

Une version polyatomique de l'algorithme Frank-Wolfe pour résoudre le problème LASSO en grandes dimensions

Adrian Jarret, PhD student @EPFL/LCAV

en collaboration avec Matthieu Simeoni, Julien Fageot, Martin Vetterli

EPFL

GRETSI 2022

LCAV

Programme

- I. Quelques rappels sur le LASSO
- II. La version classique de l'algorithme Frank-Wolfe
- III. Frank-Wolfe Polyatomique:
une méthode efficace pour le problème LASSO

I. Le problème LASSO

Reconstruction linéaire avec hypothèse de parcimonie

$$\arg \min_{\mathbf{x} \in \mathbb{R}^N} \frac{1}{2} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1$$

$$\left[\begin{array}{l} \mathbf{y} \in \mathbb{R}^L \\ \mathbf{H} \in \mathbb{R}^{L \times N} \\ \lambda > 0 \end{array} \right.$$

[Tibshirani, 1996]

I. Le problème LASSO

Reconstruction linéaire avec hypothèse de parcimonie

$$\arg \min_{\mathbf{x} \in \mathbb{R}^N} \frac{1}{2} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1$$

$$\left[\begin{array}{l} \mathbf{y} \in \mathbb{R}^L \\ \mathbf{H} \in \mathbb{R}^{L \times N} \\ \lambda > 0 \end{array} \right.$$

[Tibshirani, 1996]

- Modèle: $\mathbf{y} = \mathbf{H}\mathbf{x}_0 + \mathbf{w}$

- Solution: $\mathbf{x}^* \approx \mathbf{x}_0$

I. Le problème LASSO

Reconstruction linéaire avec hypothèse de parcimonie

$$\arg \min_{\mathbf{x} \in \mathbb{R}^N} \frac{1}{2} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1$$

$$\left[\begin{array}{l} \mathbf{y} \in \mathbb{R}^L \\ \mathbf{H} \in \mathbb{R}^{L \times N} \\ \lambda > 0 \end{array} \right.$$

[Tibshirani, 1996]

- Model: $\mathbf{y} = \mathbf{H}\mathbf{x}_0 + \mathbf{w}$

- Solution: $\mathbf{H}\mathbf{x}^* \approx \mathbf{H}\mathbf{x}_0$

I. Le problème LASSO

Reconstruction linéaire avec hypothèse de parcimonie

$$\arg \min_{\mathbf{x} \in \mathbb{R}^N} \frac{1}{2} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1$$

$$\begin{cases} \mathbf{y} \in \mathbb{R}^L \\ \mathbf{H} \in \mathbb{R}^{L \times N} \\ \lambda > 0 \end{cases}$$

[Tibshirani, 1996]

Théorème de la représentation:

Les solutions forment un sous-ensemble non vide, convexe et compact de \mathbb{R}^N dont les points extrêmes sont au plus L-parcimonieux:

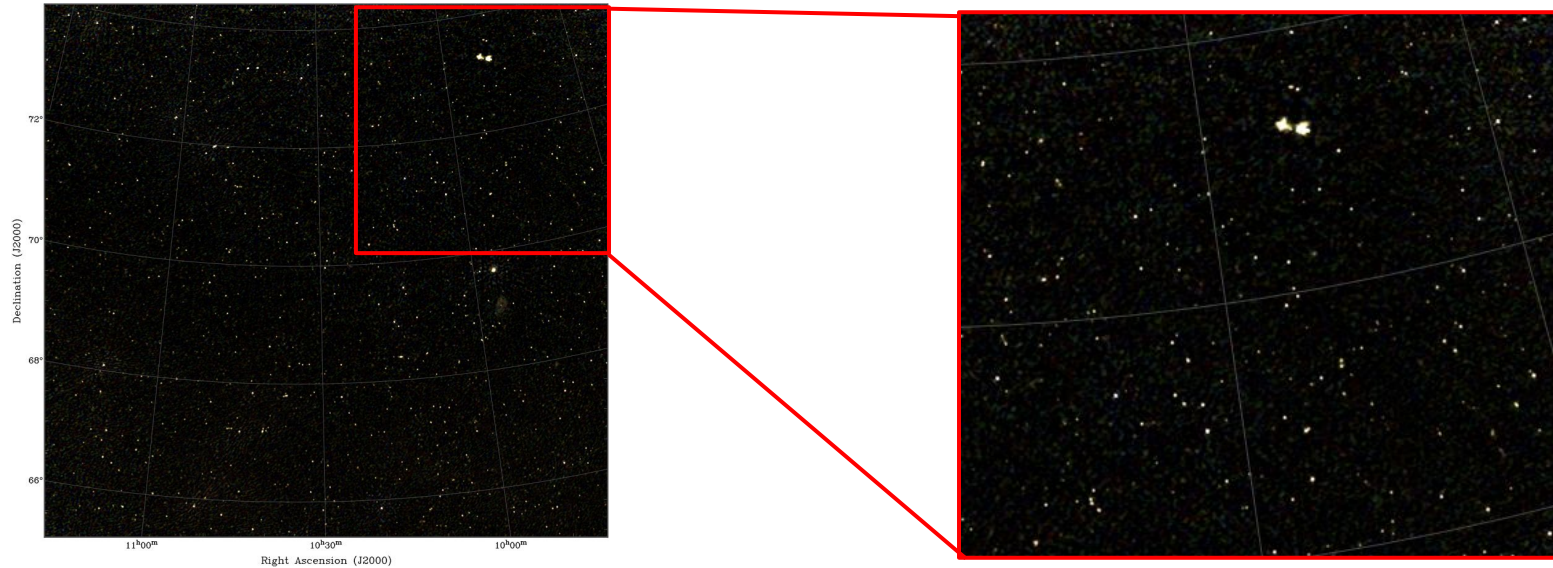
$$\mathbf{x}_{\text{sparse}} = \sum_{k=1}^L a_k \mathbf{e}_{n_k}$$

