palma Lopes)

[Thèse M A AlHajj (2020)]



# Méthodes parcimonieuses et acquisition comprimée

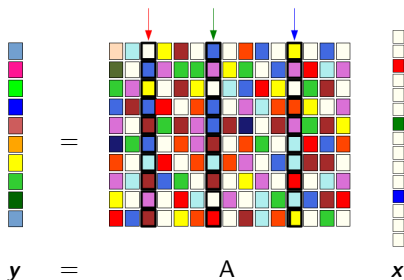
- 1 Définitions, problèmes inverses parcimonieux
- 2 Algorithmes de calcul
- 3 Quelques résultats en CND
- 4 Acquisition comprimée (compressed sensing) : principes et quelques applications

# Méthodes parcimonieuses et acquisition comprimée

- 1 Définitions, problèmes inverses parcimonieux
- 2 Algorithmes de calcul
- 3 Quelques résultats en CND
- 4 Acquisition comprimée (compressed sensing) : principes et quelques applications

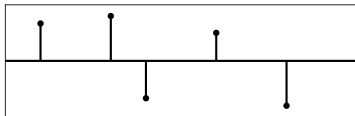
## Hypothèse de compressibilité / parcimonie

Un signal / une image / un ensemble de données peut être représenté (de manière exacte ou approchée) avec un *faible nombre* de coefficients choisis dans un *espace de représentation approprié*.

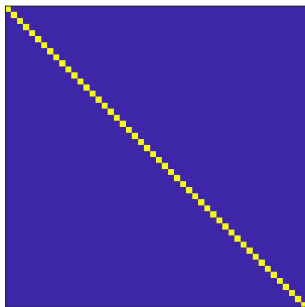


$y = Ax$ , où  $x$  est essentiellement composé de valeurs nulles : **synthèse parcimonieuse**

## Modèles parcimonieux “naturels”

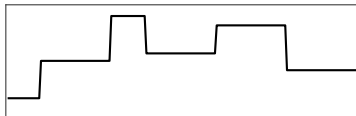


$y$

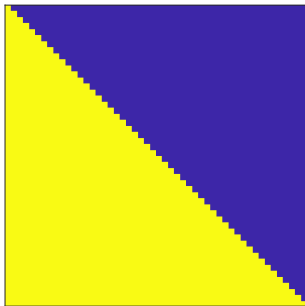


$A$

## Modèles parcimonieux “naturels”

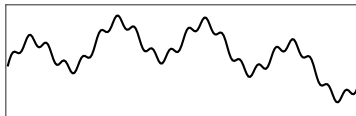


$y$

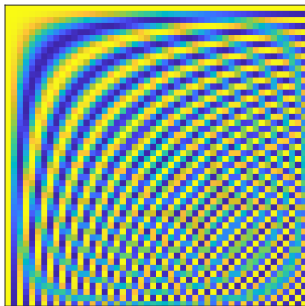


$A$

## Modèles parcimonieux “naturels”



$y$

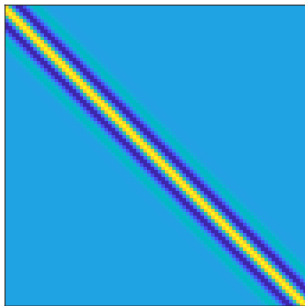


$A$

## Modèles parcimonieux “naturels”

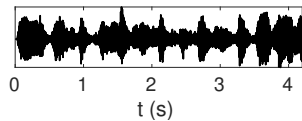


$y$

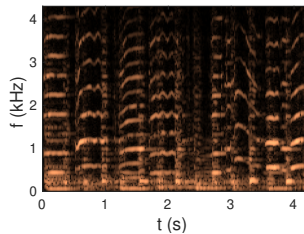


$A$

## Modèles parcimonieux issus de “transformées”



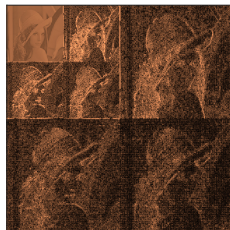
$x = \text{extrait de daphne.wav}$



$|\text{STFT}(x)|$



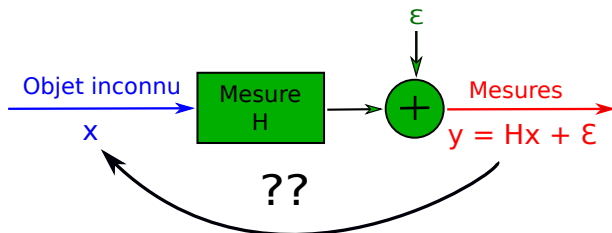
$x$



$|\text{DWT}(x)|$



## Problèmes inverses parcimonieux



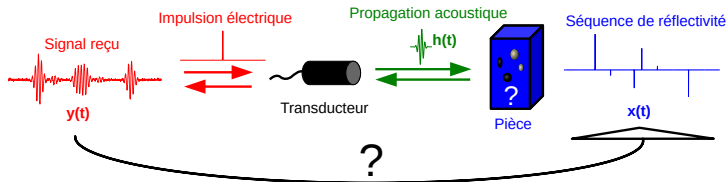
Manque d'information dans  $y \Rightarrow$  **contraintes supplémentaires sur  $x$ ...**

**Parcimonie** :  $x$  peut être approché par un faible nombre d'éléments ... bien choisis.

## Problèmes inverses parcimonieux : exemples

### ■ Déconvolution parcimonieuse

(sismique réflexion [Mendel 83], CND [Zala 92])  
[Thèse E. Carcreff, 2014]



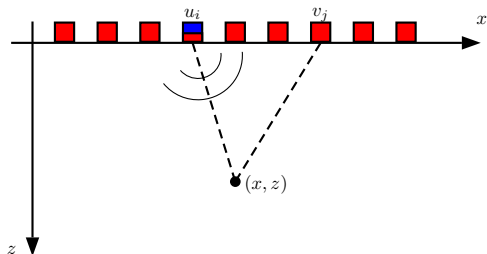
$$y = h * x + \epsilon = Hx + \epsilon$$

- $H$  = matrice de formes d'onde
- $x$  = séquence 1D de réflectivité
- $x \neq 0 \rightsquigarrow$  détection et localisation des interfaces / défauts

## Problèmes inverses parcimonieux : exemples

### ■ Reconstruction / déconvolution d'images

(ex : imagerie ultrasonore pour le CND [Thèse N. Laroche, 2021])



$$\mathbf{H} = \begin{matrix} \left. \begin{array}{c} \text{Receiver} \\ \#1 \end{array} \right\} \begin{array}{c} \mathbf{H}_{1,1}^1 \quad \cdots \quad \mathbf{H}_{1,1}^L \quad \cdots \quad \mathbf{H}_{1,1}^{N_e} \\ \vdots \quad \ddots \quad \vdots \quad \ddots \quad \vdots \\ \mathbf{H}_{N_{e1},1}^1 \quad \cdots \quad \cdots \quad \cdots \quad \mathbf{H}_{N_{e1},1}^{N_e} \\ \vdots \quad \ddots \quad \vdots \quad \ddots \quad \vdots \end{array} \right\} \text{Emitter} \\ \#1 \\ \left. \begin{array}{c} \text{Receiver} \\ \#j \end{array} \right\} \begin{array}{c} \vdots \quad \cdots \quad \cdots \quad \cdots \quad \vdots \\ \mathbf{H}_{1,j}^1 \quad \ddots \quad \mathbf{H}_{1,j}^L \quad \ddots \quad \mathbf{H}_{1,j}^{N_e} \\ \vdots \quad \cdots \quad \cdots \quad \cdots \quad \vdots \\ \mathbf{H}_{N_{e1},j}^1 \quad \cdots \quad \cdots \quad \cdots \quad \mathbf{H}_{N_{e1},j}^{N_e} \\ \vdots \quad \ddots \quad \vdots \quad \ddots \quad \vdots \end{array} \right\} \text{Emitter} \\ \#1 \\ \left. \begin{array}{c} \text{Receiver} \\ \#N_{e1} \end{array} \right\} \begin{array}{c} \mathbf{H}_{1,N_{e1}}^1 \quad \cdots \quad \cdots \quad \cdots \quad \mathbf{H}_{1,N_{e1}}^{N_e} \\ \vdots \quad \ddots \quad \vdots \quad \ddots \quad \vdots \\ \mathbf{H}_{N_{e1},N_{e1}}^1 \quad \cdots \quad \mathbf{H}_{N_{e1},N_{e1}}^L \quad \cdots \quad \mathbf{H}_{N_{e1},N_{e1}}^{N_e} \\ \vdots \quad \ddots \quad \vdots \quad \ddots \quad \vdots \end{array} \right\} \text{Emitter} \\ \#N_{e1} \end{matrix}$$

- $\mathbf{H}$  = matrice de formes d'onde pour tout couple  $(u_i, v_j)$
- $\mathbf{x}$  = carte de réflectivité (2D, 3D)
- $\mathbf{x} \neq 0 \rightsquigarrow$  détection et localisation de réflecteurs ponctuels

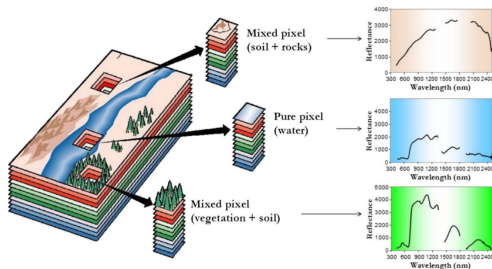
### Autres modèles :

- Gradient de l'image parcimonieux
- Parcimonie en ondelettes:  $\mathbf{W}\mathbf{x} = \mathbf{u}$  parcimonieux

# Problèmes inverses parcimonieux : exemples

## ■ Démélange spectral / séparation de sources

[Iordache *et al.*, 11], [Thèse R. Ben Mhenni]



$$s_1(\lambda) \simeq a_{\text{soil}} s_{\text{soil}}(\lambda) + a_{\text{rock}} s_{\text{rock}}(\lambda)$$

$$s_2(\lambda) \simeq a_{\text{water}} s_{\text{water}}(\lambda)$$

$$s_3(\lambda) \simeq a_{\text{tree}} s_{\text{tree}}(\lambda) + a_{\text{soil}} s_{\text{soil}}(\lambda)$$

$$y = Dx + \epsilon$$

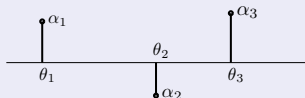
- $S = [s_1^{\text{ref}}, \dots, s_N^{\text{ref}}]$  : bibliothèque de spectres purs
- $a \neq 0 \rightsquigarrow$  détection des composants présents + proportions du mélange

# Problèmes inverses parcimonieux : quelques exemples

## ■ Identification de modèles non linéaires [Tang et al., 13]

### Modèle non linéaire

$$\mathbf{y} \simeq \sum_{k=1}^K \alpha_k \mathcal{M}(\theta_k)$$



→  $\theta, \alpha, K?$

(ex: analyse spectrale)

$$\theta_p^d = p \Delta\theta \rightarrow$$

### Modèle discrétisé linéaire + parcimonie

$$\mathbf{y} \simeq \sum_{p=1}^P x_p \mathcal{M}(\theta_p^d) = \mathbf{M} \mathbf{x}$$



$\mathbf{x}$  parcimonieux :  $\mathcal{S} = \{p | x_p \neq 0\}$

$$\rightsquigarrow \hat{\boldsymbol{\theta}} = \{\theta_p^d\}_{p \in \mathcal{Q}}$$

$$\rightsquigarrow \hat{\boldsymbol{\alpha}} = \{x_p\}_{p \in \mathcal{Q}}$$

$$\rightsquigarrow \hat{K} = \text{Card } \mathcal{Q}$$

Mais aussi...

