Analyse et Représentation des Signaux (ARSIG)

IV. Représentations parcimonieuses

ECN option DATASIM

Sébastien Bourguignon

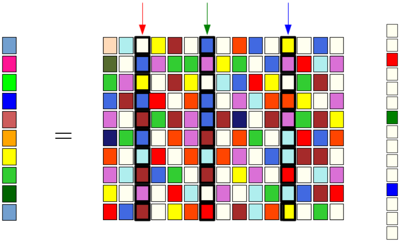
Sebastien.Bourguignon@ec-nantes.fr

Année 2021-2022

Hypothèse de compressibilité / parcimonie

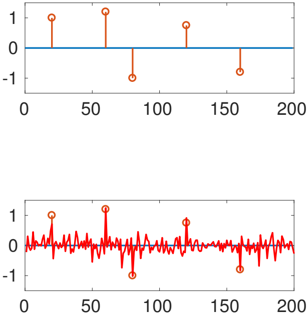
Un signal / une image / un ensemble de données peut être représenté de manière exacte ou approchée avec un faible nombre de coefficients choisis dans un espace de représentation approprié.

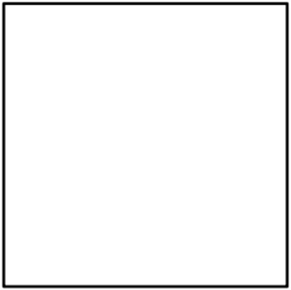
Synthèse parcimonieuse : $\boldsymbol{y} \simeq \mathrm{A}\boldsymbol{x}$, $\boldsymbol{x}$ essentiellement composé de valeurs nulles



$$\boldsymbol{y} = \mathrm{A} \quad \boldsymbol{x}$$

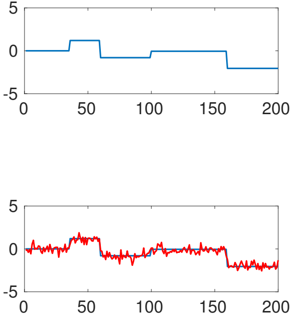
Modèles parcimonieux "naturels"

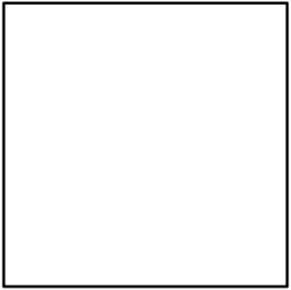




A

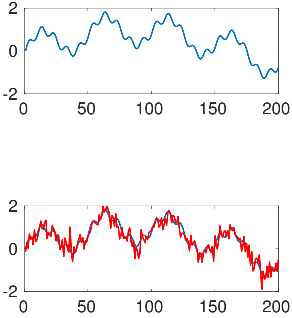
Modèles parcimonieux “naturels”

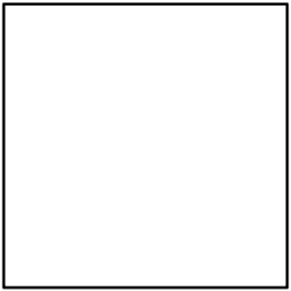




A

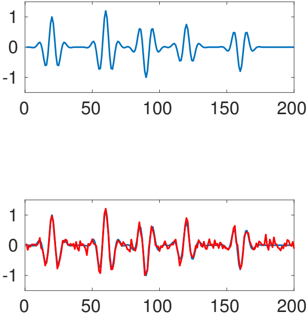
Modèles parcimonieux “naturels”

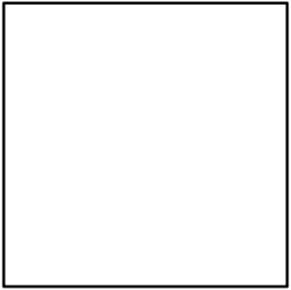




A

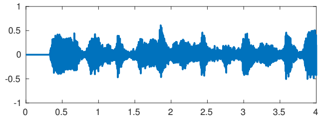
Modèles parcimonieux “naturels”



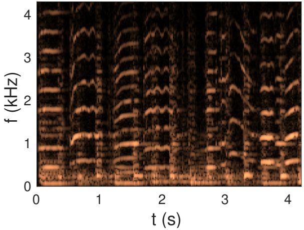


A

Modèles parcimonieux issus de “transformées”