[Unser et al., 2016]

I. Le problème LASSO

Le LASSO en Radio Interférométrie: intérêts et limitations

[Image: C. Riseley, G. Gurkan, G. Heald and the MSSS team.]

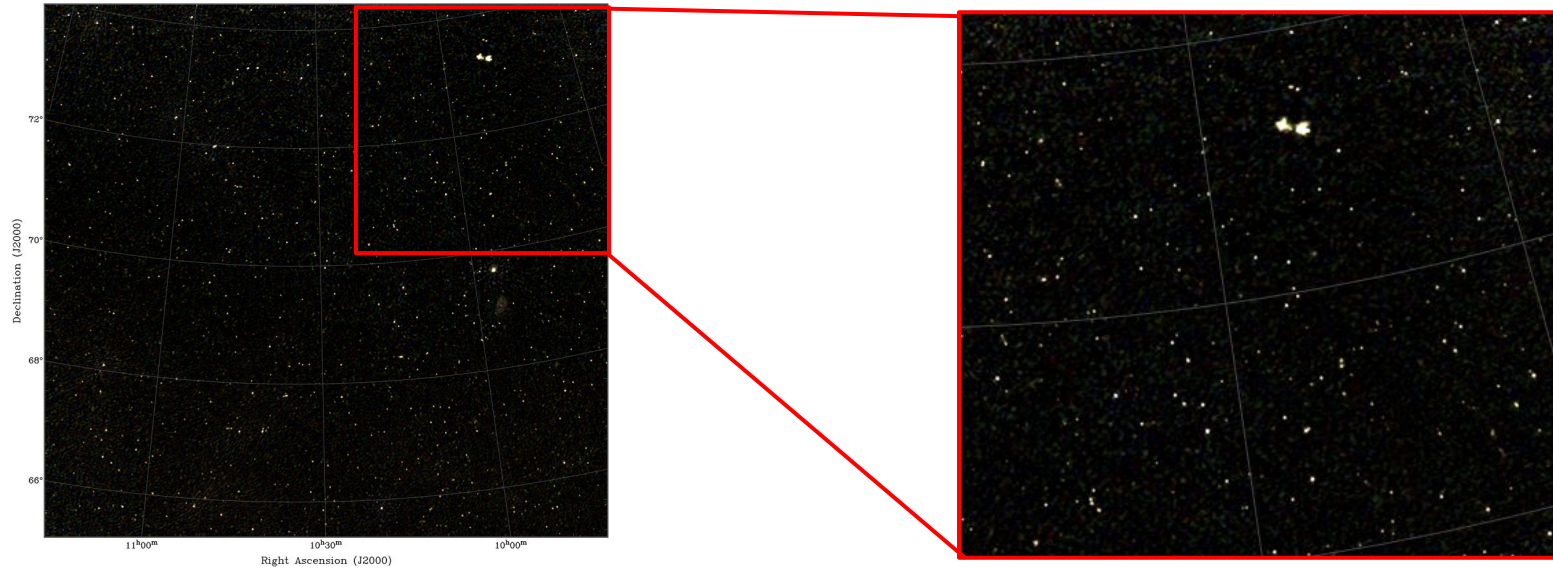


Application en RA:
Modèle de sources ponctuelles

I. Le problème LASSO

Le LASSO en Radio Interférométrie: intérêts et limitations

[Image: C. Riseley, G. Gurkan, G. Heald and the MSSS team.]