- **Sélection de variables en statistiques** [Miller, 02]
- **Finance (optimisation de portefeuille)** [Gao & Li, 13]
- Apprentissage de dictionnaires / *sparse coding* [Rubinstein et al. 10]
- ...

# Méthodes parcimonieuses et acquisition comprimée

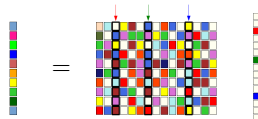
- 1 Définitions, problèmes inverses parcimonieux
- 2 Algorithmes de calcul**
- 3 Quelques résultats en CND
- 4 Acquisition comprimée (compressed sensing) : principes et quelques applications

# Vocabulaire

- $\mathbf{y} \in \mathbb{R}^N$  vecteur de **données**
- $\mathbf{x} \in \mathbb{R}^P$  **coefficients de la représentation** (avec souvent  $P > N$ )
- $A = [\mathbf{a}_1, \dots, \mathbf{a}_P]$ : **dictionnaire** de  $P$  colonnes = **atomes**
- **Représentation** parcimonieuse (exacte) :  $\mathbf{y} = A\mathbf{x}$  avec  $\mathbf{x}$  parcimonieux
- **Approximation** parcimonieuse (avec bruit) :  $\mathbf{y} \simeq A\mathbf{x}$  avec  $\mathbf{x}$  parcimonieux

$$A\mathbf{x} = \sum_{p=1}^P x_p \mathbf{a}_p \text{ avec } \mathbf{x} \text{ parcimonieux : } x_p \neq 0 \text{ pour } p \in \underbrace{\{p_1, \dots, p_K\}}_{\text{support de } \mathbf{x}}$$

$\rightsquigarrow$  sélection de colonnes actives



- “Norme”  $\ell_0$  :  $\|\mathbf{x}\|_0 = \text{Card}\{p | x_p \neq 0\}$



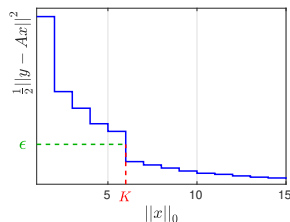
## Formulation de problèmes d'optimisation

Objectif : on cherche le meilleur  $x$  parcimonieux tel que  $y \simeq Ax$

$$\rightsquigarrow \min_x \frac{1}{2} \|y - Ax\|^2 \quad \text{s.c.} \quad \|x\|_0 \leq K$$

$$\rightsquigarrow \min_x \|x\|_0 \quad \text{s.c.} \quad \frac{1}{2} \|y - Ax\|^2 \leq \epsilon$$

$$\rightsquigarrow \min_x \frac{1}{2} \|y - Ax\|^2 + \mu \|x\|_0$$



- Complexité *combinatoire* = sélection du support de  $x$  :  $S = \{p | x_p \neq 0\}$

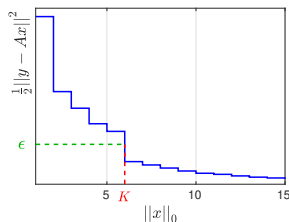
## Formulation de problèmes d'optimisation

Objectif : on cherche le meilleur  $x$  parcimonieux tel que  $y \simeq Ax$

$$\rightsquigarrow \min_x \frac{1}{2} \|y - Ax\|^2 \text{ s.c. } \|x\|_0 \leq K$$

$$\rightsquigarrow \min_x \|x\|_0 \text{ s.c. } \frac{1}{2} \|y - Ax\|^2 \leq \epsilon$$

$$\rightsquigarrow \min_x \frac{1}{2} \|y - Ax\|^2 + \mu \|x\|_0$$



- Complexité *combinatoire* = sélection du **support** de  $x$  :  $S = \{p | x_p \neq 0\}$

- **Optimisation exacte accessible en petite dimension**

$\rightsquigarrow$  algorithmes Branch-and-Bound [Bertsimas *et al.* 16, B. *et al.* 16, ...  
thèses R. Ben Mhenni, 2020, G. Samain en cours]

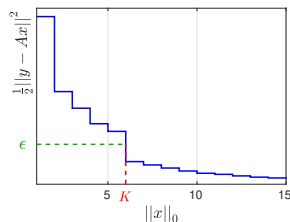
## Formulation de problèmes d'optimisation

Objectif : on cherche le **meilleur**  $x$  **parcimonieux** tel que  $y \simeq Ax$

$$\rightsquigarrow \min_x \frac{1}{2} \|y - Ax\|^2 \quad \text{s.c.} \quad \|x\|_0 \leq K$$

$$\rightsquigarrow \min_x \|x\|_0 \quad \text{s.c.} \quad \frac{1}{2} \|y - Ax\|^2 \leq \epsilon$$

$$\rightsquigarrow \min_x \frac{1}{2} \|y - Ax\|^2 + \mu \|x\|_0$$



- Complexité *combinatoire* = sélection du **support** de  $x$  :  $S = \{p | x_p \neq 0\}$
- **Optimisation exacte accessible en petite dimension**
  - $\rightsquigarrow$  algorithmes Branch-and-Bound [Bertsimas *et al.* 16, B. *et al.* 16, ...  
thèses R. Ben Mhenni, 2020, G. Samain en cours]
- Sinon :

$\rightsquigarrow$  **optimisation locale** : heuristiques d'exploration combinatoire (partielle)

$\rightsquigarrow$  **relaxations** : résolution d'un problème plus simple

## Optimisation parcimonieuse : heuristiques

$$\min_x \frac{1}{2} \|\mathbf{y} - \mathbf{Ax}\|^2 \quad \text{s.c.} \quad \|\mathbf{x}\|_0 \leq K, \quad \min_x \|\mathbf{x}\|_0 \quad \text{s.c.} \quad \frac{1}{2} \|\mathbf{y} - \mathbf{Ax}\|^2 \leq \epsilon$$

### ■ Construction itérative d'une solution parcimonieuse (algorithmes "gloutons")

Forward selection [Efroymsen, 60]. . .

1. Support  $S^{(0)} = \emptyset$ ,  $\mathbf{x}^{(0)} = 0$ , résidu  $\mathbf{r}^{(0)} = \mathbf{y}$
2. Itération  $t$  : choix d'une composante  $k^{(t)} \notin S^{(t-1)} \rightarrow S^{(t)} = S^{(t-1)} \cup \{k^{(t)}\}$
3. Calcul des amplitudes  $\mathbf{x}_{S^{(t)}}$  et mise à jour du résidu  $\mathbf{r}^{(t)} = \mathbf{y} - \mathbf{A}_{S^{(t)}} \mathbf{x}_{S^{(t)}}$
4.  $t \leftarrow t + 1$  et retour en 2. jusqu'à  $\|\mathbf{r}^{(t)}\| < \epsilon$  ou  $\text{Card } S^{(t)} = K$



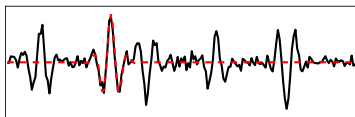
## Optimisation parcimonieuse : heuristiques

$$\min_x \frac{1}{2} \|y - Ax\|^2 \text{ s.c. } \|x\|_0 \leq K, \quad \min_x \|x\|_0 \text{ s.c. } \frac{1}{2} \|y - Ax\|^2 \leq \epsilon$$

### ■ Construction itérative d'une solution parcimonieuse (algorithmes "gloutons")

Forward selection [Efroymsen, 60]. . .

1. Support  $S^{(0)} = \emptyset$ ,  $x^{(0)} = 0$ , résidu  $r^{(0)} = y$
2. Itération  $t$  : choix d'une composante  $k^{(t)} \notin S^{(t-1)} \rightarrow S^{(t)} = S^{(t-1)} \cup \{k^{(t)}\}$
3. Calcul des amplitudes  $x_{S^{(t)}}$  et mise à jour du résidu  $r^{(t)} = y - A_{S^{(t)}} x_{S^{(t)}}$
4.  $t \leftarrow t + 1$  et retour en 2. jusqu'à  $\|r^{(t)}\| < \epsilon$  ou  $\text{Card } S^{(t)} = K$



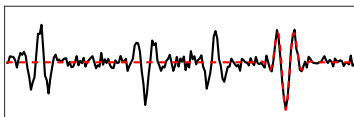
## Optimisation parcimonieuse : heuristiques

$$\min_x \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2 \quad \text{s.c.} \quad \|\mathbf{x}\|_0 \leq K, \quad \min_x \|\mathbf{x}\|_0 \quad \text{s.c.} \quad \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2 \leq \epsilon$$

### ■ Construction itérative d'une solution parcimonieuse (algorithmes "gloutons")

Forward selection [Efroymsen, 60]. . .

1. Support  $S^{(0)} = \emptyset$ ,  $\mathbf{x}^{(0)} = 0$ , résidu  $\mathbf{r}^{(0)} = \mathbf{y}$
2. Itération  $t$  : choix d'une composante  $k^{(t)} \notin S^{(t-1)} \rightarrow S^{(t)} = S^{(t-1)} \cup \{k^{(t)}\}$
3. Calcul des amplitudes  $\mathbf{x}_{S^{(t)}}$  et mise à jour du résidu  $\mathbf{r}^{(t)} = \mathbf{y} - \mathbf{A}_{S^{(t)}} \mathbf{x}_{S^{(t)}}$
4.  $t \leftarrow t + 1$  et retour en 2. jusqu'à  $\|\mathbf{r}^{(t)}\| < \epsilon$  ou  $\text{Card } S^{(t)} = K$



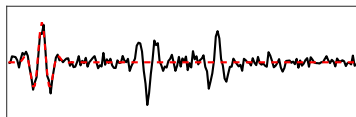
## Optimisation parcimonieuse : heuristiques

$$\min_x \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2 \quad \text{s.c.} \quad \|\mathbf{x}\|_0 \leq K, \quad \min_x \|\mathbf{x}\|_0 \quad \text{s.c.} \quad \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2 \leq \epsilon$$

### ■ Construction itérative d'une solution parcimonieuse (algorithmes "gloutons")

Forward selection [Efroymsen, 60]. . .