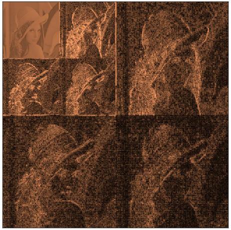


$\boldsymbol{x}$ = extrait de `daphne.wav`



$|\mathrm{STFT}(\boldsymbol{x})|$

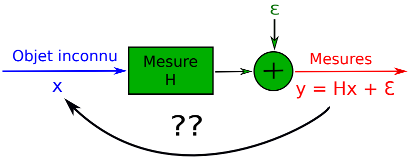
$\boldsymbol{x}$



$|\mathrm{DWT}(\boldsymbol{x})|$

Représentations parcimonieuses : pourquoi

- Débruitage
- Compression
- Extraction de descripteurs
- Problèmes inverses



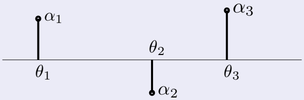
  - $\boldsymbol{x}$ parcimonieux directement (train d'impulsions)
  - ou admet une représentation dans un dictionnaire : $\boldsymbol{x} = A\boldsymbol{u}$

Mais aussi...

- Identification de modèles non linéaires [Tang *et al.*, 13]

Modèle non linéaire

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

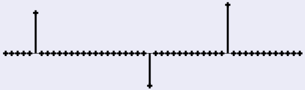


$$\rightarrow \widehat{\boldsymbol{\theta}},\ \widehat{\boldsymbol{\alpha}},\ \widehat{K}?$$

$$\xrightarrow{\theta_p^d = p\,\Delta_\theta}$$

Modèle discrétisé linéaire + parcimonie

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



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

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

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

$\rightsquigarrow \ \widehat{K} = \mathrm{Card}\ \mathcal{Q}$

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

Calcul de la solution parcimonieuse

- Tous ces problèmes reposent sur la définition / la construction d'un espace de représentation A adapté aux données
  - Connaissance *a priori* des propriétés du signal / de l'image / des données
  - Apprentissage de représentations

Calcul de la solution parcimonieuse : étant donnés $\boldsymbol{y}$ et A, comment trouver $\boldsymbol{x}$?

  - Simple si A orthogonal
  - En pratique, rarement le cas!

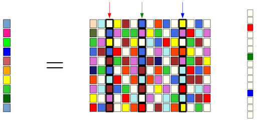
Vocabulaire

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

  Approximation parcimonieuse (avec bruit) : $\boldsymbol{y} \simeq \mathrm{A}\boldsymbol{x}$ avec $\boldsymbol{x}$ parcimonieux

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

$\rightsquigarrow$ sélection de colonnes actives

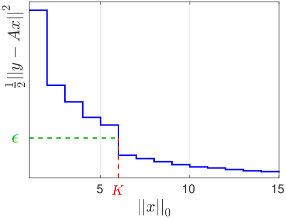


- "Norme" ℓ_0 : $\|\boldsymbol{x}\|_0 = \mathrm{Card}\{p | x_p \neq 0\}$

Formulation de problèmes d'optimisation

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

- $\rightsquigarrow \min_{\boldsymbol{x}} \frac{1}{2}\|\boldsymbol{y} - A\boldsymbol{x}\|^2$ s.c. $\|\boldsymbol{x}\|_0 \leq K$
- $\rightsquigarrow \min_{\boldsymbol{x}} \|\boldsymbol{x}\|_0$ s.c. $\frac{1}{2}\|\boldsymbol{y} - A\boldsymbol{x}\|^2 \leq \epsilon$
- $\rightsquigarrow \min_{\boldsymbol{x}} \frac{1}{2}\|\boldsymbol{y} - A\boldsymbol{x}\|^2 + \mu\|\boldsymbol{x}\|_0$



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

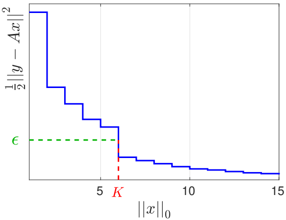
Formulation de problèmes d'optimisation

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

$\rightsquigarrow$ $\min_{\boldsymbol{x}} \frac{1}{2}\|\boldsymbol{y} - A\boldsymbol{x}\|^2$ s.c. $\|\boldsymbol{x}\|_0 \leq K$

$\rightsquigarrow$ $\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_0$ s.c. $\frac{1}{2}\|\boldsymbol{y} - A\boldsymbol{x}\|^2 \leq \epsilon$