Application en RA:

Modèle de sources ponctuelles

Limitations:

Méthodes numériques coûteuses

II. L'algorithme Frank-Wolfe classique

Un algorithme d'optimisation convexe avec un comportement glouton

$$\min_{\mathbf{x} \in \mathcal{D}} f(\mathbf{x})$$

Algorithm 1: Vanilla Frank-Wolfe Algorithm (V-FW)

Initialize $\mathbf{x}_0 \in \mathcal{D}$

for $k = 1, 2 \dots$ **do**

1) Find an update direction:

$$\mathbf{s}_k \in \arg \min_{\mathbf{s} \in \mathcal{D}} \langle \nabla f(\mathbf{x}_k), \mathbf{s} \rangle$$

2.a) Step size: $\gamma_k \leftarrow \frac{2}{k+2}$

2.b) Reweight: $\mathbf{x}_{k+1} \leftarrow (1 - \gamma_k)\mathbf{x}_k + \gamma_k\mathbf{s}_k$

end

[Frank,Wolfe,1956]

II. L'algorithme Frank-Wolfe classique

Un algorithme d'optimisation convexe avec un comportement glouton

$$\min_{\mathbf{x} \in \mathcal{D}} f(\mathbf{x})$$

Algorithm 1: Vanilla Frank-Wolfe Algorithm (V-FW)

Initialize $\mathbf{x}_0 \in \mathcal{D}$

for $k = 1, 2 \dots$ **do**

1) Find an update direction:

$$\mathbf{s}_k \in \arg \min_{\mathbf{s} \in \mathcal{D}} \langle \nabla f(\mathbf{x}_k), \mathbf{s} \rangle$$

2.a) Step size: $\gamma_k \leftarrow \frac{2}{k+2}$

2.b) Reweight: $\mathbf{x}_{k+1} \leftarrow (1 - \gamma_k)\mathbf{x}_k + \gamma_k \mathbf{s}_k$

end

Création
d'atome

“Atom”

[Frank,Wolfe,1956]

II. L'algorithme Frank-Wolfe classique

Un algorithme d'optimisation convexe avec un comportement glouton

$$\min_{\mathbf{x} \in \mathcal{D}} f(\mathbf{x})$$

Algorithm 1: Vanilla Frank-Wolfe Algorithm (V-FW)

Initialize $\mathbf{x}_0 \in \mathcal{D}$

for $k = 1, 2 \dots$ **do**

1) Find an update direction:

$$\mathbf{s}_k \in \arg \min_{\mathbf{s} \in \mathcal{D}} \langle \nabla f(\mathbf{x}_k), \mathbf{s} \rangle$$

2.a) Step size: $\gamma_k \leftarrow \frac{2}{k+2}$

2.b) Reweight: $\mathbf{x}_{k+1} \leftarrow (1 - \gamma_k)\mathbf{x}_k + \gamma_k\mathbf{s}_k$

end

$$\mathbf{x}_{k+1} = \sum_{i=0}^k \alpha_{i,k} \mathbf{s}_i$$

Estimation
des poids

[Frank,Wolfe,1956]

II. The classical Frank-Wolfe algorithm

A convex optimization algorithm with a greedy behavior

$$\min_{\mathbf{x} \in \mathcal{D}} f(\mathbf{x})$$

Algorithm 1: Vanilla Frank-Wolfe Algorithm (V-FW)

Initialize $\mathbf{x}_0 \in \mathcal{D}$

for $k = 1, 2 \dots$ **do**

1) Find an update direction:

$$\mathbf{s}_k \in \arg \min_{\mathbf{s} \in \mathcal{D}} \langle \nabla f(\mathbf{x}_k), \mathbf{s} \rangle$$

2.a) Step size: $\gamma_k \leftarrow \frac{2}{k+2}$

2.b) Reweight: $\mathbf{x}_{k+1} \leftarrow (1 - \gamma_k)\mathbf{x}_k + \gamma_k \mathbf{s}_k$

end

Update
direction

Weights
estimation

“Atom”

$$\mathbf{x}_{k+1} = \sum_{i=0}^k \alpha_{i,k} \mathbf{s}_i$$

[Frank,Wolfe,1956]

II. L'algorithme Frank-Wolfe classique

Un exemple: minimisation sur un polygone

$$\min_{\mathbf{x} \in \mathcal{D}} f(\mathbf{x})$$

Algorithm 1: Vanilla Frank-Wolfe Algorithm (V-FW)

Initialize $\mathbf{x}_0 \in \mathcal{D}$

for $k = 1, 2 \dots$ **do**

1) Find an update direction:

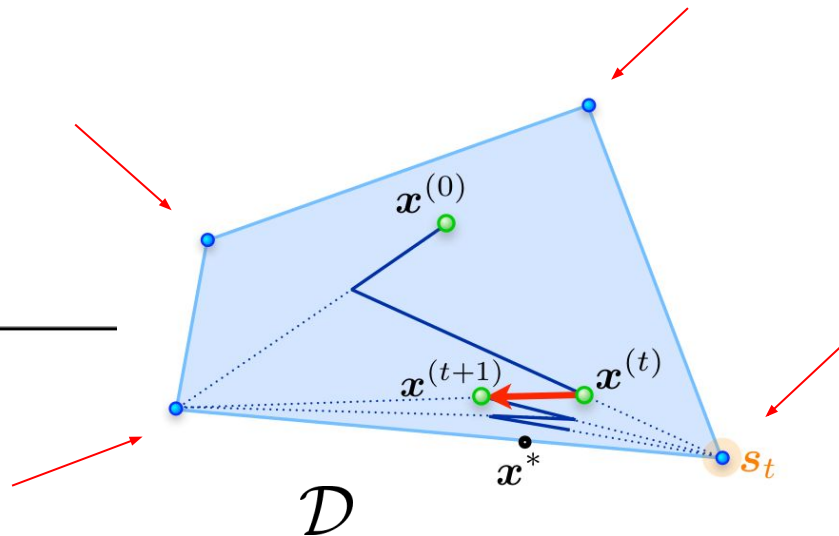
$$\mathbf{s}_k \in \arg \min_{\mathbf{s} \in \mathcal{D}} \langle \nabla f(\mathbf{x}_k), \mathbf{s} \rangle$$