1. Support  $S^{(0)} = \emptyset$ ,  $\mathbf{x}^{(0)} = 0$ , résidu  $\mathbf{r}^{(0)} = \mathbf{y}$
2. Itération  $t$  : choix d'une composante  $k^{(t)} \notin S^{(t-1)} \rightarrow S^{(t)} = S^{(t-1)} \cup \{k^{(t)}\}$
3. Calcul des amplitudes  $\mathbf{x}_{S^{(t)}}$  et mise à jour du résidu  $\mathbf{r}^{(t)} = \mathbf{y} - \mathbf{A}_{S^{(t)}} \mathbf{x}_{S^{(t)}}$
4.  $t \leftarrow t + 1$  et retour en 2. jusqu'à  $\|\mathbf{r}^{(t)}\| < \epsilon$  ou  $\text{Card } S^{(t)} = K$



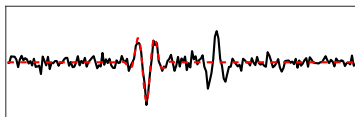
## Optimisation parcimonieuse : heuristiques

$$\min_x \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2 \quad \text{s.c.} \quad \|\mathbf{x}\|_0 \leq K, \quad \min_x \|\mathbf{x}\|_0 \quad \text{s.c.} \quad \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2 \leq \epsilon$$

### ■ Construction itérative d'une solution parcimonieuse (algorithmes "gloutons")

Forward selection [Efroymsen, 60]. . .

1. Support  $S^{(0)} = \emptyset$ ,  $\mathbf{x}^{(0)} = 0$ , résidu  $\mathbf{r}^{(0)} = \mathbf{y}$
2. Itération  $t$  : choix d'une composante  $k^{(t)} \notin S^{(t-1)} \rightarrow S^{(t)} = S^{(t-1)} \cup \{k^{(t)}\}$
3. Calcul des amplitudes  $\mathbf{x}_{S^{(t)}}$  et mise à jour du résidu  $\mathbf{r}^{(t)} = \mathbf{y} - \mathbf{A}_{S^{(t)}} \mathbf{x}_{S^{(t)}}$
4.  $t \leftarrow t + 1$  et retour en 2. jusqu'à  $\|\mathbf{r}^{(t)}\| < \epsilon$  ou  $\text{Card } S^{(t)} = K$





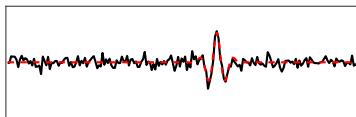
## Optimisation parcimonieuse : heuristiques

$$\min_x \frac{1}{2} \|y - Ax\|^2 \text{ s.c. } \|x\|_0 \leq K, \quad \min_x \|x\|_0 \text{ s.c. } \frac{1}{2} \|y - Ax\|^2 \leq \epsilon$$

### ■ Construction itérative d'une solution parcimonieuse (algorithmes "gloutons")

Forward selection [Efroymsen, 60]. . .

1. Support  $S^{(0)} = \emptyset$ ,  $x^{(0)} = 0$ , résidu  $r^{(0)} = y$
2. Itération  $t$  : choix d'une composante  $k^{(t)} \notin S^{(t-1)} \rightarrow S^{(t)} = S^{(t-1)} \cup \{k^{(t)}\}$
3. Calcul des amplitudes  $x_{S^{(t)}}$  et mise à jour du résidu  $r^{(t)} = y - A_{S^{(t)}} x_{S^{(t)}}$
4.  $t \leftarrow t + 1$  et retour en 2. jusqu'à  $\|r^{(t)}\| < \epsilon$  ou  $\text{Card } S^{(t)} = K$



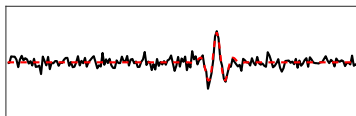
## Optimisation parcimonieuse : heuristiques

$$\min_x \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2 \text{ s.c. } \|\mathbf{x}\|_0 \leq K, \quad \min_x \|\mathbf{x}\|_0 \text{ s.c. } \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2 \leq \epsilon$$

### ■ Construction itérative d'une solution parcimonieuse (algorithmes "gloutons")

Forward selection [Efroymsen, 60]...

1. Support  $S^{(0)} = \emptyset$ ,  $\mathbf{x}^{(0)} = 0$ , résidu  $\mathbf{r}^{(0)} = \mathbf{y}$
2. Itération  $t$  : choix d'une composante  $k^{(t)} \notin S^{(t-1)} \rightarrow S^{(t)} = S^{(t-1)} \cup \{k^{(t)}\}$
3. Calcul des amplitudes  $\mathbf{x}_{S^{(t)}}$  et mise à jour du résidu  $\mathbf{r}^{(t)} = \mathbf{y} - \mathbf{A}_{S^{(t)}} \mathbf{x}_{S^{(t)}}$
4.  $t \leftarrow t + 1$  et retour en 2. jusqu'à  $\|\mathbf{r}^{(t)}\| < \epsilon$  ou  $\text{Card } S^{(t)} = K$



### ■ Exemple : Orthogonal Matching Pursuit [Pati et al., 93]

2. Choix de la composante la plus adaptée au résidu :  $k^{(t)} = \arg \max_j |\mathbf{a}_j^T \mathbf{r}^{(t-1)}|$
3. Optimisation des amplitudes :  $\mathbf{x}_{S^{(t)}} = \arg \min_{\mathbf{x}_S} \|\mathbf{y} - \mathbf{A}_{S^{(t)}} \mathbf{x}_{S^{(t)}}\|^2$

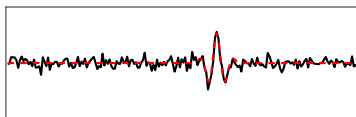
## Optimisation parcimonieuse : heuristiques

$$\min_x \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2 \text{ s.c. } \|\mathbf{x}\|_0 \leq K, \quad \min_x \|\mathbf{x}\|_0 \text{ s.c. } \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2 \leq \epsilon$$

### ■ Construction itérative d'une solution parcimonieuse (algorithmes "gloutons")

Forward selection [Efroymsen, 60]...

1. Support  $S^{(0)} = \emptyset$ ,  $\mathbf{x}^{(0)} = 0$ , résidu  $\mathbf{r}^{(0)} = \mathbf{y}$
2. Itération  $t$  : choix d'une composante  $k^{(t)} \notin S^{(t-1)} \rightarrow S^{(t)} = S^{(t-1)} \cup \{k^{(t)}\}$
3. Calcul des amplitudes  $\mathbf{x}_{S^{(t)}}$  et mise à jour du résidu  $\mathbf{r}^{(t)} = \mathbf{y} - \mathbf{A}_{S^{(t)}}\mathbf{x}_{S^{(t)}}$
4.  $t \leftarrow t + 1$  et retour en 2. jusqu'à  $\|\mathbf{r}^{(t)}\| < \epsilon$  ou  $\text{Card } S^{(t)} = K$

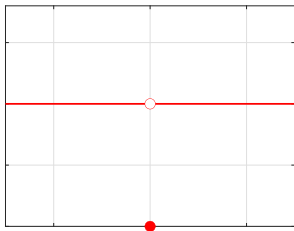


### ■ Nombreuses heuristiques [OLS, SBR, COSAMP, ...]

- Choix sous le compromis qualité / complexité calculatoire
- Interférences entre composantes  $\leadsto$  chances faibles de retrouver l'optimum global

## Optimisation parcimonieuse : relaxations continues

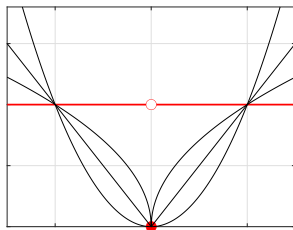
$$\min_{\mathbf{x}} \frac{1}{2} \|\mathbf{y} - \mathbf{Ax}\|^2 + \lambda \underbrace{\|\mathbf{x}\|_0}_{\sum_p 1_{\{x_p \neq 0\}}}$$



## Optimisation parcimonieuse : relaxations continues

$$\min_{\mathbf{x}} \frac{1}{2} \|\mathbf{y} - \mathbf{Ax}\|^2 + \lambda \underbrace{\|\mathbf{x}\|_0}_{\sum_p 1_{\{x_p \neq 0\}}} \quad \rightsquigarrow \quad \min_{\mathbf{x}} \frac{1}{2} \|\mathbf{y} - \mathbf{Ax}\|^2 + \lambda \underbrace{\phi(\mathbf{x})}_{\sum_p \varphi(x_p)}$$

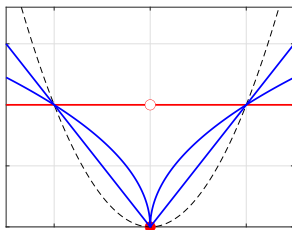
avec  $\varphi$  continue, paire, croissante sur  $\mathbb{R}^+$



## Optimisation parcimonieuse : relaxations continues

$$\min_x \frac{1}{2} \|y - Ax\|^2 + \lambda \underbrace{\|x\|_0}_{\sum_p 1_{\{x_p \neq 0\}}} \quad \rightsquigarrow \quad \min_x \frac{1}{2} \|y - Ax\|^2 + \lambda \underbrace{\phi(x)}_{\sum_p \varphi(x_p)}$$

avec  $\varphi$  continue, paire, croissante sur  $\mathbb{R}^+$



La solution est parcimonieuse si et seulement si  $\varphi$  non différentiable en 0

[Moulin & Liu 98]

## Optimisation continue non différentiable

$$\min_{\mathbf{x}} \frac{1}{2} \|\mathbf{y} - \mathbf{Ax}\|^2 + \lambda \sum_p \varphi(x_p), \varphi \text{ non différentiable en } 0$$

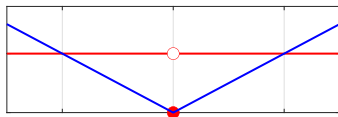
**LE sujet chaud depuis 10 ans / algorithmes parcimonieux! [Bach et al. 12]**

## Optimisation continue non différentiable

$$\min_{\mathbf{x}} \frac{1}{2} \|\mathbf{y} - \mathbf{Ax}\|^2 + \lambda \sum_p \varphi(x_p), \varphi \text{ non différentiable en } 0$$

LE sujet chaud depuis 10 ans / algorithmes parcimonieux! [Bach et al. 12]

■ Norme  $\ell_1$  ( $\varphi(x_p) = |x_p|$ ) : critère convexe, géométrie simple



- Beaucoup de stratégies, qui fournissent la même solution
- Efficacité calculatoire dépendante du problème (taille, parcimonie, A)

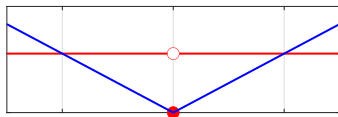


## Optimisation continue non différentiable

$$\min_x \frac{1}{2} \|y - Ax\|^2 + \lambda \sum_p \varphi(x_p), \varphi \text{ non différentiable en } 0$$

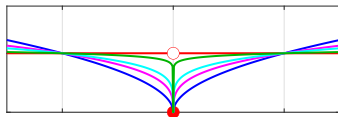
LE sujet chaud depuis 10 ans / algorithmes parcimonieux! [Bach et al. 12]

■ Norme  $\ell_1$  ( $\varphi(x_p) = |x_p|$ ) : critère convexe, géométrie simple



- Beaucoup de stratégies, qui fournissent la même solution
- Efficacité calculatoire dépendante du problème (taille, parcimonie, A)

■ Fonctions non convexes

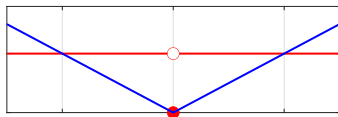


## Optimisation continue non différentiable

$$\min_x \frac{1}{2} \|y - Ax\|^2 + \lambda \sum_p \varphi(x_p), \varphi \text{ non différentiable en } 0$$

