Annexe B

Algorithme de Gauss-Seidel pour optimisation d'un critère quadratique

Lorsqu'on cherche à optimiser un critère d'une manière itérative, il existe deux approches:

- les méthodes qui mettent à jour l'ensemble des variables à chaque itérations, comme par exemples des méthodes du type gradient ; et
- les méthodes qui mettent à jour seulement une variable à chaque itérations.

L'objet de cette annexe est étudier la méthode de Gausse-Seidel qui est une méthode de la deuxième approche pour le cas du critère MC ou celui d'un critère régularisé.

B.1 Introduction

Considérons le critère

$$J(\boldsymbol{x}) = \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|^2$$

et son gradient :

$$\nabla J = -2\boldsymbol{A}^t(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x})$$

Rappelons le principe d'un algorithme du gradient pour calculer la solution :

$$\begin{aligned}
\boldsymbol{x}^{(0)} &= \mathbf{0} \\
\boldsymbol{x}^{(k+1)} &= \boldsymbol{x}^{(k)} + \alpha^{(k)} \nabla J\left(\boldsymbol{x}^{(k)}\right) \\
&= \boldsymbol{x}^{(k)} + \alpha^{(k)} \boldsymbol{A}^t\left(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}^{(k)}\right) = \boldsymbol{x}^{(k)} + \alpha^{(k)}\boldsymbol{A}^t\boldsymbol{y} - \alpha^{(k)}\boldsymbol{A}^t\boldsymbol{A}\boldsymbol{x}^{(k)}
\end{aligned}$$

Supposons maintenant, qu'à l'itération k, on voudrait remettre à jour seulement la variable x_k et écrivons alors le critère en fonction de cette seule variable :

$$\begin{aligned}
J(x_k) &= \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|^2 \\
&= \sum_{i=1}^{M}\left(y_i - \sum_{j=1}^{N} a_{ij}x_j\right)^2 \\
&= \sum_{i=1}^{M}\left(y_i - \sum_{j\neq k}^{N} a_{ij}x_j - a_{ik}x_k\right)^2 \\
&= \sum_{i=1}^{M}\left[\left(y_i - \sum_{j\neq k}^{N} a_{ij}x_j\right)^2 - 2a_{ik}\left(y_i - \sum_{j\neq k}^{N} a_{ij}x_j\right)x_k + a_{ik}^2 x_k^2\right] \\
\frac{\partial J}{\partial x_k} &= -2\sum_{i=1}^{M} a_{ik}\left(y_i - \sum_{j=1}^{N} a_{ij}x_j\right) \\
&= -2\sum_{i=1}^{M} a_{ik}\left(y_i - \sum_{j\neq k}^{N} a_{ij}x_j - a_{ik}x_k\right) \\
&= -2\sum_{i=1}^{M}\left[a_{ik}\left(y_i - \sum_{j\neq k}^{N} a_{ij}x_j\right) - a_{ik}^2 x_k\right]
\end{aligned}$$

Le critère $J(x_k)$ est quadratique en x_k et sa dérivée est linéaire. On peut alors calculer explicitement la solution en annulant la dérivée :

$$\begin{aligned}
\frac{\partial J}{\partial x_k} = 0 \longrightarrow x_k &= \frac{1}{\sum_{i=1}^{M} a_{ik}^2}\sum_{i=1}^{M} a_{ik}\left(y_i - \sum_{j\neq k}^{N} a_{ij}x_j\right) \\
&= \frac{1}{\sum_{i=1}^{M} a_{ik}^2}\sum_{i=1}^{M} a_{ik}\left(y_i - \sum_{j=1}^{N} a_{ij}x_j + a_{ik}x_k\right) \\
&= x_k + \frac{1}{\sum_{i=1}^{M} a_{ik}^2}\sum_{i=1}^{M} a_{ik}\left(y_i - \sum_{j=1}^{N} a_{ij}x_j\right)
\end{aligned}$$

$$
\begin{aligned}
&= x_k + \frac{1}{< \boldsymbol{a}_{*k}, \boldsymbol{a}_{*k} >} \sum_{i=1}^{M} a_{ik} \left(y_i - [\boldsymbol{A}\boldsymbol{x}^{(k)}]_i \right) \\
&= x_k + \frac{1}{< \boldsymbol{a}_{*k}, \boldsymbol{a}_{*k} >} \left\langle \boldsymbol{a}_{*k}, (\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}^{(k)}) \right\rangle
\end{aligned}
$$

Pour simplifier ces écritures, notons par

$$\boldsymbol{x} = \boldsymbol{x}_{(-k)} + \boldsymbol{x}_{(k)}$$

avec

$$\boldsymbol{x}_{(-k)} = [x_1, \cdots, x_{k-1}, 0, x_{k+1}, \cdots, x_N]^t, \quad \text{et} \quad \boldsymbol{x}_{(k)} = [0, \cdots, 0, x_k, 0, \cdots, 0]^t$$