2.a) Step size: $\gamma_k \leftarrow \frac{2}{k+2}$

2.b) Reweight: $\mathbf{x}_{k+1} \leftarrow (1 - \gamma_k)\mathbf{x}_k + \gamma_k\mathbf{s}_k$

end

Points extrêmes
parcimonieux



[Lacoste-Julien et al.,
2013]

II. L'algorithme Frank-Wolfe classique

Forme des atomes générés pour le LASSO

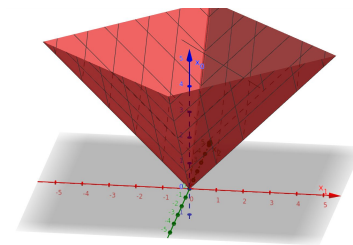
- Sous-problème linéaire:

$$\mathbf{s}_k \in \arg \min_{\mathbf{s} \in \mathcal{D}} \langle \nabla f(\mathbf{x}_k), \mathbf{s} \rangle$$

- Atomes identifiés par FW:

$$\mathbf{s}_k = \pm M \mathbf{e}_{i_k}$$

$$M > 0$$



II. L'algorithme Frank-Wolfe classique

Forme des atomes générés pour le LASSO

- Sous-problème linéaire:

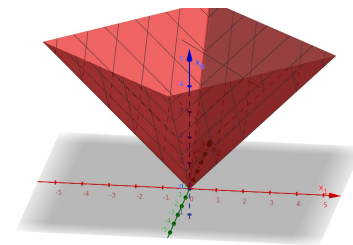
$$\mathbf{s}_k \in \arg \min_{\mathbf{s} \in \mathcal{D}} \langle \nabla f(\mathbf{x}_k), \mathbf{s} \rangle$$

- Atomes identifiés par FW:

$$\mathbf{s}_k = \pm M \mathbf{e}_{i_k}$$

$$M > 0$$

- En résumé: 1 atome/itération, redondance, oscillations
évaluation sous-optimale des poids, ...



II. The classical Frank-Wolfe algorithm

Shape of the atoms generated by the LASSO

- Linear subproblem:

$$\mathbf{s}_k = \arg \max_{\mathbf{s} \in \mathcal{D}} \langle \boldsymbol{\eta}_k, \mathbf{s} \rangle$$

- Atoms identified by FW:

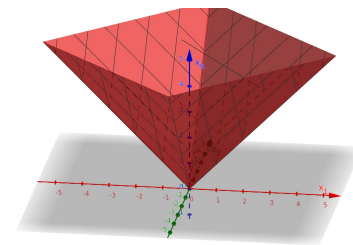
$$\mathbf{s}_k = \pm M \mathbf{e}_{i_k}$$

$$M > 0$$

- Bottom line: 1 atom/iteration, redundancy, oscillations
suboptimal estimation of the weights, ...

Empirical dual
certificate:

$$\boldsymbol{\eta}_k = \frac{1}{\lambda} \mathbf{G}^T (\mathbf{y} - \mathbf{G} \mathbf{x}_k) \in \mathbb{R}^N$$



III. Frank-Wolfe Polyatomique pour le LASSO

Comment accélérer FW ?

[J., Fageot, Simeoni, 2021]

FW Classique

FW Polyatomique



III. Frank-Wolfe Polyatomique pour le LASSO

Comment accélérer FW ?

[J., Fageot, Simeoni, 2021]

FW Classique

1. Seulement un atome par itération

$$\mathbf{s}_k$$

FW Polyatomique

1. Autoriser plusieurs atomes

$$\mathcal{I}_k = \{\mathbf{s}_1^{(k)}, \mathbf{s}_2^{(k)}, \dots\}$$

III. Frank-Wolfe Polyatomique pour le LASSO

Comment accélérer FW ?

[J., Fageot, Simeoni, 2021]

FW Classique

1. Seulement un atome par itération

$$\mathbf{s}_k$$

2. Combinaison convexe des poids

FW Polyatomique

1. Autoriser plusieurs atomes

$$\mathcal{I}_k = \{\mathbf{s}_1^{(k)}, \mathbf{s}_2^{(k)}, \dots\}$$

2. Ré-estimation de tous les poids

III. Frank-Wolfe Polyatomique pour le LASSO

L'algorithme

[J., Fageot, Simeoni, 2021]

