L'algorithme de Frank-Wolfe: une version polyatomique pour la résolution rapide du problème LASSO

Adrian Jarret, PhD student @EPFL/LCAV

en collaboration avec Matthieu Simeoni, Julien Fageot, Martin Vetterli

EPFL

CANUM 2020

LCAV

Programme

I. Quelques mots sur le LASSO

II. L'algorithme de Frank-Wolfe: version classique

III. Frank-Wolfe Polyatomique:

une méthode efficace pour le problème LASSO

I. Rappels sur le problème LASSO

Reconstruction linéaire et hypothèse de parcimonie

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

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

[Tibshirani, 1996]

I. Rappels sur le problème LASSO

Reconstruction linéaire et hypothèse de parcimonie

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

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

[Tibshirani, 1996]

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

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

I. Rappels sur le problème LASSO

Reconstruction linéaire et hypothèse de parcimonie

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

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

[Tibshirani, 1996]

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

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

I. Rappels sur le problème LASSO

Reconstruction linéaire et hypothèse de parcimonie

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

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

[Tibshirani, 1996]

Théorème de représentation:

L'ensemble des solutions est un sous-ensemble non vide, convexe et compact de $\mathbb{R}^N$, dont les points extrêmes sont au plus L-sparse:

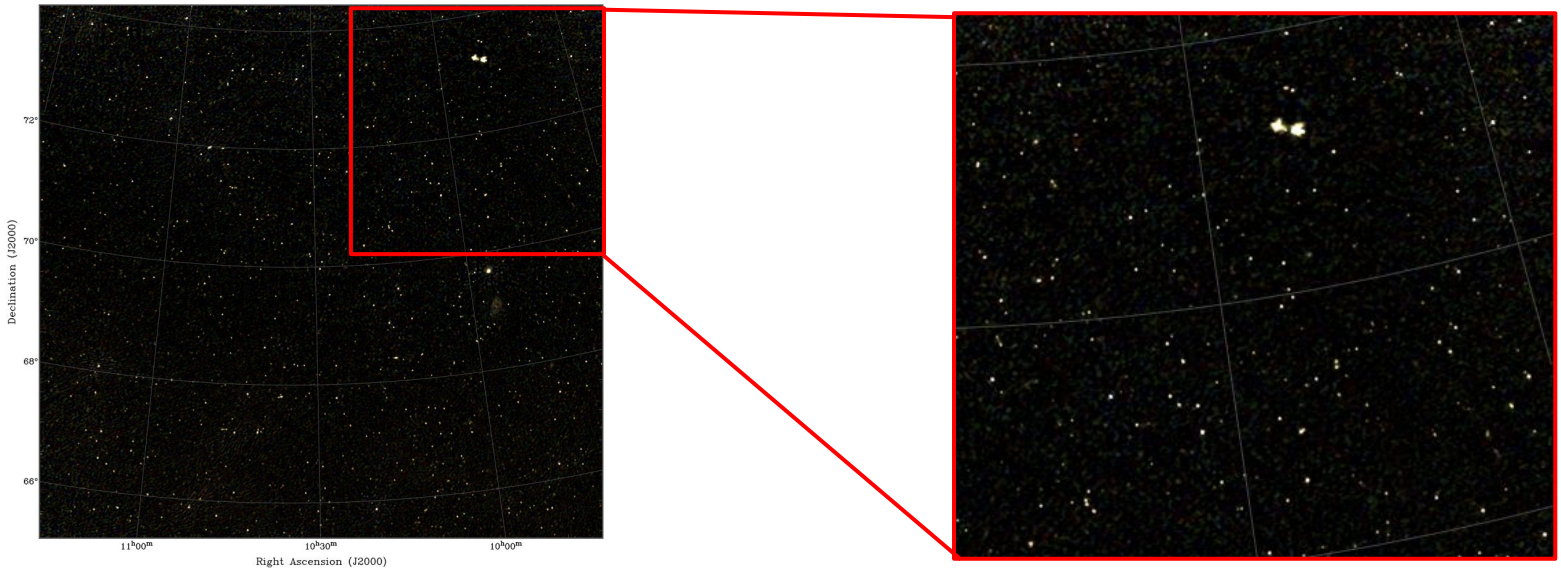
$$\mathbf{x}_{\text{sparse}} = \sum_{k=1}^{L} a_k \mathbf{e}_{n_k}$$