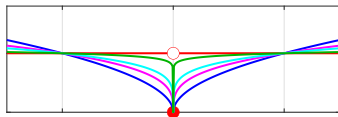
LE sujet chaud depuis 10 ans / algorithmes parcimonieux! [Bach et al. 12]

■ Norme  $\ell_1$  ( $\varphi(x_p) = |x_p|$ ) : critère convexe, géométrie simple



- Beaucoup de stratégies, qui fournissent la même solution
- Efficacité calculatoire dépendante du problème (taille, parcimonie, A)

■ Fonctions non convexes

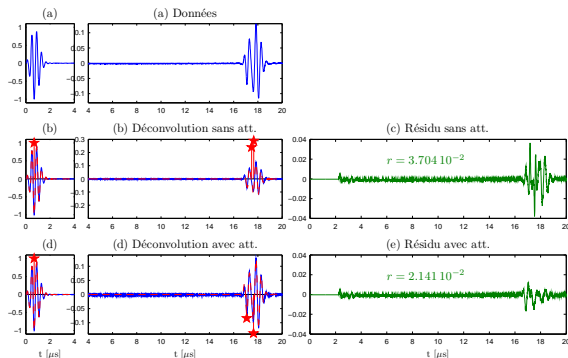
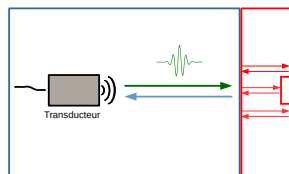


- Algorithmes d'optimisation locale : pas de garantie d'optimalité globale... algorithmes proximaux, ADMM, IRL1, ...
- ... mais fonction de coût plus proche de  $\ell_0$
- En pratique, résultats souvent meilleurs que  $\ell_1$ ...

# Méthodes parcimonieuses et acquisition comprimée

- 1 Définitions, problèmes inverses parcimonieux
- 2 Algorithmes de calcul
- 3 Quelques résultats en CND**
- 4 Acquisition comprimée (compressed sensing) : principes et quelques applications

## ■ Déconvolution d'A-scans : détection de FBH



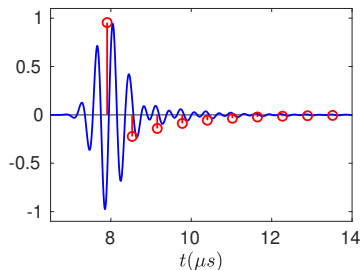
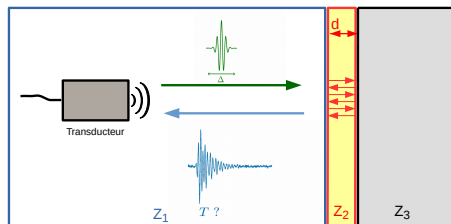
Haut : données  $\mathbf{y}$ .

Centre :  $\min \frac{1}{2} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2$  s.c.  $\|\mathbf{x}\|_0 \leq 3$ .

Bas :  $\min \frac{1}{2} \|\mathbf{y} - \mathbf{H}\mathbf{H}_a\mathbf{x}\|^2$  s.c.  $\|\mathbf{x}\|_0 \leq 3$ .

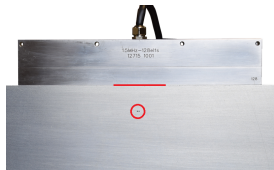
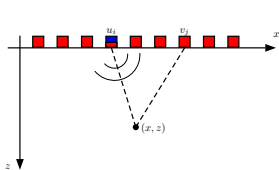
[Cacreff *et al.*, 2014, Thèse E. Carcreff, 2014]

## ■ Déconvolution d'A scans : mesure d'épaisseur



$$\hat{d} = 1.99 \text{ mm} \pm 0.0002 \text{ mm}$$

## ■ Imagerie ultrasonore : inversion de données *Full Matrix Capture*



$$\underbrace{\mathbf{y}}_{\text{Ensemble de A-scans } y_{i,j}} = \mathbf{A} \underbrace{\mathbf{o}}_{\text{image de réflectivité}} + \boldsymbol{\epsilon}$$
$$\min_x \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{o}\|^2 + \lambda_1 \|\mathbf{o}\|_1 + \lambda_2 \|\mathbf{D}\mathbf{o}\|_1 \quad (*)$$

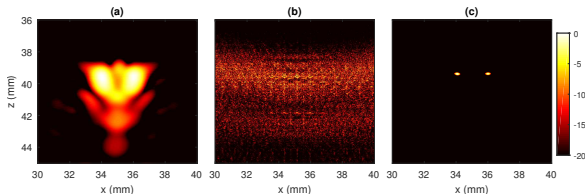
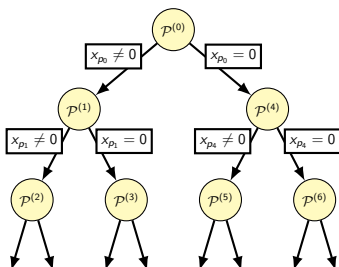


Image TFM, inversion non régularisée et solution de (\*)

[Laroche *et al.*, IEEE TUFFC, 2020]

## Optimisation exacte : *Branch-and-Bound* [Land & Doig, 60]

- **Arbre de décision** pour les variables binaires
- À chaque nœud, on décide si une variable est nulle / non-nulle, les autres sont *relâchées*  $\leadsto$  **borne inférieure** sur la valeur optimale
- Si cette borne est supérieure à la meilleure solution connue, la branche est élaguée



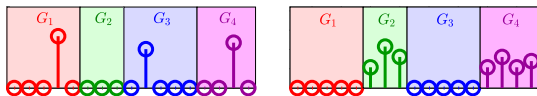
- ▷ Sur quelle variable brancher?
- ▷ Quel côté explorer en premier?
- ▷ Quel nœud explorer en premier?  
 $\leadsto$  heuristiques gloutonnes
- ▷ Calcul des relaxations?  
 $\leadsto$  problèmes en norme  $\ell_1$
- ▷ (Heuristiques, coupes, ...)

Résolution de problèmes inverses difficiles jusqu'à  $P \sim 1000$ ,  $K \sim 10 - 20$

[Thèses R. Ben Mhenni (2020), G. Samain]

## Généralisations : parcimonie structurée

### ■ Prise en compte de contraintes liant les différentes variables



- Parcimonie dans des groupes :  $\|\mathbf{x}_{G_i}\| \leq K_i$
- Sélection parcimonieuse de groupes :  $x_i = 0 \Leftrightarrow x_j \neq 0, j \in G_i$   
ex : *Group LASSO* : sélection parcimonieuse de groupes de variables

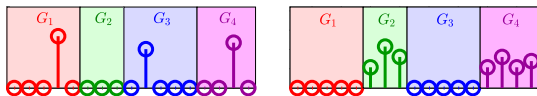
$$\min_{\mathbf{x}} \frac{1}{2} \|\mathbf{y} - \mathbf{Ax}\|^2 + \lambda \sum_j \|\mathbf{x}\|_{G_j}$$

...

↪ Généralisations des algorithmes gloutons et relaxations (normes mixtes) ...



### ■ Prise en compte de contraintes liant les différentes variables



- Parcimonie dans des groupes :  $\|\mathbf{x}_{G_i}\| \leq K_i$
- Sélection parcimonieuse de groupes :  $x_i = 0 \Leftrightarrow x_j \neq 0, j \in G_i$   
ex : *Group LASSO* : sélection parcimonieuse de groupes de variables

$$\min_{\mathbf{x}} \frac{1}{2} \|\mathbf{y} - \mathbf{Ax}\|^2 + \lambda \sum_j \|\mathbf{x}\|_{G_j}$$

...

- ~ Généralisations des algorithmes gloutons et relaxations (normes mixtes) ...  
et des approches exactes! [Thèse G. Samain]

## Généralisations : apprentissage de dictionnaires

Pour un ensemble de données  $\mathbf{y}_1, \dots, \mathbf{y}_M$ , on cherche un dictionnaire commun de représentation  $A$  tel que

$$\mathbf{y}_m = A\mathbf{x}_m, \mathbf{x}_m \text{ parcimonieux}$$

- Problème conjoint en  $(A, \mathbf{x}_1, \dots, \mathbf{x}_M)$
- Pas de solution simple, beaucoup de méthodes approchées

Applications  $\uparrow\uparrow$  en statistiques / machine learning (*sparse coding*) ...

# Méthodes parcimonieuses et acquisition comprimée

- 1 Définitions, problèmes inverses parcimonieux
- 2 Algorithmes de calcul
- 3 Quelques résultats en CND
- 4 Acquisition comprimée (compressed sensing) : principes et quelques applications

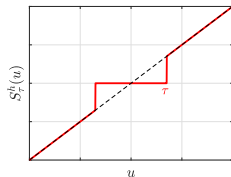
## Parenthèse : conditions d'optimalité des solutions en norme $\ell_1$

■ Supposons  $A$  orthogonale ( $A$  carrée,  $A^T A = I$ ).

$$\min_{\mathbf{x}} \frac{1}{2} \|\mathbf{y} - A\mathbf{x}\|^2 + \lambda \|\mathbf{x}\|_0 \Leftrightarrow \min_{\mathbf{x}} \frac{1}{2} \left\| A^T \mathbf{y} - \mathbf{x} \right\|^2 + \lambda \|\mathbf{x}\|_0$$

$$\Leftrightarrow \forall p, \min_{x_p} \frac{1}{2} (\mathbf{a}_p^T \mathbf{y} - x_p)^2 + \lambda \mathbf{1}_{\{x_p \neq 0\}}$$

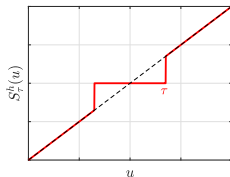
$$\Leftrightarrow \forall p, \hat{x}_p = \mathcal{S}_\tau^h(\mathbf{a}_p^T \mathbf{y}) \text{ (hard thresholding)}$$



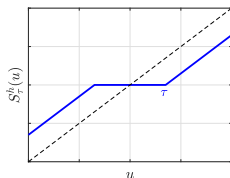
## Parenthèse : conditions d'optimalité des solutions en norme $\ell_1$

■ Supposons  $A$  orthogonale ( $A$  carrée,  $A^T A = I$ ).