Algorithm 2: Polyatomic FW (P-FW) of quality $\delta > 0$

Initialize: $\mathbf{x}_0 \leftarrow 0, \mathcal{S}_0 \leftarrow \emptyset$

for $k = 1, 2 \dots$ **do**

$\gamma_k \leftarrow 2/(k+2)$

1".a) Polyatomic exploration:

$$\mathcal{I}_k = \{1 \leq j \leq N : |\boldsymbol{\eta}_k|_j \geq \|\boldsymbol{\eta}_k\|_\infty - \delta\gamma_k\}$$

$$\mathbf{s}_k \leftarrow (\sum_{i \in \mathcal{I}_k} \mathbf{e}_i) / \text{Card}(\mathcal{I}_k)$$

1".b) Update active indices: $\mathcal{S}_k \leftarrow \mathcal{S}_{k-1} \cup \mathcal{I}_k$

2".a) Set accuracy threshold: $\varepsilon_k = \varepsilon_0 \gamma_k$


2".b) Update active weights:

$$\mathbf{x}_{k+1/2} \leftarrow (1 - \gamma_k)\mathbf{x}_k + \gamma_k \mathbf{s}_k$$

$$\mathbf{x}_{k+1} \leftarrow \text{partial_correction}(\mathbf{x}_{k+1/2}, \mathcal{S}_k, \varepsilon_k)$$

$k \leftarrow k + 1$

end

$$\boldsymbol{\eta}_k = \frac{1}{\lambda} \mathbf{G}^T(\mathbf{y} - \mathbf{G}\mathbf{x}_k) \in \mathbb{R}^N$$


III. Frank-Wolfe Polyatomique pour le LASSO

L'algorithme

[J., Fageot, Simeoni, 2021]

Algorithm 2: Polyatomic FW (P-FW) of quality $\delta > 0$

Initialize: $\mathbf{x}_0 \leftarrow 0, \mathcal{S}_0 \leftarrow \emptyset$

for $k = 1, 2 \dots$ **do**

$\gamma_k \leftarrow 2/(k+2)$

1".a) Polyatomic exploration:

$$\mathcal{I}_k = \{1 \leq j \leq N : |\boldsymbol{\eta}_k|_j \geq \|\boldsymbol{\eta}_k\|_\infty - \delta\gamma_k\}$$

1".b) Update active indices: $\mathcal{S}_k \leftarrow \mathcal{S}_{k-1} \cup \mathcal{I}_k$

2".a) Set accuracy threshold: $\varepsilon_k = \varepsilon_0\gamma_k$

2".b) Update active weights:

$$\mathbf{x}_{k+1/2} \leftarrow (1 - \gamma_k)\mathbf{x}_k + \gamma_k\mathbf{s}_k$$

$$\mathbf{x}_{k+1} \leftarrow \text{partial_correction}(\mathbf{x}_{k+1/2}, \mathcal{S}_k, \varepsilon_k)$$

$k \leftarrow k + 1$

end

$$\boldsymbol{\eta}_k = \frac{1}{\lambda} \mathbf{G}^T(\mathbf{y} - \mathbf{G}\mathbf{x}_k) \in \mathbb{R}^N$$

Estimation
des atomes

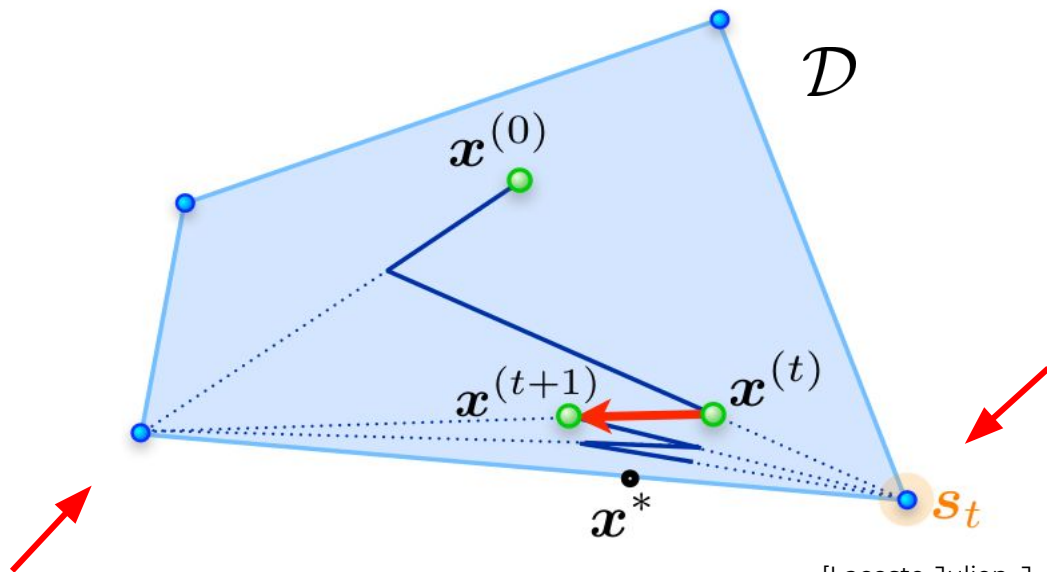
III. Frank-Wolfe Polyatomique pour le LASSO