[Unser et al., 2016]

I. Rappels sur le problème LASSO

Le LASSO en radio interférométrie: intérêts et limitations

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

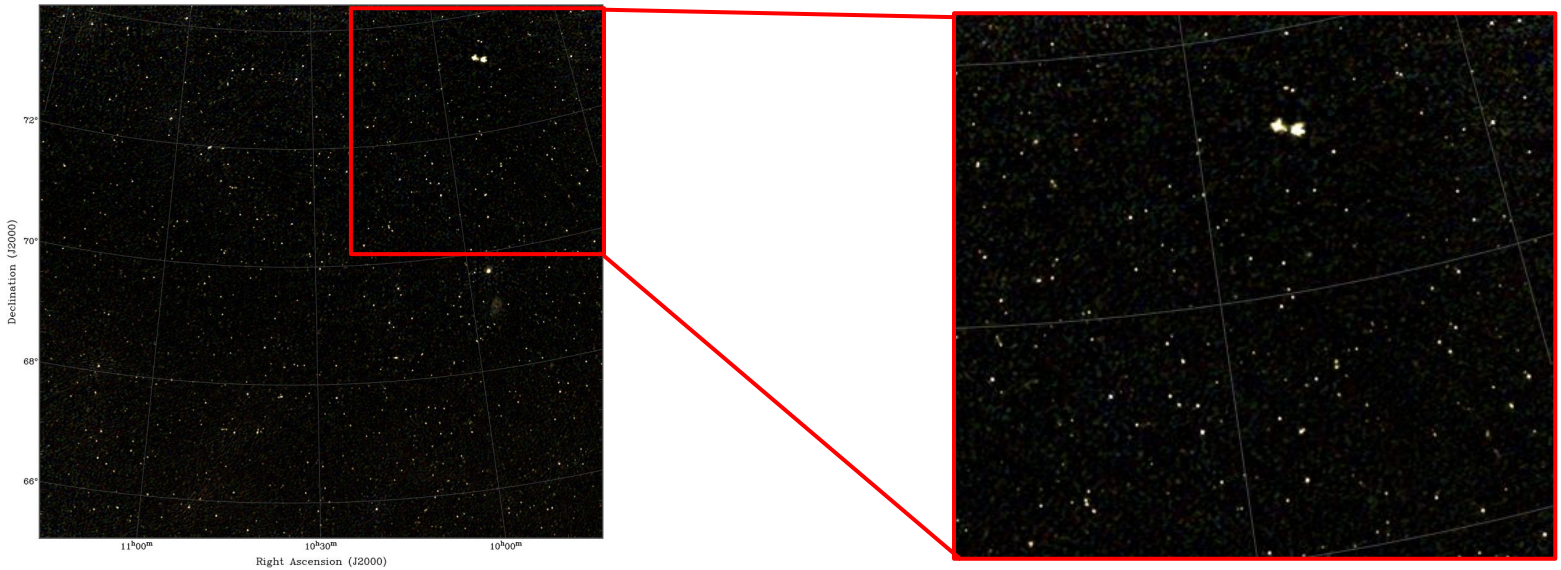


Application en RA:
Modélisation de sources ponctuelles

I. Rappels sur 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:
Modélisation de sources ponctuelles

Limitations:
Méthodes usuelles numériquement coûteuses

II. L'algorithme Frank-Wolfe classique

Un algorithme d'optimisation convexe au comportement glouton

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

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

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

for $k = 1, 2 \cdots$ **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 au comportement glouton