$\rightsquigarrow$ $\min_{\boldsymbol{x}} \frac{1}{2}\|\boldsymbol{y} - A\boldsymbol{x}\|^2 + \mu\|\boldsymbol{x}\|_0$

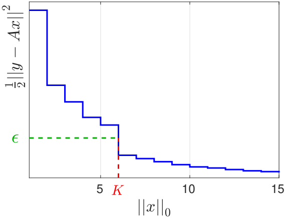


- Complexité *combinatoire* = sélection du support de $\boldsymbol{x}$: $S = \{p | x_p \neq 0\}$
- **Optimisation exacte accessible en *petite* dimension**
  - $\rightsquigarrow$ algorithmes Branch-and-Bound [Bienstock 96, Ben Mhenni *et al.* 2020, ...]

Formulation de problèmes d'optimisation

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

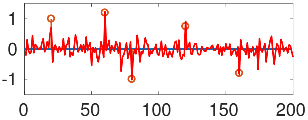
- $\rightsquigarrow$ $\min_{\boldsymbol{x}} \frac{1}{2}\|\boldsymbol{y} - A\boldsymbol{x}\|^2$ s.c. $\|\boldsymbol{x}\|_0 \leq K$
- $\rightsquigarrow$ $\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_0$ s.c. $\frac{1}{2}\|\boldsymbol{y} - A\boldsymbol{x}\|^2 \leq \epsilon$
- $\rightsquigarrow$ $\min_{\boldsymbol{x}} \frac{1}{2}\|\boldsymbol{y} - A\boldsymbol{x}\|^2 + \mu\|\boldsymbol{x}\|_0$



- Complexité *combinatoire* = sélection du support de $\boldsymbol{x}$: $S = \{p | x_p \neq 0\}$
- **Optimisation exacte accessible en *petite* dimension**
  - $\rightsquigarrow$ algorithmes Branch-and-Bound [Bienstock 96, Ben Mhenni *et al.* 2020, ...]
- Sinon :
  - $\rightsquigarrow$ **optimisation *locale* : heuristiques d'exploration combinatoire (partielle)**
  - $\rightsquigarrow$ **relaxations : résolution d'un problème plus simple**

Problème basique : parcimonie dans l'espace des données

- On suppose $\boldsymbol{y} \simeq \boldsymbol{x}$, avec $\boldsymbol{x}$ parcimonieux ($A = I$)



TD : Comment obtient-on $\widehat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \|\boldsymbol{y} - \boldsymbol{x}\|^2$ tel que $\|\boldsymbol{x}\|_0 \leq K$?

$=$ meilleure approximation de $\boldsymbol{y}$ par $\boldsymbol{x}$ à seulement K valeurs non nulles?

Parcimonie dans une base orthogonale

- On suppose $\boldsymbol{y} \simeq \mathrm{A}\boldsymbol{x}$, avec $\boldsymbol{x}$ parcimonieux et A carrée orthonormale

  Ex : Transformée de Fourier Discrète, TCD, ondelettes orthogonales, ...

$$\min_{\boldsymbol{x}} \tfrac{1}{2}\|\boldsymbol{y} - \mathrm{A}\boldsymbol{x}\|^2 \text{ s.c. } \|\boldsymbol{x}\|_0 \leq K \iff \min_{\boldsymbol{x}} \tfrac{1}{2}\left\|\mathrm{A}^T\boldsymbol{y} - \boldsymbol{x}\right\|^2 \text{ s.c. } \|\boldsymbol{x}\|_0 \leq K$$

- Procédure d'estimation :
  - passage dans l'espace transformé : $\boldsymbol{z} = \mathrm{A}^T\boldsymbol{y}$
  - seuillage des composantes de $\boldsymbol{z} \rightsquigarrow \widehat{\boldsymbol{x}}$

Cas général : A transformation quelconque, éventuellement redondante

- On suppose $\boldsymbol{y} \simeq \mathrm{A}\boldsymbol{x}$, avec $\boldsymbol{x}$ parcimonieux, A non orthogonale et redondante

A redondante : $N_{\text{colonnes}} > N_{\text{lignes}} \rightsquigarrow$ possibilités de modélisation $\nearrow$