Choix multiple d'atomes

[J., Fageot, Simeoni, 2021]

$$\mathcal{I}_k : \quad \langle \nabla f(\mathbf{x}_k), \mathbf{s}_k \rangle \in \left[\min_{\mathbf{s} \in \mathcal{D}} \langle \nabla f(\mathbf{x}_k), \mathbf{s} \rangle, \min_{\mathbf{s} \in \mathcal{D}} \langle \nabla f(\mathbf{x}_k), \mathbf{s} \rangle + \delta \gamma_k \right]$$

$$\mathcal{S}_k \leftarrow \mathcal{S}_{k-1} \cup \mathcal{I}_k$$



[Lacoste-Julien, J., 2013]

III. Frank-Wolfe Polyatomique pour le LASSO

Une nouvelle stratégie d'estimation des poids

[J., Fageot, Simeoni, 202

1]

Algorithm 2: Polyatomic FW (P-FW) of quality $\delta > 0$

Initialize: $\mathbf{x}_0 \leftarrow 0, \mathcal{S}_0 \leftarrow \emptyset$

for $k = 1, 2 \dots$ **do**

$\gamma_k \leftarrow 2/(k+2)$

1".a) Polyatomic exploration:

$$\mathcal{I}_k = \{1 \leq j \leq N : |\boldsymbol{\eta}_k|_j \geq \|\boldsymbol{\eta}_k\|_\infty - \delta\gamma_k\}$$

$$\mathbf{s}_k \leftarrow (\sum_{i \in \mathcal{I}_k} \mathbf{e}_i) / \text{Card}(\mathcal{I}_k)$$

1".b) Update active indices: $\mathcal{S}_k \leftarrow \mathcal{S}_{k-1} \cup \mathcal{I}_k$

2".a) Set accuracy threshold: $\varepsilon_k = \varepsilon_0 \gamma_k$

2".b) Update active weights:

$$\mathbf{x}_{k+1} \leftarrow \text{partial_correction}(\mathbf{x}_{k+1/2}, \mathcal{S}_k, \varepsilon_k)$$

$k \leftarrow k + 1$

end

$$\boldsymbol{\eta}_k = \frac{1}{\lambda} \mathbf{G}^T (\mathbf{y} - \mathbf{G}\mathbf{x}_k) \in \mathbb{R}^N$$

Correction
partielle

III. Frank-Wolfe Polyatomique pour le LASSO

Correction partielle des poids

- Ensemble des indices actifs:

$$\mathcal{S}_k \subset \{1, \dots, N\}, \quad |\mathcal{S}_k| \ll N$$

Réduction de
la dimension

- “Correction complète”:

$$\mathbf{x}_{k+1} \leftarrow \arg \min_{\text{Supp}(\mathbf{x}) \subset \mathcal{S}_k} \frac{1}{2} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1$$

LASSO à
support réduit

[Jaggi, 2013]

III. Frank-Wolfe Polyatomique pour le LASSO

Correction partielle des poids

- Ensemble des indices actifs:

$$\mathcal{S}_k \subset \{1, \dots, N\}, \quad |\mathcal{S}_k| \ll N$$

Réduction de la dimension

- “Correction complète”:

$$\mathbf{x}_{k+1} \leftarrow \arg \min_{\text{Supp}(\mathbf{x}) \subset \mathcal{S}_k} \frac{1}{2} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1$$

LASSO à support réduit

[Jaggi, 2013]

- Correction partielle:
 - Arrêt prématuré
 - Précision adaptative

Deuxième contribution de FW Polyatomique

III. Frank-Wolfe Polyatomique pour le LASSO

Garantie de convergence

Théorème 1: (*Convergence de FW Polyatomique*)

- Soit :
$$\left\{ \begin{array}{l} f(\mathbf{x}) = \frac{1}{2}\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1 \\ \mathbf{x}^* \in \arg \min f(\mathbf{x}) \end{array} \right.$$

- Alors :

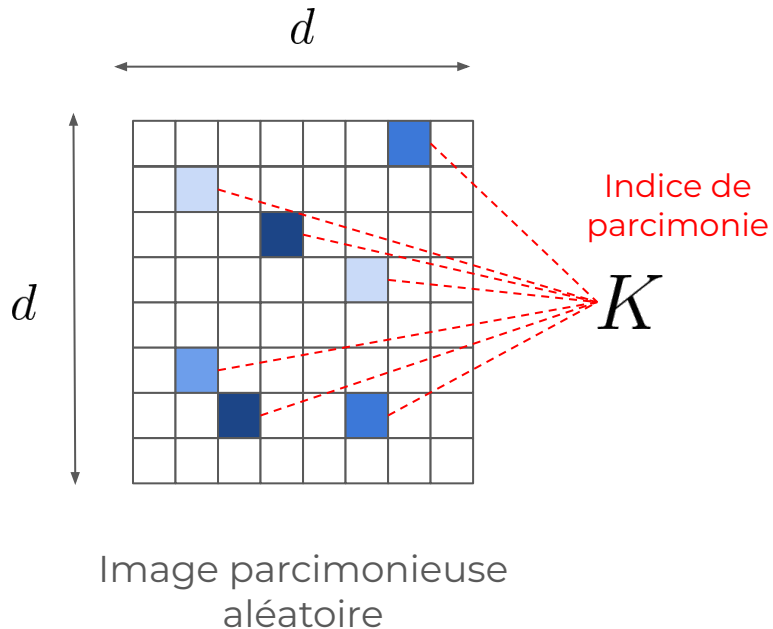
$$f(\mathbf{x}_k) - f(\mathbf{x}^*) \leq \frac{2}{k+2}(C_f + 2\delta)$$

