ENS Paris-Saclay

Plassier Vincent vincent.plassier@ens-paris- saclay.fr

TP OPTIMISATION Minimisation par optimisation proximale

27 mars 2018

1 Implémentation de Ista

Dans un premier temps, nous allons implémenter l'algorithme ISTA. Pour cela nous avons besoin du lemme suivant :

Lemme 1.1 *Nous avons,* $\forall \gamma,\ \lambda \geqslant 0$, $prox_{\lambda g}^{\gamma} = prox_{g}^{\lambda\gamma}$.

Dans cette partie nous allons nous concentrer sur la preuve de ce lemme.

D'après le cours nous avons l'existence et l'unicité de $prox_{\lambda h}^{\gamma}$ pour toute fonction h de $\mathbb{R}^d \longmapsto \mathbb{R}$. On en déduit l'existence d'un unique couple $(y_0, y_1) \in \mathbb{R}^d \times \mathbb{R}^d$ vérifiant :

$$y_0 = \underset{\mathbb{R}^d}{\operatorname{argmin}}\{g(y) + (2\lambda\gamma)^{-1}\left\|x-y\right\|^2\} \text{ et } y_1 = \underset{\mathbb{R}^d}{\operatorname{argmin}}\{\lambda g(y) + (2\gamma)^{-1}\left\|x-y\right\|^2\}$$

Par définition de y_0, on a $\forall y \in \mathbb{R}^d$,

$$g(y_0) + (2\lambda\gamma)^{-1}\left\|x-y_0\right\|^2 \leqslant g(y) + (2\lambda\gamma)^{-1}\left\|x-y\right\|^2$$

En multipliant cette inégalité par λ et en prenant l'infimum à gauche sur y on obtient,

$$y_0 \in \underset{\mathbb{R}^d}{\operatorname{argmin}}\{\lambda g(y) + (2\gamma)^{-1}\left\|x-y\right\|^2\}$$

Par uncité de y_1, on obtient : $y_0 = y_1$. D'où le résultat annoncé.
Grâce à se résultat, nous implémentons l'algorithme avec l'étape :

$$x_{k+1} = prox_{g}^{\lambda\gamma}\{x_k - \gamma\nabla f(x_k)\}$$

2 Acquisition comprimée

Dans le cadre d'un signal comprimé, on s'interesse désormais au minimum de la fonction F définit par :

$$F(x) = f(x) + g(x), \text{ avec } f : x \longmapsto \left\|y - Ax\right\|^2 \text{ et } g : x \longmapsto \sum_{i=1}^{d}\left\|x_i\right\|.$$

Maintenant calculons la différentielle de f. Nous avons l'égalité suivante :

$$\forall (x,h) \in \mathbb{R}^d \times \mathbb{R}^d,\ f(x+h) - f(x) = 2\langle A^*Ax - A^*y | h\rangle + o(h)$$

De cette dernière égalité, nous en déduisons : $\nabla f(x) = 2A^*(Ax - y)$.
Ainsi par Cauchy-Schwarz f est $2norm(A^*A)$-lipschitzienne.

Considérons maintenant une fonction h convexe, sci et propre sur $\mathbb{R}^d$ sous la forme $h(x) = \sum_{i=1}^{d} h_i(x_i)$.

Par définition, $prox_h^\gamma = \underset{\mathbb{R}^d}{\operatorname{argmin}}\{h(y) + (2\gamma)^{-1}\|x-y\|^2\} = \underset{\mathbb{R}^d}{\operatorname{argmin}}\{\sum_{i=1}^{d}[h_i(x_i) + (2\gamma)^{-1}(x_i - y_i)^2]\}$.
Prenons $x = (prox_{h_i}^\gamma)_{i\in\{1,\dots,d\}}$, par définition de x,

$$\forall \tilde{x} \in \mathbb{R}^d,\ h_i(x_i) + (2\gamma)^{-1}(x_i - y_i)^2 \leqslant h_i(\tilde{x}_i) + (2\gamma)^{-1}(\tilde{x}_i - y_i)^2$$

$$\text{D'où, } \sum_{i=1}^{d}[h_i(x_i) + (2\gamma)^{-1}(x_i - y_i)^2] \leqslant \sum_{i=1}^{d}[h_i(\tilde{x}_i) + (2\gamma)^{-1}(\tilde{x}_i - y_i)^2]$$

En prennat $\tilde{x} = (prox_{h_i}^\gamma)_{i\in\{1,\dots,d\}}$, on en déduit l'inégalité inverse. Par unicité de $prox_h^\gamma$ on en déduit que $prox_h^\gamma = (prox_{h_i}^\gamma)_{i\in\{1,\dots,d\}}$.