On a alors

$$J(x_k) = a x_k^2 + b x_k + c$$

avec

$$
\begin{aligned}
a &= \sum_i a_{ik}^2 = \|\boldsymbol{a}_{*k}\|^2 \\
b &= -2 \sum_i a_{ik} \left(y_i - \sum_{j \neq k}^{N} a_{ij} x_j \right) = -2 \left\langle \boldsymbol{a}_{*k}, (\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}_{(-k)}) \right\rangle \\
c &= \sum_i \left(y_i - \sum_{j \neq k}^{N} a_{ij} x_j \right)^2 = \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}_{(-k)}\|^2
\end{aligned}
$$

L'algorithme de Gausse-Seidel est fondée sur cette relation et donné par :

$$
\begin{aligned}
\boldsymbol{x}^{(0)} &= \mathbf{0} \\
x_k^{(k+1)} &= x_k^{(k)} + \alpha^{(k)} \frac{\partial J}{\partial x_k} \left(\boldsymbol{x}^{(k)} \right) \\
&= x_k^{(k)} + \alpha^{(k)} \left[\boldsymbol{A}^t \left(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}^{(k)} \right) \right]_k \\
&= x_k^{(k)} + \alpha^{(k)} \sum_{i=1}^{M} a_{ik} \left(y_i - [\boldsymbol{A}\boldsymbol{x}^{(k)}]_i \right) \quad \text{avec} \quad \alpha^{(k)} = \frac{1}{< \boldsymbol{a}_{*k}, \boldsymbol{a}_{*k} >}.
\end{aligned}
$$

Rappelons à titre de comparaison la Méthode de Kaczmarz que nous l'avons vue dans le chapitre concernant l'inversion généralisée :

$$
\begin{aligned}
\boldsymbol{x}^{(0)} &= \mathbf{0} \\
\boldsymbol{x}^{(k+1)} &= \boldsymbol{x}^{(k)} + \frac{\left(y_i - [\boldsymbol{A}\boldsymbol{x}^{(k)}]_i \right)}{< \boldsymbol{a}_{i*}, \boldsymbol{a}_{i*} >} \boldsymbol{a}_{i*} \\
&= \boldsymbol{x}^{(k)} + \frac{\left(y_i - \left\langle \boldsymbol{a}_{i*}, \boldsymbol{x}^{(k)} \right\rangle \right)}{< \boldsymbol{a}_{i*}, \boldsymbol{a}_{i*} >} \boldsymbol{a}_{i*}, \quad i = 1, \cdots, M, 1 \cdots, M, 1, \cdots
\end{aligned}
$$

B.2 Le cas d'un critère régularisé

Le critère et son gradient dans ce cas sont :

$$\begin{aligned} J(\boldsymbol{x}) &= \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|^2 + \lambda\|\boldsymbol{C}\boldsymbol{x}\|^2, \\ \nabla J &= -2\boldsymbol{A}^t(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}) + 2\lambda\boldsymbol{C}^t\boldsymbol{C}\boldsymbol{x} = -2\left[\boldsymbol{A}^t\boldsymbol{y} - (\boldsymbol{A}^t\boldsymbol{A} + 2\lambda\boldsymbol{C}^t\boldsymbol{C})\boldsymbol{x}\right] \end{aligned}$$

La même démarche nous conduit :

$$\begin{aligned} J(x_k) &= \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|^2 + \lambda\|\boldsymbol{C}\boldsymbol{x}\|^2 \\ &= \sum_{i=1}^{M}\left(y_i - \sum_{j=1}^{N} a_{ij}x_j\right)^2 + \sum_{j=1}^{N}\left(\sum_{n=1}^{N} c_{jn}x_n\right)^2 \\ &= \sum_{i=1}^{M}\left[\left(y_i - \sum_{j\neq k}^{N} a_{ij}x_j\right)^2 - 2a_{ik}\left(y_i - \sum_{j\neq k}^{N} a_{ij}x_j\right)x_k + a_{ik}^2x_k^2\right] \\ &\quad +\lambda\sum_{j\neq k}^{N}\left(\sum_{n\neq k}^{N} c_{jn}x_n + c_{jk}x_k\right)^2 + \lambda\sum_{n\neq k}^{N}(c_{kn}x_n + c_{kk}x_k)^2 \\ \frac{\partial J}{\partial x_k} &= -2\sum_{i=1}^{M} a_{ik}\left(y_i - \sum_{j=1}^{N} a_{ij}x_j\right) + 2\lambda\sum_{j=1}^{N}\sum_{n=1}^{N} c_{jk}c_{jn}x_n \\ &= -2\sum_{i=1}^{M} a_{ik}\left(y_i - \sum_{j\neq k}^{N} a_{ij}x_j - a_{ik}x_k\right) + 2\lambda\sum_{n=1}^{N} x_n\sum_{j=1}^{N} c_{jk}c_{jn} \\ &= -2\sum_{i=1}^{M}\left[a_{ik}\left(y_i - \sum_{j\neq k}^{N} a_{ij}x_j\right) - a_{ik}^2x_k\right] + 2\lambda\sum_{n\neq k}^{N} x_n\sum_{j=1}^{N} c_{jk}c_{jn} + 2\lambda x_k\sum_{j=1}^{N} c_{jk}^2 \end{aligned}$$