[Jaggi, 2013]

[J., Fageot, Simeoni, 20

III. Frank-Wolfe Polyatomique pour le LASSO

Simulations: contexte expérimental



- Modèle des mesures:

$$y = \mathbf{H}\mathbf{x}_0 + \mathbf{w}$$

$$\mathbf{H} \in \mathbb{R}^{L \times N}$$

$$N = d^2 \quad L = \alpha K$$

III. Frank-Wolfe Polyatomique pour le LASSO

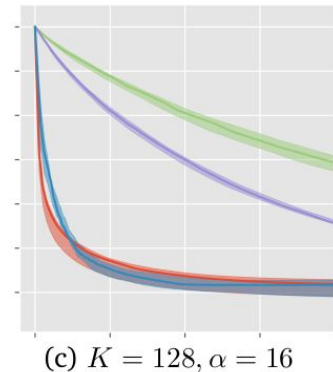
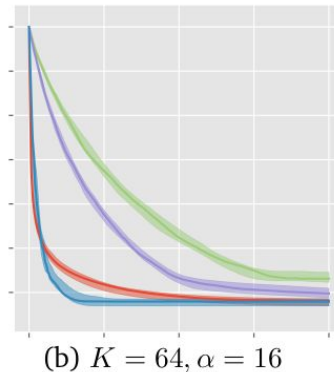
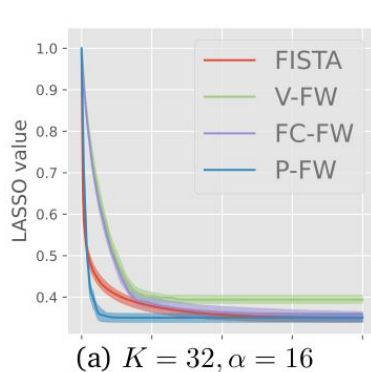
Simulations: acquisition comprimée

$$y = \mathbf{H}x_0 + w$$
$$\mathbf{H} \in \mathbb{R}^{L \times N}$$

[J., Fageot, Simeoni, 20

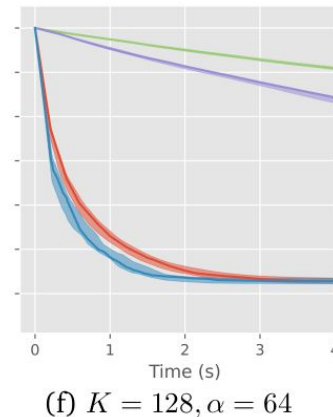
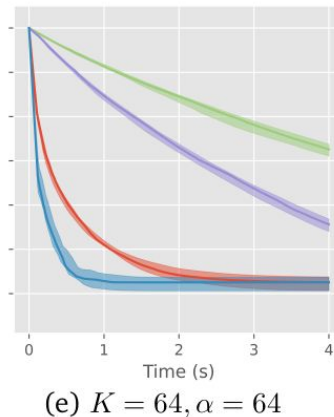
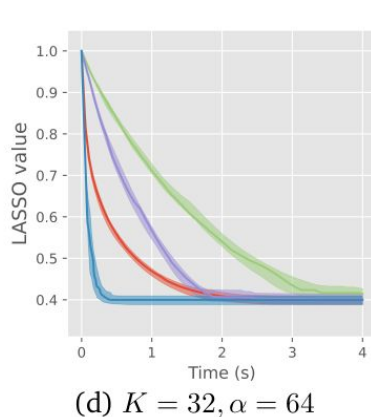
$$L = \alpha K$$

$$\alpha = 16$$



21]