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

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

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

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

Direction de recherche

1) Find an update direction:

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

"Atome"

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 au comportement glouton

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

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

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

for $k = 1, 2 \cdots$ **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

Estimation des poids

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

[Frank,Wolfe,1956]

II. L'algorithme Frank-Wolfe classique

Un exemple illustré

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

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

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

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

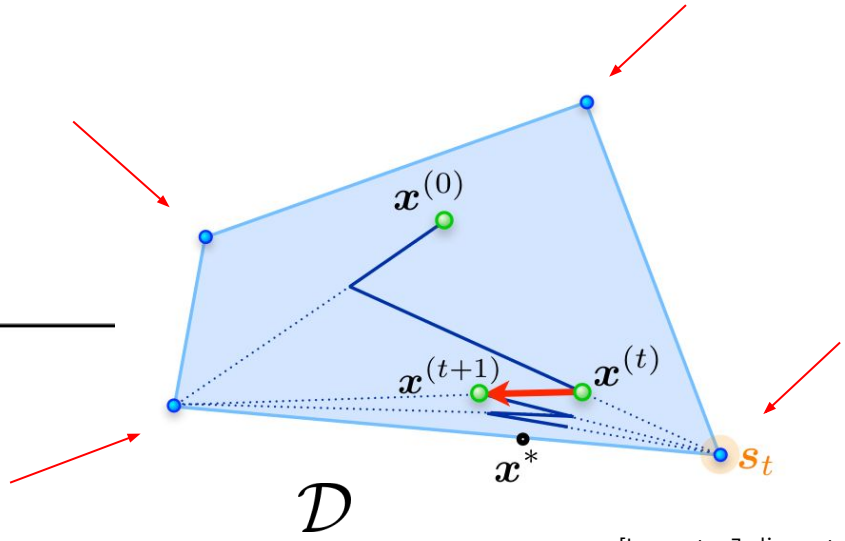
1) Find an update direction:

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

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



[Lacoste-Julien et al., 2013]

II. L'algorithme Frank-Wolfe classique

Les atomes générés par le LASSO

Certificat dual empirique:

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

- Sous-problème linéaire:

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

- Atomes identifiés par FW:

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

$$M > 0$$

II. L'algorithme Frank-Wolfe classique

Les atomes générés par le LASSO

Certificat dual empirique:

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

- Sous-problème linéaire:

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

- Atomes identifiés par FW:

$$\mathbf{s}_k = \pm M \mathbf{e}_{i_k} \qquad M > 0$$

- Bilan: 1 seul atome/itération, redondance, oscillations
  évaluation des poids sous-optimale, ...

III. Frank-Wolfe Polyatomique pour le LASSO

Ou comment accélérer FW

[**J.**,Fageot,Simeoni,2021]

Vanilla FW

FW Polyatomique

III. Frank-Wolfe Polyatomique pour le LASSO

Ou comment accélérer FW

[**J.**,Fageot,Simeoni,2021]

Vanilla FW

1. Un seul 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

Ou comment accélérer FW

[**J.**,Fageot,Simeoni,2021]

Vanilla FW

1. Un seul 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-évaluation 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 $k = 1, 2 \cdots$ **do**

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

1".a) Polyatomic exploration:

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

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

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:

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

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

$k \leftarrow k+1$

end

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 $k = 1, 2 \cdots$ **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:

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

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

$k \leftarrow k+1$

end

Estimation des atomes

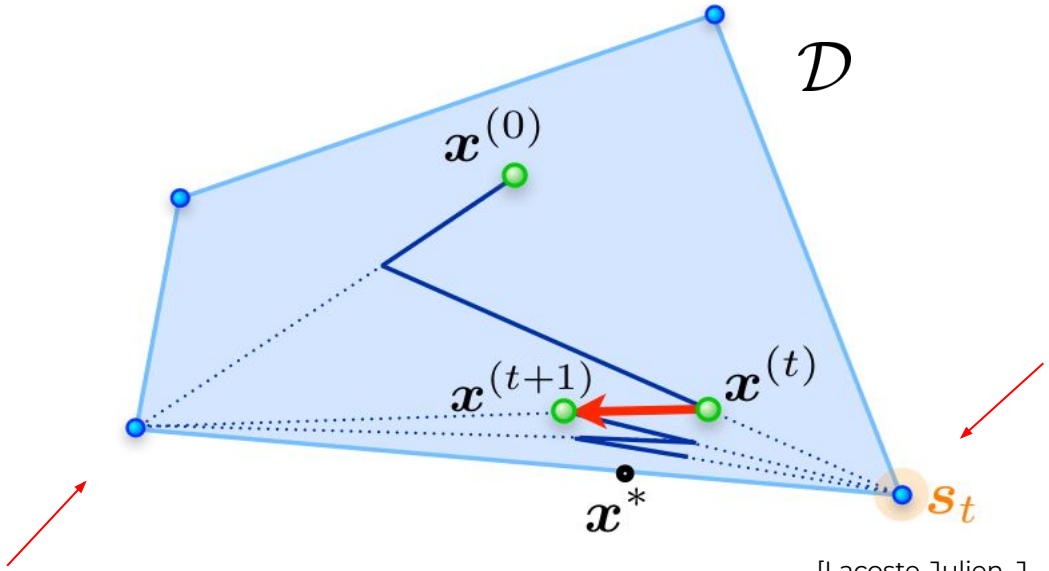
III. Frank-Wolfe Polyatomique pour le LASSO

Mise à jour polyatomique

[**J.**,Fageot,Simeoni,2021]

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

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



[Lacoste-Julien, J, 2013]

III. Frank-Wolfe Polyatomique pour le LASSO

Une nouvelle stratégie d'évaluation des poids

[**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 $k = 1, 2 \cdots$ **do**

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

1".a) Polyatomic exploration:

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

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

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:

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

$k \leftarrow k + 1$

end

Correction partielle des poids

III. Frank-Wolfe Polyatomique pour le LASSO

Correction partielle des poids

- Ensemble des indices actifs:

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

Réduction de la dimension

- Correction complète des poids:

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

LASSO à support réduit

[Jaggi, 2013]

III. Frank-Wolfe Polyatomique pour le LASSO

Correction partielle des poids

- Ensemble des indices actifs:

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

Réduction de la dimension

- Correction complète des poids:

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

LASSO à support réduit

[Jaggi, 2013]

- Correction partielle:
  - Critère d'arrêt anticipé
  - Précision croissante adaptée

Deuxième nouveauté de FW Polyatomique

[**J.**,Fageot,Simeoni, 2021]

III. Frank-Wolfe Polyatomique pour le LASSO

Garantie de convergence

Théorème 1: *(Convergence de Frank-Wolfe Polyatomique)*

- Soit :

$$\begin{cases} f(\mathbf{x}) = \frac{1}{2}\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1 \\ \mathbf{x}^* \in \arg\min f(\mathbf{x}) \end{cases}$$

- Alors :

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

[Jaggi, 2013]

[**J.**,Fageot,Simeoni,2021]

III. Frank-Wolfe Polyatomique pour le LASSO

Résultats de simulations: acquisition comprimée

[**J.**,Fageot,Simeoni,2021]