- A = union de bases de propriétés différentes $\mathrm{A} = [\mathrm{A}_1, \dots, \mathrm{A}_M]$

$$\boldsymbol{y} \simeq \begin{bmatrix} \mathrm{A}_1 & \mathrm{A}_2 & \cdot & \cdot & \cdot & \mathrm{A}_M \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \\ \cdot \\ \cdot \\ \cdot \\ \boldsymbol{x}_M \end{bmatrix} = \boldsymbol{y}_1 + \cdots + \boldsymbol{y}_M$$

  - Capter des propriétés complémentaires
  - Séparation en plusieurs composantes $\boldsymbol{y}_1, \dots, \boldsymbol{y}_M$
  - Exemple : $\mathrm{A} = [\mathrm{I}, \mathrm{DWT}^{-1}]$

- Transformées cohérentes redondantes
  (ex. temps-fréquence, paquets d'ondelettes)
  - Plus de précision dans le modèle

Cas général : A transformation quelconque, éventuellement <u>redondante</u>

- On suppose $\boldsymbol{y} \simeq \mathrm{A}\boldsymbol{x}$, avec $\boldsymbol{x}$ parcimonieux, A non orthogonale et redondante
  - Pas de formulation explicite ($\neq$ seuillage)
  - Infinité de solutions de $\boldsymbol{y} = \mathrm{A}\boldsymbol{x}$ ... non parcimonieuses

TD : Soit $\mathrm{A} = [\mathrm{A}_1, \mathrm{A}_2]$, où A_1 et A_2 orthogonales (ex: $\mathrm{A}_1 = \mathrm{I}$, $\mathrm{A}_2 = \mathrm{DWT}^{-1}$).
Quelle solution obtient-on par seuillage de $\mathrm{A}^T\boldsymbol{y}$?

Cas général : A transformation quelconque, éventuellement redondante

■ On suppose $\boldsymbol{y} \simeq \mathrm{A}\boldsymbol{x}$, avec $\boldsymbol{x}$ parcimonieux, A non orthogonale et redondante

- Pas de formulation explicite ($\neq$ seuillage)
- Infinité de solutions de $\boldsymbol{y} = \mathrm{A}\boldsymbol{x}$ ... non parcimonieuses

TD : Soit $\mathrm{A} = [\mathrm{A}_1, \mathrm{A}_2]$, où A_1 et A_2 orthogonales (ex: $\mathrm{A}_1 = \mathrm{I}$, $\mathrm{A}_2 = \mathrm{DWT}^{-1}$).
Quelle solution obtient-on par seuillage de $\mathrm{A}^T \boldsymbol{y}$?

TD : Soit A carrée, inversible, mais non orthogonale.
Quelle solution obtient-on par seuillage de $\mathrm{A}^{-1}\, \boldsymbol{y}$?

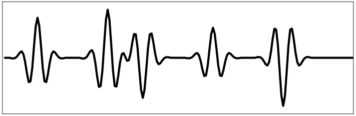
Optimisation parcimonieuse : algorithmes gloutons

$$\min_{\boldsymbol{x}} \tfrac{1}{2}\|\boldsymbol{y} - \mathrm{A}\boldsymbol{x}\|^2 \;\text{ s.c. }\; \|\boldsymbol{x}\|_0 \leq K, \quad \min_{\boldsymbol{x}} \|\boldsymbol{x}\|_0 \;\text{ s.c. }\; \tfrac{1}{2}\|\boldsymbol{y} - \mathrm{A}\boldsymbol{x}\|^2 \leq \epsilon$$

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

Forward selection [Efroymson, 60]...

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



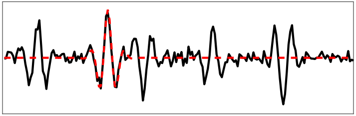
Optimisation parcimonieuse : algorithmes gloutons

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

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

Forward selection [Efroymson, 60]...

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



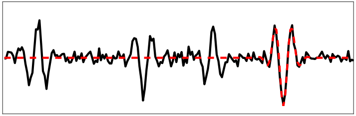
Optimisation parcimonieuse : algorithmes gloutons

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

■ Construction itérative d'une solution parcimonieuse (algorithmes “gloutons”)

Forward selection [Efroymson, 60]...