$$\begin{aligned}\min_{\mathbf{x}} \frac{1}{2} \|\mathbf{y} - A\mathbf{x}\|^2 + \lambda \|\mathbf{x}\|_0 &\Leftrightarrow \min_{\mathbf{x}} \frac{1}{2} \left\| A^T \mathbf{y} - \mathbf{x} \right\|^2 + \lambda \|\mathbf{x}\|_0 \\ &\Leftrightarrow \forall p, \min_{x_p} \frac{1}{2} (\mathbf{a}_p^T \mathbf{y} - x_p)^2 + \lambda \mathbf{1}_{\{x_p \neq 0\}} \\ &\Leftrightarrow \forall p, \hat{x}_p = \mathcal{S}_\tau^h(\mathbf{a}_p^T \mathbf{y}) \text{ (hard thresholding)}\end{aligned}$$



$$\begin{aligned}\text{et } \min_{\mathbf{x}} \frac{1}{2} \|\mathbf{y} - A\mathbf{x}\|^2 + \lambda \|\mathbf{x}\|_1 &\Leftrightarrow \min_{\mathbf{x}} \frac{1}{2} \left\| A^T \mathbf{y} - \mathbf{x} \right\|^2 + \lambda \|\mathbf{x}\|_1 \\ &\Leftrightarrow \forall p, \min_{x_p} \frac{1}{2} (\mathbf{a}_p^T \mathbf{y} - x_p)^2 + \lambda |x_p| \\ &\Leftrightarrow \forall p, \hat{x}_p = \mathcal{S}_\tau^s(\mathbf{a}_p^T \mathbf{y}) \text{ (soft thresholding)}\end{aligned}$$



- Dans les deux cas, sélection des colonnes les plus corrélées aux données :  $\ell_1 \Leftrightarrow \ell_0$

## Parentèse : conditions d'optimalité des solutions en norme $\ell_1$

### ■ Théorie connue

Considérons le cas sans bruit (modèle exact): 
$$\begin{cases} \hat{\mathbf{x}}^0 = \arg \min \|\mathbf{x}\|_0 \text{ s.t. } \mathbf{y} = \mathbf{Ax} \\ \hat{\mathbf{x}}^1 = \arg \min \|\mathbf{x}\|_1 \text{ s.t. } \mathbf{y} = \mathbf{Ax} \end{cases}$$

## Parenthèse : conditions d'optimalité des solutions en norme $\ell_1$

### ■ Théorie connue

Considérons le cas sans bruit (modèle exact): 
$$\begin{cases} \hat{\mathbf{x}}^0 = \arg \min \|\mathbf{x}\|_0 \text{ s.t. } \mathbf{y} = \mathbf{A}\mathbf{x} \\ \hat{\mathbf{x}}^1 = \arg \min \|\mathbf{x}\|_1 \text{ s.t. } \mathbf{y} = \mathbf{A}\mathbf{x} \end{cases}$$

- **Cohérence mutuelle**  $\mu(\mathbf{A}) := \max_{p \neq q} |\mathbf{a}_p^T \mathbf{a}_q|$

▶ Si  $\|\hat{\mathbf{x}}^1\|_0 < \left(1 + \frac{1}{\mu(\mathbf{A})}\right) / 2$ , alors  $\ell_1 \Leftrightarrow \ell_0$ . [Fuchs, 04]

## Parenthèse : conditions d'optimalité des solutions en norme $\ell_1$

### ■ Théorie connue

Considérons le cas sans bruit (modèle exact): 
$$\begin{cases} \hat{\mathbf{x}}^0 = \arg \min \|\mathbf{x}\|_0 \text{ s.t. } \mathbf{y} = \mathbf{A}\mathbf{x} \\ \hat{\mathbf{x}}^1 = \arg \min \|\mathbf{x}\|_1 \text{ s.t. } \mathbf{y} = \mathbf{A}\mathbf{x} \end{cases}$$

- **Cohérence mutuelle**  $\mu(\mathbf{A}) := \max_{p \neq q} |\mathbf{a}_p^T \mathbf{a}_q|$

▶ Si  $\|\hat{\mathbf{x}}^1\|_0 < \left(1 + \frac{1}{\mu(\mathbf{A})}\right) / 2$ , alors  $\ell_1 \Leftrightarrow \ell_0$ . [Fuchs, 04]

- **Exact Recovery Coefficient** pour un support donné  $\mathcal{S} = \{p | \hat{x}_p^1 \neq 0\}$ :

▶ Si  $\text{ERC}(\mathcal{S}) := 1 - \max_{\ell \notin \mathcal{S}} \|\mathbf{A}_{\mathcal{S}}^T \mathbf{a}_{\ell}\|_1 > 0$ , alors  $\ell_1 \Leftrightarrow \ell_0$ . [Tropp 06]



## Parenthèse : conditions d'optimalité des solutions en norme $\ell_1$

### ■ Théorie connue

Considérons le cas sans bruit (modèle exact): 
$$\begin{cases} \hat{\mathbf{x}}^0 = \arg \min \|\mathbf{x}\|_0 \text{ s.t. } \mathbf{y} = \mathbf{A}\mathbf{x} \\ \hat{\mathbf{x}}^1 = \arg \min \|\mathbf{x}\|_1 \text{ s.t. } \mathbf{y} = \mathbf{A}\mathbf{x} \end{cases}$$

- **Cohérence mutuelle**  $\mu(\mathbf{A}) := \max_{p \neq q} |\mathbf{a}_p^T \mathbf{a}_q|$

▶ Si  $\|\hat{\mathbf{x}}^1\|_0 < \left(1 + \frac{1}{\mu(\mathbf{A})}\right) / 2$ , alors  $\ell_1 \Leftrightarrow \ell_0$ . [Fuchs, 04]

- **Exact Recovery Coefficient** pour un support donné  $\mathcal{S} = \{p | \hat{x}_p^1 \neq 0\}$ :

▶ Si  $\text{ERC}(\mathcal{S}) := 1 - \max_{\ell \notin \mathcal{S}} \|\mathbf{A}_{\mathcal{S}}^T \mathbf{a}_{\ell}\|_1 > 0$ , alors  $\ell_1 \Leftrightarrow \ell_0$ . [Tropp 06]

- **Restricted Isometry Property (RIP)**: soit  $\delta_K > 0$  la plus petite valeur telle que

$$\forall \text{ support } S_K \text{ de taille } K, \forall \mathbf{u}, (1 - \delta_K) \|\mathbf{u}\|_2^2 \leq \|\mathbf{A}_{S_K} \mathbf{u}\|_2^2 \leq (1 + \delta_K) \|\mathbf{u}\|_2^2$$

▶ Si  $\delta_K + \delta_{2K} + \delta_{3K} < 1$ , alors  $\ell_1 \Leftrightarrow \ell_0$ . [Candes et al., 06]

- Spark, Null Space Property, ... [Eldar & Kutyniok, 12]

## Parenthèse : conditions d'optimalité des solutions en norme $\ell_1$ (fin)

■ **Théorie connue** : si  $A$  est “presque orthogonale”, alors  $\ell_1 \Leftrightarrow \ell_0$

- $\mu(A) := \max_{k \neq \ell} |\mathbf{a}_k^T \mathbf{a}_\ell|$ ,  $\|\mathbf{x}\|_0 < \left(1 + \frac{1}{\mu(A)}\right) / 2$
- $\text{ERC}(\mathcal{S}) := 1 - \max_{\ell \notin \mathcal{S}} \|\mathbf{A}_{\mathcal{S}}^T \mathbf{a}_\ell\|_1 > 0$
- $\forall K\text{-element support } S_K, \forall \mathbf{u}, (1 - \delta_K) \|\mathbf{u}\|_2^2 \leq \|\mathbf{A}_{S_K} \mathbf{u}\|_2^2 \leq (1 + \delta_K) \|\mathbf{u}\|_2^2$ 
  - ▶ Hypothèses très fortes sur  $A$ , indépendamment des données
  - ▶ Conditions suffisantes (pire cas)
  - ▶ RIP invérifiable en pratique. . .
  - ▶ *Quelques généralisations au cas bruité*

## Compressed sensing : introduction [Donoho, Candès, Tao, ...]

On veut acquérir un signal temporel / une image / des données  $\mathbf{x} \in \mathbb{R}^P$

*ex :  $\mathbf{x}$  = image de  $1024 \times 1024$  pixels ( $P = 10^6$ )*

## Compressed sensing : introduction [Donoho, Candès, Tao, ...]

On veut acquérir un signal temporel / une image / des données  $\mathbf{x} \in \mathbb{R}^P$

ex :  $\mathbf{x}$  = image de  $1024 \times 1024$  pixels ( $P = 10^6$ )

Hypothèse de compressibilité:

$\mathbf{x}$  admet une représentation parcimonieuse dans un dictionnaire  $D$  approprié

$$\mathbf{x} = D\mathbf{u}, \text{ avec } \|\mathbf{u}\|_0 = K.$$

ex : la transformée en ondelettes de  $\mathbf{x}$  contient l'essentiel de l'information dans très peu de (gros) coefficients:  $\mathbf{u} = W\mathbf{x} \Leftrightarrow \mathbf{x} = W^T\mathbf{u}$ ,  $\mathbf{u}$  parcimonieux.

## Compressed sensing : introduction [Donoho, Candès, Tao, ...]

On veut acquérir un signal temporel / une image / des données  $\mathbf{x} \in \mathbb{R}^P$

ex :  $\mathbf{x}$  = image de  $1024 \times 1024$  pixels ( $P = 10^6$ )

Hypothèse de compressibilité:

$\mathbf{x}$  admet une représentation parcimonieuse dans un dictionnaire  $D$  approprié

$$\mathbf{x} = D\mathbf{u}, \text{ avec } \|\mathbf{u}\|_0 = K.$$

ex : la transformée en ondelettes de  $\mathbf{x}$  contient l'essentiel de l'information dans très peu de (gros) coefficients:  $\mathbf{u} = W\mathbf{x} \Leftrightarrow \mathbf{x} = W^T\mathbf{u}$ ,  $\mathbf{u}$  parcimonieux.

- Acquisition classique : ensemble de  $\mathbf{x}$  (*intensité lumineuse en  $10^6$  pixels*)

## Compressed sensing : introduction [Donoho, Candès, Tao, ...]

On veut acquérir un signal temporel / une image / des données  $\mathbf{x} \in \mathbb{R}^P$

ex :  $\mathbf{x}$  = image de  $1024 \times 1024$  pixels ( $P = 10^6$ )

Hypothèse de compressibilité:

$\mathbf{x}$  admet une représentation parcimonieuse dans un dictionnaire  $D$  approprié

$$\mathbf{x} = D\mathbf{u}, \text{ avec } \|\mathbf{u}\|_0 = K.$$

ex : la transformée en ondelettes de  $\mathbf{x}$  contient l'essentiel de l'information dans très peu de (gros) coefficients:  $\mathbf{u} = W\mathbf{x} \Leftrightarrow \mathbf{x} = W^T\mathbf{u}$ ,  $\mathbf{u}$  parcimonieux.

- Acquisition classique : ensemble de  $\mathbf{x}$  (*intensité lumineuse en  $10^6$  pixels*)
- Pourtant, l'information utile est dans  $K \ll P$  coefficients. . .

## Compressed sensing : introduction [Donoho, Candès, Tao, ...]

On veut acquérir un signal temporel / une image / des données  $\mathbf{x} \in \mathbb{R}^P$

ex :  $\mathbf{x} = \text{image de } 1024 \times 1024 \text{ pixels } (P = 10^6)$

Hypothèse de compressibilité:

$\mathbf{x}$  admet une représentation parcimonieuse dans un dictionnaire  $\mathbf{D}$  approprié

$$\mathbf{x} = \mathbf{D}\mathbf{u}, \text{ avec } \|\mathbf{u}\|_0 = K.$$

ex : la transformée en ondelettes de  $\mathbf{x}$  contient l'essentiel de l'information dans très peu de (gros) coefficients:  $\mathbf{u} = \mathbf{W}\mathbf{x} \Leftrightarrow \mathbf{x} = \mathbf{W}^T\mathbf{u}$ ,  $\mathbf{u}$  parcimonieux.