Solution explicite:

$$\nabla J = \boldsymbol{0} \longrightarrow (\boldsymbol{A}^t\boldsymbol{A} + \lambda\boldsymbol{C}^t\boldsymbol{C})\boldsymbol{x} = \boldsymbol{A}^t\boldsymbol{y} \longrightarrow \boldsymbol{x} = (\boldsymbol{A}^t\boldsymbol{A} + \lambda\boldsymbol{C}^t\boldsymbol{C})^{-1}\boldsymbol{A}^t\boldsymbol{y}$$

Méthode du gradient :

$$\begin{aligned} \boldsymbol{x}^{(0)} &= \boldsymbol{0} \\ \boldsymbol{x}^{(k+1)} &= \boldsymbol{x}^{(k)} + \alpha^{(k)}\nabla J\left(\boldsymbol{x}^{(k)}\right) \\ &= \boldsymbol{x}^{(k)} + \alpha^{(k)}\left[\boldsymbol{A}^t\left(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}^{(k)}\right) - \lambda\boldsymbol{C}^t\boldsymbol{C}\boldsymbol{x}^{(k)}\right] \\ &= \boldsymbol{x}^{(k)} + \alpha^{(k)}\boldsymbol{A}^t\boldsymbol{y} - \alpha^{(k)}\left(\boldsymbol{A}^t\boldsymbol{A} + \frac{\lambda}{\alpha}\boldsymbol{C}^t\boldsymbol{C}\right)\boldsymbol{x}^{(k)} \end{aligned}$$

Algorithme du Gauss-Seidel :

$$\begin{aligned} \frac{\partial J}{\partial x_k} &= 0 \longrightarrow \\ x_k &= \frac{1}{\sum_{i=1}^{M} a_{ik}^2 + \lambda\sum_{j=1}^{N} c_{jk}^2}\sum_{i=1}^{M} a_{ik}\left(y_i - \sum_{j\neq k}^{N} a_{ij}x_j\right) - \lambda\sum_{n\neq k}^{N} x_n\sum_{j=1}^{N} c_{jk}c_{jn} \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{\sum_{i=1}^{M} a_{ik}^2 + \lambda \sum_{j=1}^{N} c_{jk}^2} \sum_{i=1}^{M} a_{ik} \left(y_i - \sum_{j=1}^{N} a_{ij} x_j + a_{ik} x_k \right) - \lambda \sum_{n=1}^{N} x_n \sum_{j=1}^{N} c_{jk} c_{jn} + \lambda x_k \sum_{j=1}^{N} c_{jk}^2 \\
&= x_k + \frac{1}{\|\boldsymbol{a}_{*k}\|^2 + \lambda \|\boldsymbol{c}_{*k}\|^2} \left\{ \sum_{i=1}^{M} a_{ik} \left(y_i - [\boldsymbol{A}\boldsymbol{x}^{(k)}]_i \right) + \lambda \sum_{j=1}^{N} c_{jk} \sum_{n=1}^{N} c_{jn} x_n \right\} \\
&= x_k + \frac{1}{\|\boldsymbol{a}_{*k}\|^2 + \lambda \|\boldsymbol{c}_{*k}\|^2} \left\{ \left[\boldsymbol{A}^t \left(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}^{(k)} \right) \right]_k + \lambda \left[\boldsymbol{C}^t \boldsymbol{C} \boldsymbol{x}^{(k)} \right]_k \right\} \\
&= x_k + \frac{1}{\|\boldsymbol{a}_{*k}\|^2 + \lambda \|\boldsymbol{c}_{*k}\|^2} \left\{ [\boldsymbol{A}^t \boldsymbol{y}]_k - \left[\left(\boldsymbol{A}^t \boldsymbol{A} + \lambda \boldsymbol{C}^t \boldsymbol{C} \right) \boldsymbol{x}^{(k)} \right) \right]_k \right\} \\
&= x_k + \frac{1}{\|\boldsymbol{a}_{*k}\|^2 + \lambda \|\boldsymbol{c}_{*k}\|^2} \left\{ \left\langle \boldsymbol{a}_{*k}, (\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}^{(k)}) \right\rangle + \lambda \left[\boldsymbol{C}^t \boldsymbol{C} \boldsymbol{x}^{(k)} \right]_k \right\}
\end{aligned}
$$

Algorithme :

$$
\begin{aligned}
\boldsymbol{x}^{(0)} &= \boldsymbol{0} \\
x_k^{(k+1)} &= x_k^{(k)} + \alpha^{(k)} + \alpha^k \left\{ \left\langle \boldsymbol{a}_{*k}, (\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}^{(k)}) \right\rangle + \lambda \left[\boldsymbol{C}^t \boldsymbol{C} \boldsymbol{x}^{(k)} \right]_k \right\} \\
&\quad \text{avec} \quad \alpha^k = \frac{1}{\|\boldsymbol{a}_{*k}\|^2 + \lambda \|\boldsymbol{c}_{*k}\|^2}
\end{aligned}
$$