$$\alpha = 64$$



$$N = 128^2$$

III. Polyatomic Frank-Wolfe for the LASSO

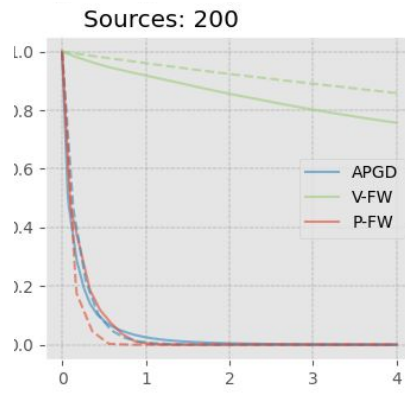
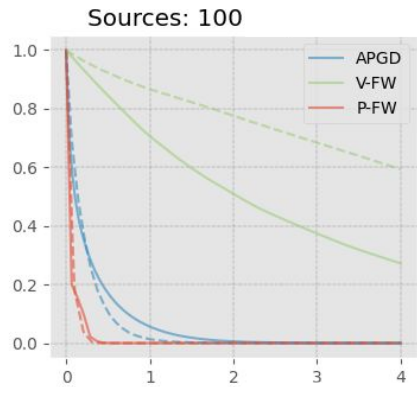
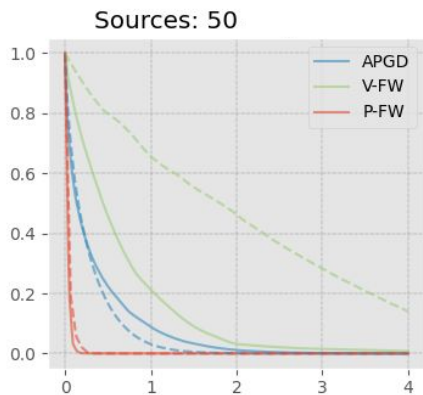
Simulations: coefficients de Fourier aléatoires

$$y = \mathbf{H}\mathbf{x}_0 + \mathbf{w}$$

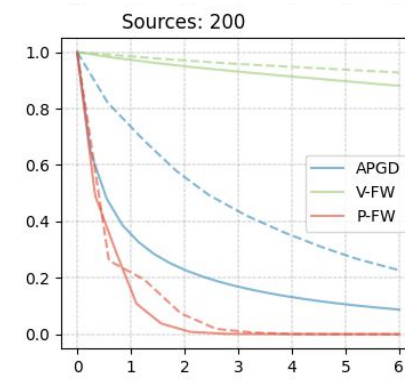
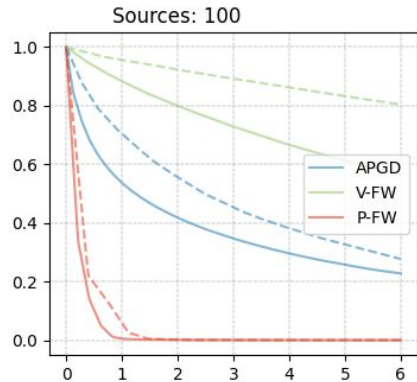
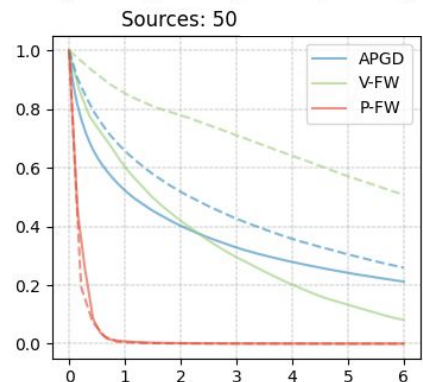
$$\mathbf{H} \in \mathbb{R}^{L \times N}$$

[J., Fageot, Simeoni, 20

$N = 101^2$



$N = 201^2$

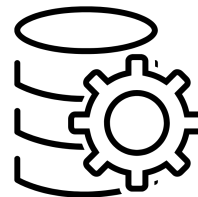
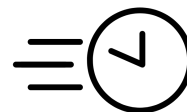


Adrian Jarret

Ce qu'il faut retenir

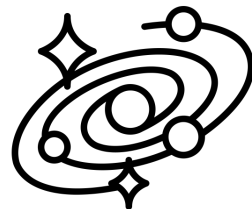
- Frank-Wolfe Polyatomique est **rapide**
 - Exploration efficace
 - Repérage rapide des degrés de liberté actifs
 - Vitesse liée à la parcimonie

- Frank-Wolfe Polyatomique est **extensible**
 - Itérées parcimonieuses
 - Retrait des atomes inutiles en cours d'exécution
 - Application en grandes dimensions



Further works

- Apply Polyatomic FW to **radio astronomy problems**
 - Simulated data, higher dimensions (ongoing)
 - Real-world data (SKA)



- Extend the algorithm to reconstruct **continuous-domain data** (ongoing)
 - Sparse atoms = **Dirac impulses**
 - Difficulty of estimating their location

$$\mathcal{F}(\mathbf{y}, \Phi s) + \mathcal{R}(s)$$

Appendix: Epigraphical lift

Differentiable LASSO problem

$$\arg \min_{\mathbf{x} \in \mathbb{R}^N} \frac{1}{2} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1$$



$$\arg \min_{(t, \mathbf{x}) \in \mathcal{D}} \frac{1}{2} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \lambda t$$

$$\mathcal{D} = \{t, \mathbf{x} \in \mathbb{R}_+ \times \mathbb{R}^N : \|\mathbf{x}\|_1 \leq t \leq M\}$$

$$\begin{cases} \mathbf{y} \in \mathbb{R}^L \\ \mathbf{H} \in \mathbb{R}^{L \times N} \\ \lambda > 0 \end{cases}$$

$$M = \frac{\|\mathbf{y}\|_2^2}{2\lambda}$$

[Harchaoui et al.,
2013]