- Acquisition classique : ensemble de  $\mathbf{x}$  (*intensité lumineuse en  $10^6$  pixels*)
- Pourtant, l'information utile est dans  $K \ll P$  coefficients. . .
- Si l'on savait par avance quels coefficients sont non nuls ( $\mathcal{S} = \{p | u_p \neq 0\}$ ), on pourrait n'acquérir que les coefficients correspondants :

$$\text{ex } \therefore, u_p = \mathbf{d}_p^T \mathbf{x}, \quad \underline{p \in \mathcal{S}}$$

## Compressed sensing : introduction [Donoho, Candès, Tao, ...]

On veut acquérir un signal temporel / une image / des données  $\mathbf{x} \in \mathbb{R}^P$

ex :  $\mathbf{x} = \text{image de } 1024 \times 1024 \text{ pixels } (P = 10^6)$

Hypothèse de compressibilité:

$\mathbf{x}$  admet une représentation parcimonieuse dans un dictionnaire  $\mathbf{D}$  approprié

$$\mathbf{x} = \mathbf{D}\mathbf{u}, \text{ avec } \|\mathbf{u}\|_0 = K.$$

ex : la transformée en ondelettes de  $\mathbf{x}$  contient l'essentiel de l'information dans très peu de (gros) coefficients:  $\mathbf{u} = \mathbf{W}\mathbf{x} \Leftrightarrow \mathbf{x} = \mathbf{W}^T\mathbf{u}$ ,  $\mathbf{u}$  parcimonieux.

- Acquisition classique : ensemble de  $\mathbf{x}$  (*intensité lumineuse en  $10^6$  pixels*)
- Pourtant, l'information utile est dans  $K \ll P$  coefficients. . .
- Si l'on savait par avance quels coefficients sont non nuls ( $\mathcal{S} = \{p | u_p \neq 0\}$ ), on pourrait n'acquérir que les coefficients correspondants :

$$\text{ex } \therefore u_p = \mathbf{d}_p^T \mathbf{x}, \quad \underline{p \in \mathcal{S}}$$

- On reconstruit alors  $\mathbf{x} = \mathbf{D}\mathbf{u}$  (avec  $\forall p \notin \mathcal{S}, u_p = 0$ )



## Matrice de mesure

■ **Idée : acquérir un nombre restreint  $N (< P)$  de mesures sous la forme  $y_n = \phi_n^T \mathbf{x}$**

- Données  $\mathbf{y} = \Phi \mathbf{x}$ ,  $\Phi$  matrice de mesure

Supposition : on sait acquérir *directement*  $\Phi \mathbf{x}$  sans acquérir  $\mathbf{x}$ . (cf + loin)

- Hypothèse de parcimonie  $\mathbf{x} \simeq \mathbf{D}\mathbf{u}$ , avec  $\|\mathbf{u}\|_0 = K$

$$\Rightarrow \mathbf{y} \simeq \Phi \mathbf{D}\mathbf{u}.$$

- ▶ on estime  $\mathbf{u}$  en résolvant le problème parcimonieux :

$$\rightsquigarrow \hat{\mathbf{u}} = \arg \min_{\mathbf{u}} \frac{1}{2} \|\mathbf{y} - \Phi \mathbf{D}\mathbf{u}\|_2^2 \quad \text{s.c.} \quad \|\mathbf{u}\|_0 \leq K$$

- ▶ puis on reconstruit  $\hat{\mathbf{x}} = \mathbf{D}\hat{\mathbf{u}}$

## Matrice de mesure

■ **Idée : acquérir un nombre restreint  $N (< P)$  de mesures sous la forme  $y_n = \phi_n^T x$**

- Données  $\mathbf{y} = \Phi \mathbf{x}$ ,  $\Phi$  matrice de mesure

Supposition : on sait acquérir *directement*  $\Phi \mathbf{x}$  sans acquérir  $\mathbf{x}$ . (cf + loin)

- Hypothèse de parcimonie  $\mathbf{x} \simeq D\mathbf{u}$ , avec  $\|\mathbf{u}\|_0 = K$

$$\Rightarrow \mathbf{y} \simeq \Phi D\mathbf{u}.$$

- ▶ on estime  $\mathbf{u}$  en résolvant le problème parcimonieux :

$$\rightsquigarrow \hat{\mathbf{u}} = \arg \min_{\mathbf{u}} \frac{1}{2} \|\mathbf{y} - \Phi D\mathbf{u}\|_2^2 \quad \text{s.c.} \quad \|\mathbf{u}\|_0 \leq K$$

- ▶ puis on reconstruit  $\hat{\mathbf{x}} = D\hat{\mathbf{u}}$

Coût réduit à l'acquisition, mais coût numérique de reconstruction

## Matrice de mesure

■ **Idée : acquérir un nombre restreint  $N (< P)$  de mesures sous la forme  $y_n = \phi_n^T x$**

- Données  $y = \Phi x$ ,  $\Phi$  matrice de mesure

Supposition : on sait acquérir *directement*  $\Phi x$  sans acquérir  $x$ . (cf + loin)

- Hypothèse de parcimonie  $x \simeq D\mathbf{u}$ , avec  $\|\mathbf{u}\|_0 = K$

$$\Rightarrow y \simeq \Phi D\mathbf{u}.$$

- ▶ on estime  $\mathbf{u}$  en résolvant le problème parcimonieux :

$$\rightsquigarrow \hat{\mathbf{u}} = \arg \min_{\mathbf{u}} \frac{1}{2} \|\mathbf{y} - \Phi D\mathbf{u}\|_2^2 \quad \text{s.c.} \quad \|\mathbf{u}\|_0 \leq K$$

- ▶ puis on reconstruit  $\hat{x} = D\hat{\mathbf{u}}$

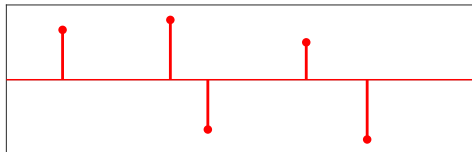
Coût réduit à l'acquisition, mais coût numérique de reconstruction

↪ **Sensing design: comment choisir la matrice de mesure  $\Phi$ ?**

## Choix de la matrice de mesure. Exemple : impulsions

- Supposons que  $x$  est un train d'impulsions :

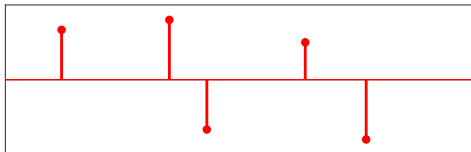
$$x[p] = \sum_{k=1}^K a_k \delta[p - p_k], \text{ d'amplitudes } a_k \text{ et positions } p_k \text{ inconnues}$$



## Choix de la matrice de mesure. Exemple : impulsions

- Supposons que  $x$  est un train d'impulsions :

$$x[p] = \sum_{k=1}^K a_k \delta[p - p_k], \text{ d'amplitudes } a_k \text{ et positions } p_k \text{ inconnues}$$

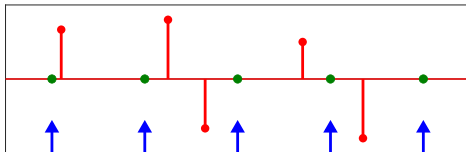


- Je ne peux acquérir que  $N \ll P$  mesures de la forme  $\phi_n^T x$  pour reconstruire  $x$ .  
Comment choisir  $\phi_n$ ?

## Choix de la matrice de mesure. Exemple : impulsions

- Supposons que  $x$  est un train d'impulsions :

$$x[p] = \sum_{k=1}^K a_k \delta[p - p_k], \text{ d'amplitudes } a_k \text{ et positions } p_k \text{ inconnues}$$

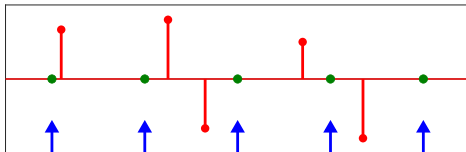


- Je ne peux acquérir que  $N \ll P$  mesures de la forme  $\phi_n^T x$  pour reconstruire  $x$ .  
Comment choisir  $\phi_n$ ?
- Pire choix : prendre  $N$  échantillons temporels différents ( $\phi_n = \delta[n - m]$ )!

## Choix de la matrice de mesure. Exemple : impulsions

- Supposons que  $x$  est un train d'impulsions :

$$x[p] = \sum_{k=1}^K a_k \delta[p - p_k], \text{ d'amplitudes } a_k \text{ et positions } p_k \text{ inconnues}$$



- Je ne peux acquérir que  $N \ll P$  mesures de la forme  $\phi_n^T x$  pour reconstruire  $x$ .  
Comment choisir  $\phi_n$ ?
- Pire choix : prendre  $N$  échantillons temporels différents ( $\phi_n = \delta[n - m]$ )!
- Intuition : construire  $N$  combinaisons linéaires qui contiennent toutes de l'information sur chaque impulsion (ex. coefficients de Fourier...)

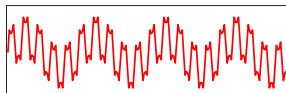
## Choix de la matrice de mesure. Exemple : sinusoides

- Supposons que  $x$  est une somme de *quelques* exponentielles complexes :

$$x[p] = \sum_{k=1}^K a_k e^{j2\pi p \lambda_k}, \text{ d'amplitudes } a_k \in \mathbb{C} \text{ et fréquences } \lambda_k \text{ inconnues}$$

- De manière équivalente,  $x$  a une transformée de Fourier **parcimonieuse** :

$$\mathbf{u} = \mathbf{F}x \text{ parcimonieux (ou } \mathbf{x} = \mathbf{F}^T \mathbf{u}), \text{ avec } \mathbf{F} = \{e^{j2\pi \frac{nk}{N}}\}_{n,k}$$



$x$



$\mathbf{u} = \mathbf{F}x$

- Je ne peux acquérir que  $N \ll P$  mesures de la forme  $\phi_n^T x$  pour reconstruire  $x$ .  
Comment choisir  $\phi_n$ ?



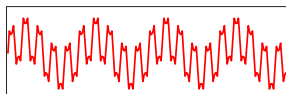
## Choix de la matrice de mesure. Exemple : sinusoides

- Supposons que  $x$  est une somme de *quelques* exponentielles complexes :

$$x[p] = \sum_{k=1}^K a_k e^{j2\pi p \lambda_k}, \text{ , d'amplitudes } a_k \in \mathbb{C} \text{ et fréquences } \lambda_k \text{ inconnues}$$

- De manière équivalente,  $x$  a une transformée de Fourier **parcimonieuse** :

$$\mathbf{u} = \mathbf{F}x \text{ parcimonieux (ou } \mathbf{x} = \mathbf{F}^T \mathbf{u}), \text{ avec } \mathbf{F} = \{e^{j2\pi \frac{nk}{N}}\}_{n,k}$$



$x$



$\mathbf{u} = \mathbf{F}x$

- Je ne peux acquérir **que**  $N \ll P$  mesures de la forme  $\phi_n^T x$  pour reconstruire  $x$ .  
**Comment choisir  $\phi_n$ ?**
- Pire choix** : prendre  $N$  valeurs de la transformée de Fourier!