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



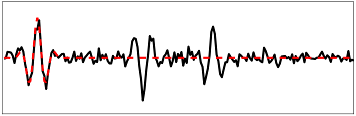
Optimisation parcimonieuse : algorithmes gloutons

$$\boxed{\min_{\boldsymbol{x}} \frac{1}{2}\|\boldsymbol{y}-\mathrm{A}\boldsymbol{x}\|^2 \ \text{ s.c. } \ \|\boldsymbol{x}\|_0 \le K, \quad \min_{\boldsymbol{x}} \|\boldsymbol{x}\|_0 \ \text{ s.c. } \ \frac{1}{2}\|\boldsymbol{y}-\mathrm{A}\boldsymbol{x}\|^2 \le \epsilon}$$

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

Forward selection [Efroymson, 60]...

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



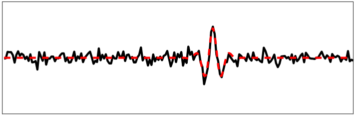
Optimisation parcimonieuse : algorithmes gloutons

$$\min_{\boldsymbol{x}} \frac{1}{2}\|\boldsymbol{y} - \mathrm{A}\boldsymbol{x}\|^2 \;\text{ s.c. }\; \|\boldsymbol{x}\|_0 \leq K, \quad \min_{\boldsymbol{x}} \|\boldsymbol{x}\|_0 \;\text{ s.c. }\; \frac{1}{2}\|\boldsymbol{y} - \mathrm{A}\boldsymbol{x}\|^2 \leq \epsilon$$

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

Forward selection [Efroymson, 60]...

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



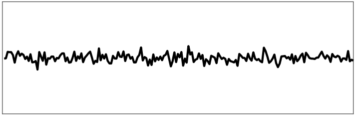
Optimisation parcimonieuse : algorithmes gloutons

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

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

Forward selection [Efroymson, 60]...

1. Support $S^{(0)} = \emptyset$, $\boldsymbol{x}^{(0)} = 0$, résidu $\boldsymbol{r}^{(0)} = \boldsymbol{y}$
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 $\boldsymbol{x}_{S^{(t)}}$ et mise à jour du résidu $\boldsymbol{r}^{(t)} = \boldsymbol{y} - \mathrm{A}_{S^{(t)}} \boldsymbol{x}_{S^{(t)}}$
4. $t \leftarrow t + 1$ et retour en 2. jusqu'à $\left\|\boldsymbol{r}^{(t)}\right\| < \epsilon$ ou Card $\boldsymbol{S}^{(t)} = K$

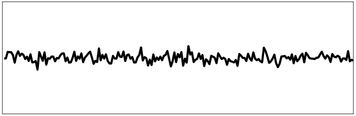


$$\min_{\boldsymbol{x}} \frac{1}{2}\|\boldsymbol{y} - A\boldsymbol{x}\|^2 \text{ s.c. } \|\boldsymbol{x}\|_0 \leq K, \quad \min_{\boldsymbol{x}} \|\boldsymbol{x}\|_0 \text{ s.c. } \frac{1}{2}\|\boldsymbol{y} - A\boldsymbol{x}\|^2 \leq \epsilon$$

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

  Forward selection [Efroymson, 60]...

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



- Exemple : Orthogonal Matching Pursuit [Pati *et al.*, 93]

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

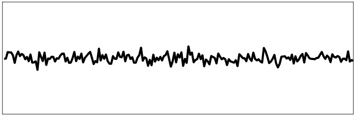
Optimisation parcimonieuse : algorithmes gloutons

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

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

Forward selection [Efroymson, 60]...

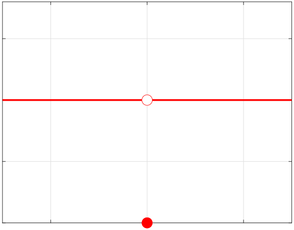
1. Support $S^{(0)} = \emptyset$, $\boldsymbol{x}^{(0)} = 0$, résidu $\boldsymbol{r}^{(0)} = \boldsymbol{y}$
2. Itération t : choix d'une composante $k^{(t)} \notin S^{(t-1)} \rightarrow S^{(t)} = S^{(t-1)} \cup \{k^{(t)}\}$
3. Calcul des amplitudes $\boldsymbol{x}_{S^{(t)}}$ et mise à jour du résidu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



- Nombreuses heuristiques [OLS, SBR, COSAMP, ...]
  - Choix sous le compromis qualité / complexité calculatoire
  - Interférences entre composantes $\rightsquigarrow$ chances faibles de retrouver l'optimum global