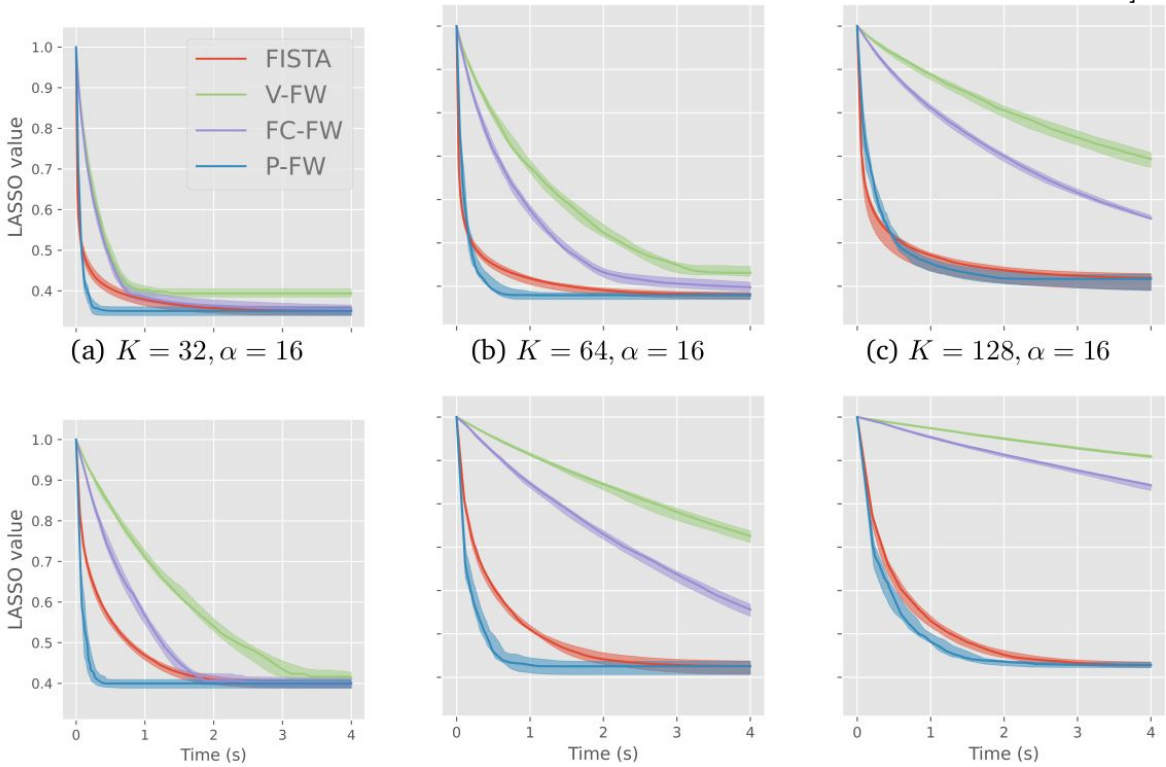
$$\mathbf{y} = \mathbf{H}\mathbf{x}_0 + \mathbf{w}$$