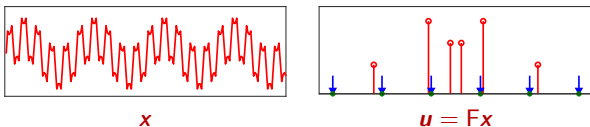
## Choix de la matrice de mesure. Exemple : sinusoides

- Supposons que  $x$  est une somme de *quelques* exponentielles complexes :

$$x[p] = \sum_{k=1}^K a_k e^{j2\pi p \lambda_k}, \text{ d'amplitudes } a_k \in \mathbb{C} \text{ et fréquences } \lambda_k \text{ inconnues}$$

- De manière équivalente,  $x$  a une transformée de Fourier **parcimonieuse** :

$$\mathbf{u} = \mathbf{F}x \text{ parcimonieux (ou } \mathbf{x} = \mathbf{F}^T \mathbf{u}), \text{ avec } \mathbf{F} = \{e^{j2\pi \frac{nk}{N}}\}_{n,k}$$



- Je ne peux acquérir **que**  $N \ll P$  mesures de la forme  $\phi_n^T x$  pour reconstruire  $x$ .  
**Comment choisir  $\phi_n$ ?**
- Pire choix** : prendre  $N$  valeurs de la transformée de Fourier!
- Intuition** : prendre  $N$  échantillons du signal...

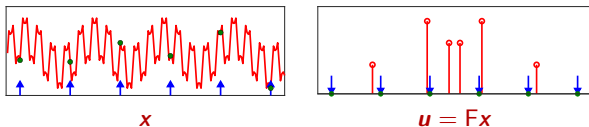
## Choix de la matrice de mesure. Exemple : sinusoides

- Supposons que  $x$  est une somme de *quelques* exponentielles complexes :

$$x[p] = \sum_{k=1}^K a_k e^{j2\pi p \lambda_k}, \text{ d'amplitudes } a_k \in \mathbb{C} \text{ et fréquences } \lambda_k \text{ inconnues}$$

- De manière équivalente,  $x$  a une transformée de Fourier **parcimonieuse** :

$$\mathbf{u} = \mathbf{F}x \text{ parcimonieux (ou } \mathbf{x} = \mathbf{F}^T \mathbf{u}), \text{ avec } \mathbf{F} = \{e^{j2\pi \frac{nk}{N}}\}_{n,k}$$



- Je ne peux acquérir que  $N \ll P$  mesures de la forme  $\phi_n^T x$  pour reconstruire  $x$ .  
Comment choisir  $\phi_n$ ?
- Pire choix : prendre  $N$  valeurs de la transformée de Fourier!
- Intuition : prendre  $N$  échantillons du signal...

## Choix de la matrice de mesure. Retour au cas général

■ **Idée : acquérir un nombre restreint  $N (< P)$  de mesures sous la forme  $y_n = \phi_n^T x$**

- Données  $y = \Phi x$ ,  $\Phi$  matrice de mesure

Supposition : on sait acquérir *directement*  $\Phi x$  sans acquérir  $x$ . (cf + loin)

- Hypothèse de parcimonie  $x \simeq D u$ , avec  $\|u\|_0 = K$

$$\Rightarrow y \simeq \Phi D u.$$

- ▶ on estime  $u$  en résolvant le problème parcimonieux :

$$\rightsquigarrow \hat{u} = \arg \min_u \frac{1}{2} \|y - \Phi D u\|_2^2 \quad \text{s.c.} \quad \|u\|_0 \leq K$$

- ▶ puis on reconstruit  $\hat{x} = D \hat{u}$

↪ **Sensing design: comment choisir la matrice de mesure  $\Phi$ ?**

## Choix de la matrice de mesure. Retour au cas général

■ **Idée : acquérir un nombre restreint  $N (< P)$  de mesures sous la forme  $y_n = \phi_n^T x$**

- Données  $y = \Phi x$ ,  $\Phi$  matrice de mesure

Supposition : on sait acquérir *directement*  $\Phi x$  sans acquérir  $x$ . (cf + loin)

- Hypothèse de parcimonie  $x \simeq D\mathbf{u}$ , avec  $\|\mathbf{u}\|_0 = K$

$$\Rightarrow y \simeq \Phi D\mathbf{u}.$$

- ▶ on estime  $\mathbf{u}$  en résolvant le problème parcimonieux :

$$\rightsquigarrow \hat{\mathbf{u}} = \arg \min_{\mathbf{u}} \frac{1}{2} \|\mathbf{y} - \Phi D\mathbf{u}\|_2^2 \quad \text{s.c.} \quad \|\mathbf{u}\|_0 \leq K$$

- ▶ puis on reconstruit  $\hat{x} = D\hat{\mathbf{u}}$

↪ **Sensing design: comment choisir la matrice de mesure  $\Phi$ ?**

**Solution : rendre les mesures incohérentes avec le dictionnaire D!**

## Choix de la matrice de mesure. Retour au cas général

■ **Idée : acquérir un nombre restreint  $N (< P)$  de mesures sous la forme  $y_n = \phi_n^T x$**

- Données  $y = \Phi x$ ,  $\Phi$  matrice de mesure

Supposition : on sait acquérir *directement*  $\Phi x$  sans acquérir  $x$ . (cf + loin)

- Hypothèse de parcimonie  $x \simeq D\mathbf{u}$ , avec  $\|\mathbf{u}\|_0 = K$

$$\Rightarrow y \simeq \Phi D\mathbf{u}.$$

- ▶ on estime  $\mathbf{u}$  en résolvant le problème parcimonieux :

$$\rightsquigarrow \hat{\mathbf{u}} = \arg \min_{\mathbf{u}} \frac{1}{2} \|\mathbf{y} - \Phi D\mathbf{u}\|_2^2 \quad \text{s.c.} \quad \|\mathbf{u}\|_0 \leq K$$

- ▶ puis on reconstruit  $\hat{x} = D\hat{\mathbf{u}}$

↪ **Sensing design: comment choisir la matrice de mesure  $\Phi$ ?**

**Solution : rendre les mesures incohérentes avec le dictionnaire  $D$ !**

- ▶ *chaque* mesure capture une information *différente* sur *toutes* les composantes

## Choix de la matrice de mesure. Retour au cas général

■ **Idée : acquérir un nombre restreint  $N (< P)$  de mesures sous la forme  $y_n = \phi_n^T x$**

- Données  $y = \Phi x$ ,  $\Phi$  matrice de mesure

Supposition : on sait acquérir *directement*  $\Phi x$  sans acquérir  $x$ . (cf + loin)

- Hypothèse de parcimonie  $x \simeq D u$ , avec  $\|u\|_0 = K$

$$\Rightarrow y \simeq \Phi D u.$$

- ▶ on estime  $u$  en résolvant le problème parcimonieux :

$$\rightsquigarrow \hat{u} = \arg \min_u \frac{1}{2} \|y - \Phi D u\|_2^2 \quad \text{s.c.} \quad \|u\|_0 \leq K$$

- ▶ puis on reconstruit  $\hat{x} = D \hat{u}$

↪ **Sensing design: comment choisir la matrice de mesure  $\Phi$ ?**

**Solution : rendre les mesures incohérentes avec le dictionnaire  $D$ !**

- ▶ *chaque* mesure capture une information *différente* sur *toutes* les composantes
- ▶ la matrice résultante  $A = \Phi D$  est plus à même de vérifier les conditions RIP

## Choix de la matrice de mesure. Retour au cas général

■ **Idée : acquérir un nombre restreint  $N (< P)$  de mesures sous la forme  $y_n = \phi_n^T x$**

- Données  $y = \Phi x$ ,  $\Phi$  matrice de mesure

Supposition : on sait acquérir *directement*  $\Phi x$  sans acquérir  $x$ . (cf + loin)

- Hypothèse de parcimonie  $x \simeq Du$ , avec  $\|u\|_0 = K$

$$\Rightarrow y \simeq \Phi Du.$$

- ▶ on estime  $u$  en résolvant le problème parcimonieux :

$$\rightsquigarrow \hat{u} = \arg \min_u \frac{1}{2} \|y - \Phi Du\|_2^2 \text{ s.c. } \|u\|_0 \leq K$$

- ▶ puis on reconstruit  $\hat{x} = D\hat{u}$

↪ **Sensing design: comment choisir la matrice de mesure  $\Phi$ ?**

**Solution : rendre les mesures incohérentes avec le dictionnaire  $D$ !**

- ▶ *chaque* mesure capture une information *différente* sur *toutes* les composantes
- ▶ la matrice résultante  $A = \Phi D$  est plus à même de vérifier les conditions RIP
- ▶  $\Phi =$  matrice pseudo-aléatoire  $\Rightarrow$  RIP avec une forte probabilité

$$A^T A = D^T \underbrace{\Phi^T \Phi}_{\sim I} D \sim I$$



### Résumé :

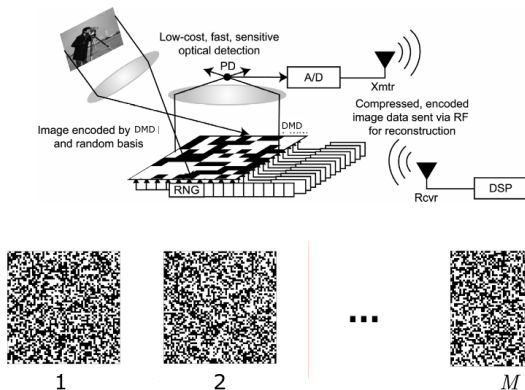
- $x$  admet une décomposition parcimonieuse  $x = Du$ ,  $u$  parcimonieux
- on construit une matrice de mesure  $\Phi$  telle que
  - ▶ les colonnes de  $\Phi$  soient *incohérentes* avec  $D$
  - ▶ **l'acquisition de  $\Phi x$  au lieu de  $x$  soit possible techniquement ...**

Exemple : caméra à un pixel! [Baraniuk *et al.*, Rice Univ.]