Optimisation parcimonieuse : relaxations continues

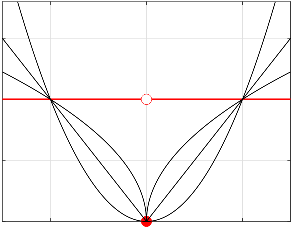
$$\min_{\boldsymbol{x}} \frac{1}{2}\|\boldsymbol{y} - \mathrm{A}\boldsymbol{x}\|^2 + \lambda \underbrace{\|\boldsymbol{x}\|_0}_{\sum_p 1_{\{x_p \neq 0\}}}$$



Optimisation parcimonieuse : relaxations continues

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

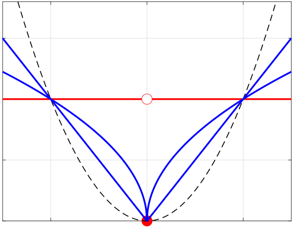
avec φ continue, paire, croissante sur $\mathbb{R}^+$



Optimisation parcimonieuse : relaxations continues

$$\min_{\boldsymbol{x}} \tfrac{1}{2}\|\boldsymbol{y} - \mathrm{A}\boldsymbol{x}\|^2 + \lambda \underbrace{\|\boldsymbol{x}\|_0}_{\sum_p 1_{\{x_p \neq 0\}}} \quad \rightsquigarrow \quad \min_{\boldsymbol{x}} \tfrac{1}{2}\|\boldsymbol{y} - \mathrm{A}\boldsymbol{x}\|^2 + \lambda \underbrace{\phi(\boldsymbol{x})}_{\sum_p \varphi(x_p)}$$

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



La solution est parcimonieuse si et seulement si φ non différentiable en 0

[Moulin & Liu 98]

TD : Parcimonie et non-différentiabilité

La solution est parcimonieuse si et seulement si φ non différentiable en 0

[Moulin & Liu 98]

Optimisation continue non différentiable

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

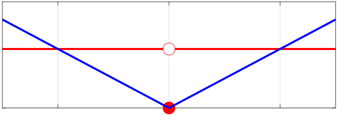
LE sujet chaud depuis 10 ans / algorithmes parcimonieux! [Bach *et al.* 12]

Optimisation continue non différentiable

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

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

■ Norme ℓ_1 $(\varphi(x_p) = |x_p|)$: critère convexe, géométrie simple



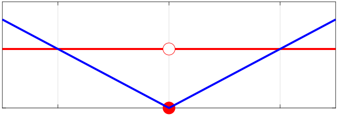
- Beaucoup de stratégies, qui fournissent la même solution
- Efficacité calculatoire dépendante du problème (taille, parcimonie, A)

Optimisation continue non différentiable

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

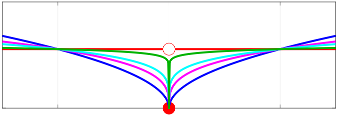
LE sujet chaud depuis 10 ans / algorithmes parcimonieux! [Bach *et al.* 12]

- Norme ℓ_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

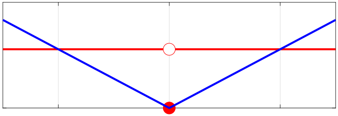


Optimisation continue non différentiable

$$\min_{\boldsymbol{x}} \tfrac{1}{2}\|\boldsymbol{y} - \mathrm{A}\boldsymbol{x}\|^2 + \lambda \sum_p \varphi(x_p), \ \varphi \text{ non différentiable en } 0$$

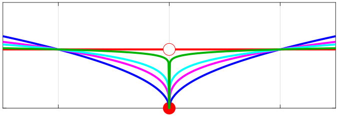
LE sujet chaud depuis 10 ans / algorithmes parcimonieux! [Bach *et al.* 12]

■ Norme ℓ_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*(algorithmes proximaux, ADMM, IRL1, . . .)
- Pas de garantie d'optimalité globale. . . mais fonction de coût plus proche de ℓ_0
- En pratique, résultats souvent meilleurs que ℓ_1. . .