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

$$L = \alpha K$$

$$N = 128^2$$

$$\approx 1.6 \times 10^4$$



(a) $K = 32, \alpha = 16$ (b) $K = 64, \alpha = 16$ (c) $K = 128, \alpha = 16$

(d) $K = 32, \alpha = 64$ (e) $K = 64, \alpha = 64$ (f) $K = 128, \alpha = 64$

III. Frank-Wolfe Polyatomique pour le LASSO

Résultats de simulations: mesures de Fourier aléatoires

[**J.**,Fageot,Simeoni,20

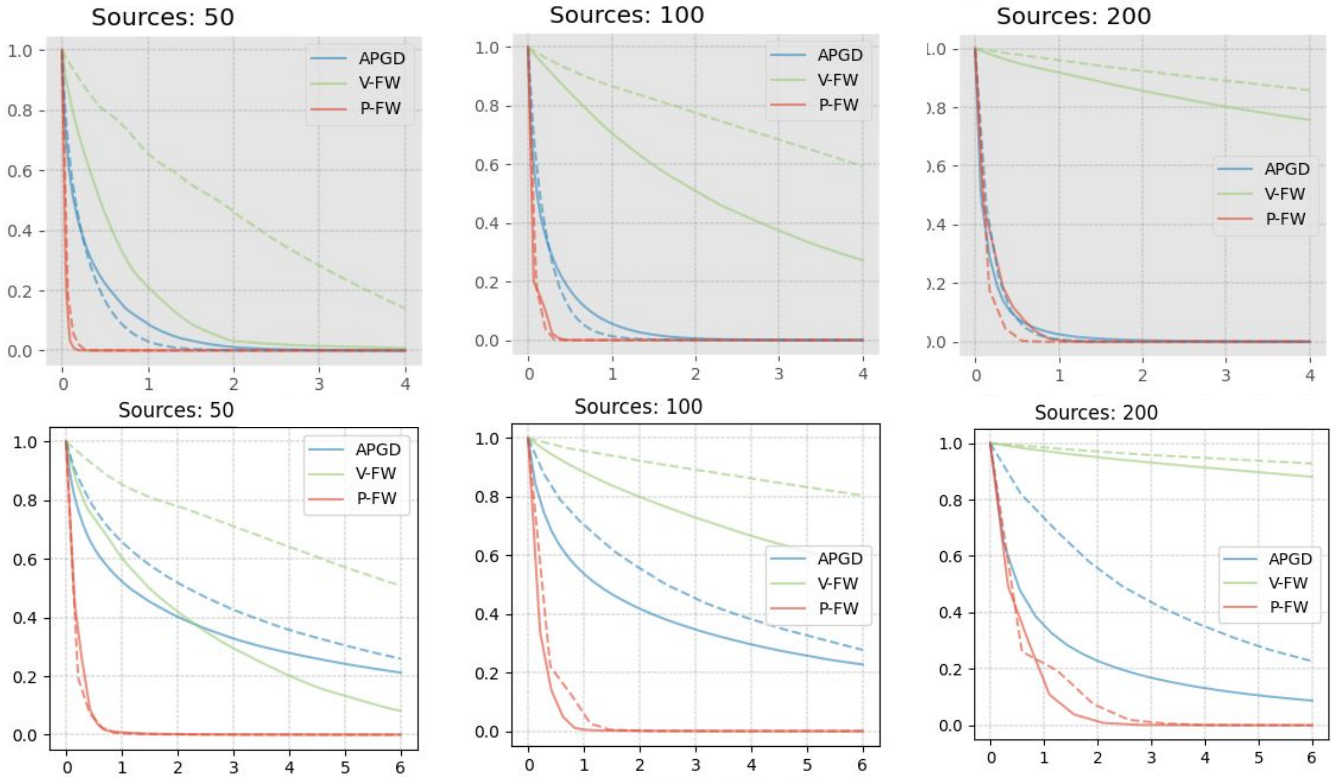
$$\mathbf{y} = \mathbf{H}\mathbf{x}_0 + \mathbf{w}$$

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

$$L = \alpha K$$

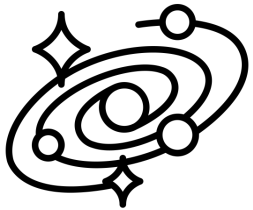
$$N = 101^2$$

$$N = 201^2$$



Perspectives

- Appliquer FW Polyatomique à des problèmes de RA
  - Données simulées, grandes dimensions
  - Mesures réelles (SKA)



- Étendre l'algorithme à la reconstruction de signaux continus
  - Reconstruction de Diracs
  - Comment identifier les nouveaux atomes ?

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

Appendix: Epigraphical lift

Differentiable LASSO problem

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

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

$$\Updownarrow$$

$$\boxed{\begin{array}{c} \arg\min_{(t,\mathbf{X})\in\mathcal{D}} \quad \frac{1}{2}\|\mathbf{y}-\mathbf{H}\mathbf{x}\|_2^2 + \lambda t \\ \mathcal{D} = \left\{t, \mathbf{x} \in \mathbb{R}_+ \times \mathbb{R}^N : \|\mathbf{x}\|_1 \leq t \leq M\right\} \end{array}}$$

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

[Harchaoui et al., 2013]