---

<sup>0</sup><https://www.ima.umn.edu/materials/2006-2007/ND6.4-15.07/3890/baraniuk-IMA-CScamera-june07.pdf>

## Single Pixel Camera

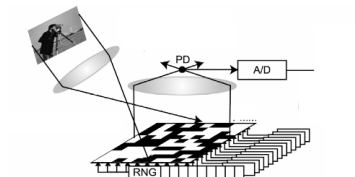


- Masques binaires sur chaque pixel
- Chaque acquisition  $y_n =$  une valeur mélangeant aléatoirement une partie des pixels

$$y_n = \mathbf{b}_n^T \mathbf{x}, \mathbf{b} \in \{0; 1\}^P, n = 1, \dots, M$$

<sup>0</sup><https://www.ima.umn.edu/materials/2006-2007/ND6.4-15.07/3890/baraniuk-IMA-CScamera-june07.pdf>

## First Image Acquisition



target  
65536 pixels



11000 measurements  
(16%)



1300 measurements  
(2%)



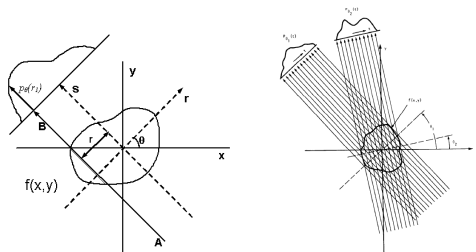
- Masques binaires sur chaque pixel
- Chaque acquisition  $y_n =$  une valeur mélangeant aléatoirement une partie des pixels

$$y_n = \mathbf{b}_n^T \mathbf{x}, \mathbf{b} \in \{0; 1\}^P, n = 1, \dots, M$$

<sup>0</sup><https://www.ima.umn.edu/materials/2006-2007/ND6.4-15.07/3890/baraniuk-IMA-CScamera-june07.pdf>

## Quelques cadres "naturels" d'acquisition : tomographie

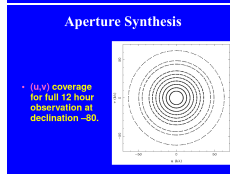
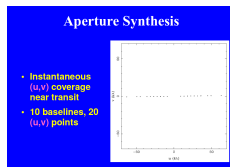
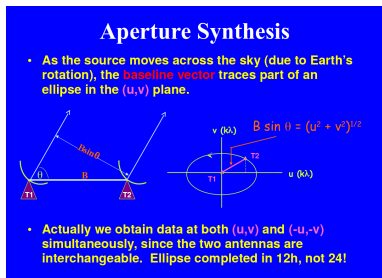
### ■ Tomographie à rayons X = acquisition de la transformée de Radon de l'image



[Kak & Slaney, 88]

- Si toutes les projections  $(r, \theta)$  sont acquises:  $\mathbf{p} = \mathbf{R}\mathbf{x}$
- Sous-échantillonnage:  $\mathbf{p} = \mathbf{M}\mathbf{R}\mathbf{x}$ , où  $\mathbf{M}$  sélectionne des points  $(r, \theta)$

## ■ Synthèse d'ouverture / radiointérométrie



- Fonction de cohérence entre deux antennes = une mesure de la transformée de Fourier (2D)
- Données  $y = MFx$

<sup>0</sup>Interferometry in Radio Astronomy, Tony Wong, ATNF

## Quelques cadres "naturels" d'acquisition : radioastronomie

### ■ Synthèse d'ouverture / radiointerférométrie

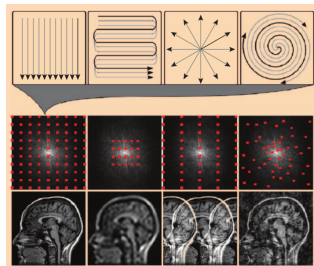
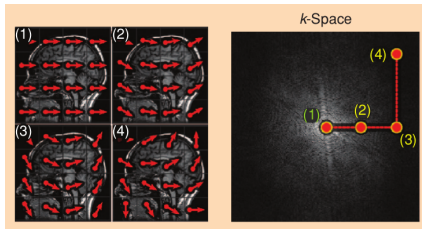


Atacama Large Millimeter/submillimeter Array (ALMA)



Square Kilometre Array (SKA),  
artist view

## Quelques cadres "naturels" d'acquisition : IRM

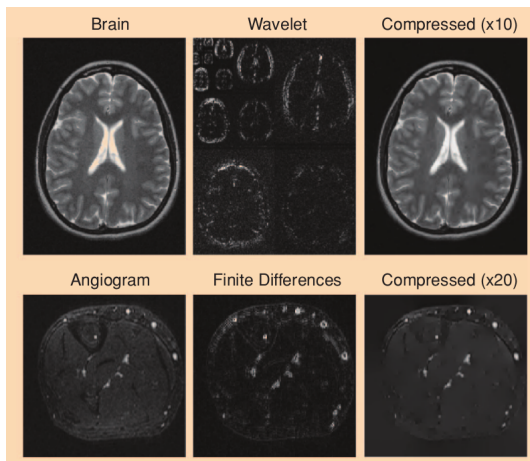


- Imagerie par Résonance Magnétique = acquisition dans le plan de Fourier
- Données  $y = MFx$

<sup>0</sup>M. Lustig, D. L. Donoho, J. M. Santos and J. M. Pauly, "Compressed Sensing MRI," in IEEE Signal Processing Magazine, vol. 25, no. 2, pp. 72-82, March 2008



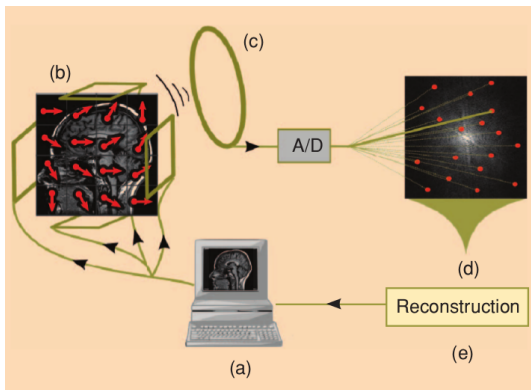
## Quelques cadres "naturels" d'acquisition : IRM



Parcimonie des images IRM dans des espaces appropriés

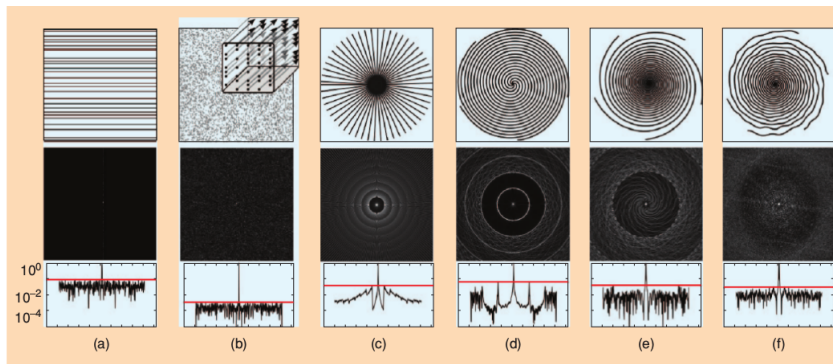
<sup>0</sup>M. Lustig, D. L. Donoho, J. M. Santos and J. M. Pauly, "Compressed Sensing MRI," in IEEE Signal Processing Magazine, vol. 25, no. 2, pp. 72-82, March 2008

## Quelques cadres "naturels" d'acquisition : Imagerie par Résonance Magnétique



<sup>0</sup>M. Lustig, D. L. Donoho, J. M. Santos and J. M. Pauly, "Compressed Sensing MRI," in IEEE Signal Processing Magazine, vol. 25, no. 2, pp. 72-82, March 2008

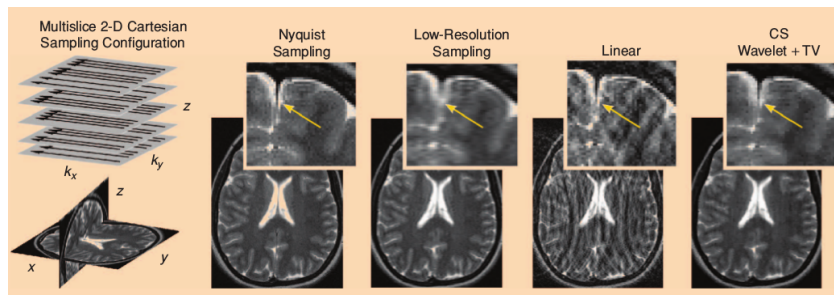
## Quelques cadres "naturels" d'acquisition : IRM



[FIG6] PSFs of various sampling trajectories: (a) random lines in 2-D, (b) random points in 2-D or cross section of random lines in 3-D, (c) radial, (d) uniform spirals, (e) variable density spirals, and (f) variable density perturbed spirals. The height of the red lines measures coherence.

<sup>0</sup>M. Lustig, D. L. Donoho, J. M. Santos and J. M. Pauly, "Compressed Sensing MRI," in IEEE Signal Processing Magazine, vol. 25, no. 2, pp. 72-82, March 2008

## Quelques cadres "naturels" d'acquisition : IRM

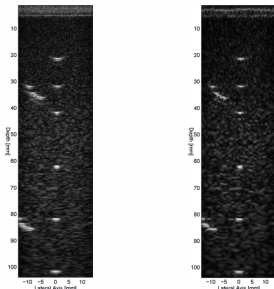


<sup>0</sup>M. Lustig, D. L. Donoho, J. M. Santos and J. M. Pauly, "Compressed Sensing MRI," in IEEE Signal Processing Magazine, vol. 25, no. 2, pp. 72-82, March 2008

# Et le CND dans tout ça??

## ■ Imagerie ultrasonore

- Acquisition directe de coefficients de Fourier des A-scans [Spaulding *et al.*, 2015]



(a) Conventional beamforming (b) CS subarray beamforming  
**Fig. 6:** Phantom image using (a) conventionally beamforming and (b) and subarray beamforming with square wave mixing and CS reconstruction (29x sample reduction).

## ■ Au-delà

- Imagerie US: optimisation de séquences de tir
- Applications à d'autres modalités de mesure (EM, thermique, ...)
- Positionnement de capteurs

### ■ Parcimonie “historique”

- S. Mallat (2009). *A wavelet tour of signal processing: The sparse way*. Elsevier/Academic Press.

### ■ Algorithmes

- F. Bach, R. Jenatton, J. Mairal, and G. Obozinski (2011). *Optimization with Sparsity-Inducing Penalties*. Foundations and Trends in Machine Learning.
- A. Miller (2002). *Subset Selection in Regression*, Chapman & Hall/CRC Monographs on Statistics & Applied Probability, CRC Press.
- $\ell_1$  ou  $\ell_0$  : D. Bertsimas, A. King and R. Mazumder (2016). *Best Subset Selection via a Modern Optimization Lens*, Annals of Statistics.
- Une réponse orientée : T. Hastie, R. Tibshirani and R. Tibshirani (2020). *Best Subset, Forward Stepwise or Lasso? Analysis and Recommendations Based on Extensive Comparisons*. Statistical Science.

### ■ Compressed sensing

- <http://dsp.rice.edu/cs/>
- D.L. Donoho (2006). *Compressed sensing*. IEEE Transactions on Information Theory.
- E.J. Candès & M.B. Wakin (2008). *An Introduction To Compressive Sampling*. IEEE Signal Processing Magazine.
- Y. C. Eldar and G. Kutyniok (2012). *Compressed Sensing: Theory and Applications*, Cambridge University Press.

### ■ Tutoriels (G. Peyré)

- Numerical tours <http://www.numerical-tours.com/>
- Mathematical tours <https://mathematical-tours.github.io/>?

### ■ Bibliothèques / toolboxes

- SPAMS (SPArse Modeling Software) :  
<http://spams-devel.gforge.inria.fr/>
- UNLocBox (optimisation convexe, parcimonie structurée) :  
<https://epfl-lts2.github.io/unlocbox-html/>
- Pages web des auteurs

Sebastien.Bourguignon@ec-nantes.fr