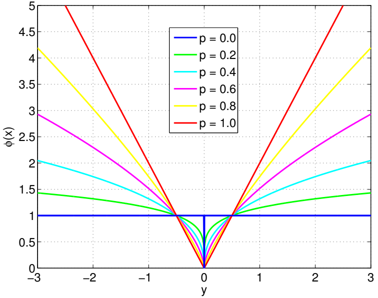
TD : Parcimonie, critères pénalisés et fonctions de seuil

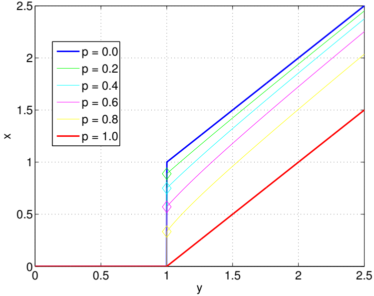
- Revenons au problème scalaire :

$$\min_x f(x), f(x) = (y-x)^2 + \lambda\varphi(x)$$

- φ induit de la parcimonie $\Leftrightarrow \varphi'(0) \neq 0$
- alors $\psi : y \xrightarrow{\psi} \widehat{x}$ est une fonction de seuillage

Fonctions de pénalisation et fonctions de seuillage associées pour $\varphi(x) = |x|^p$



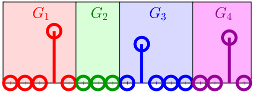


$\phi(x) = \|\boldsymbol{x}\|_0$ $\qquad$ $\psi(z) = z \quad$ si $|z| > \sqrt{\lambda}$, 0 sinon

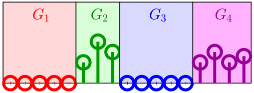
$\phi(x) = \|\boldsymbol{x}\|_1 = \sum_k |x_k|$ $\qquad$ $\psi(z) = z - \lambda/2$ si $|z| > \lambda/2$, 0 sinon

Généralisations : parcimonie structurée

- **Prise en compte de contraintes liant les différentes variables**
  - Parcimonie dans des groupes : $\|\boldsymbol{x}_{G_i}\|_0 \ \leq K_i$



  - Sélection parcimonieuse de groupes : $x_i = 0 \Leftrightarrow x_j \neq 0, j \in G_i$



ex : *Group LASSO*

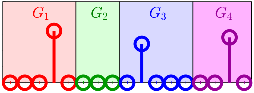
$$\min_{\boldsymbol{x}} \ \tfrac{1}{2}\|\boldsymbol{y} - \mathrm{A}\boldsymbol{x}\|^2 + \lambda \sum_j \|\boldsymbol{x}\|_{G_j}$$

  - ...

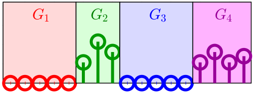
$\rightsquigarrow$ Généralisations des algorithmes gloutons et relaxations (normes mixtes) ...

Généralisations : parcimonie structurée

- **Prise en compte de contraintes liant les différentes variables**
  - Parcimonie dans des groupes : $\|\boldsymbol{x}_{G_i}\|_0 \leq K_i$



  - Sélection parcimonieuse de groupes : $x_i = 0 \Leftrightarrow x_j \neq 0, j \in G_i$



    ex : *Group LASSO*

$$\min_{\boldsymbol{x}} \; \tfrac{1}{2}\|\boldsymbol{y} - \mathrm{A}\boldsymbol{x}\|^2 + \lambda \sum_j \|\boldsymbol{x}\|_{G_j}$$

  - ...

$\rightsquigarrow$ Généralisations des algorithmes gloutons et relaxations (normes mixtes) ...
et des approches exactes en norme ℓ_0! [Thèse G. Samain]

Généralisations : apprentissage de dictionnaires

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

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

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

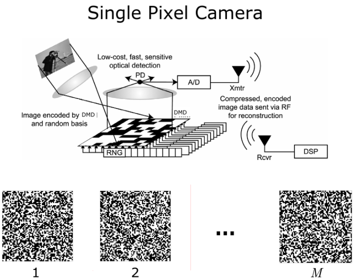
Applications $\uparrow\uparrow$ en statistiques / machine learning (*sparse coding*) . . .

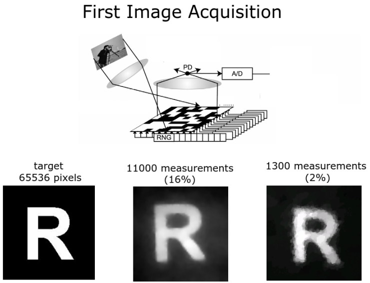
Application à l'acquisition comprimée (compressed sensing)

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

Application à l'acquisition comprimée (compressed sensing)

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





Cf. cours SIBIO!

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

Ressources I

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.
- ℓ_1 ou ℓ_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.

Ressources II

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