Considérons la fonction $\Phi(t) = |t|$, pour tout $t \in \mathbb{R}$.
Montrons que pour tout $u \in \mathbb{R}$ et $\gamma \leq 0$, $t_* = \underset{\mathbb{R}^d}{\operatorname{argmin}}\{\|t\| + (2\gamma)^{-1}(t-u)^2]\}$ si et seulement si $t_* \in u - \gamma\partial\Phi(t_*)$.
Nous avons la série d'équivalences suivantes :
$t_* = \underset{\mathbb{R}^d}{\operatorname{argmin}}\{\|t\| + (2\gamma)^{-1}(t-u)^2]\}$

$\iff$ t_* minimum de $\tilde{\Phi_\gamma}$

$\iff$ $0 \in \partial\tilde{\Phi}_\gamma(t_*) = t_* - u + \partial\Phi(t_*)$ d'après la proposition 8 du cours

$\iff$ $t_* \in u - \gamma\partial\Phi(t_*)$

Déterminons désormais la sous-différentielle de Φ sur $\mathbb{R}$.
Par définition de la sous-différentielle d'une application, pour $x \in \mathbb{R}$,

$$x \in \partial\mathbb{R} \iff |u| - |x| \geqslant (u-x)y$$

Cela équivaut à :

$$\begin{aligned} & |u| - uy \geqslant |x| - xy \\ \iff & o(u) = u(sign(u) - y) \geqslant x(sign(x) - y) \end{aligned}$$

Donc si $x \neq 0$, la dernière égalité montre que $y = sign(x)$
En condensant, nous avons : $\partial\Phi(x) = sign(x)$
Si $x = 0$, alors $\forall u$, $u(sign(u) - y) \geqslant 0$, donc $1 \geqslant y$ et $-1 \leqslant y$.
dans ce cas, on vérifie que $\partial\Phi(0) = [-1, 1]$.

D'après la question 4.b., nous avons la relation : $t_* = prox_\Phi^\gamma(u) \in u - \gamma\partial\Phi(t_*)$.
Cas 1 : si $t_* \leqslant 0$, alors $t_* = u + \gamma$ possible si $u + \gamma \geqslant 0$.
Cas 2 : si $t_* \geqslant 0$, alors $t_* = u - \gamma$ possible si $u \geqslant \gamma$.

Cas 3 : si $t_* = 0$, alors $t_* \in [u-\gamma, u+\gamma]$, cela n'est seulement possible pour $|u| \leqslant \gamma$.
Nous pouvons vérifier que :

$$prox_{\Phi}^{\gamma}(u) = \left(1 - \frac{\lambda}{|u|}\right)^{+} u$$

D'après la question 4.a., nous avons :

$$prox_{g}^{\gamma}(u) = (prox_{\Phi}^{\gamma})_{i\in\{1,\ldots,d\}}$$

La question 4.d. conclut :

$$prox_{g}^{\gamma}(u) = \left((1 - \frac{\lambda}{|x_i|})^{+} x_i\right)_{i\in\{1,\ldots,d\}}$$

.

Nous pouvons désormais implémenter les fonctions $prox_g^{\gamma}$ afin de tester nos algorithmes. Commençons par appliquer la fonction ISTA pour minimiser la fonction F. Nous obtenons l'image suivantes pour $\lambda_{l1} = 1$ et $n_{it} = 1000$:

Figure 1 – Image obtennue pour $\lambda_{l1} = 1$ et $n_{it} = 1000$

Nous constatons que l'image reste extêment bruité, en effet la valeur de λ_{l1} n'est pas optimale et le nombre n_it n'est pas suffisamment grand pour avoir une image nette.

Interessons-nous maintenant à la vitesse de convergence de l'algorithme ISTA. D'après notre cours nous savons que cet algorithme à une convergence de l'ordre de $O(\frac{1}{k})$, où k est le nombre d'itérations. Numériquement, montrons que $F(x_k) = F(x_*) + O(\frac{1}{k})$. Pour cela, nous utilisons la fonction $numpy.polyfit$ afin d'approcher $kF(x_k)$ par un polynôme de degré 1 (noté P). En effet, il nous suffit de monter que $kF(x_k) = kF(x_*) + O(1)$ pour conclure. Donc intuitivement le coefficient dominant de P sera proche de $F(x_*)$. Nous vérifions que $kF(x_k) - P(k)$ reste borné en traçant la courbe en fonction de k.

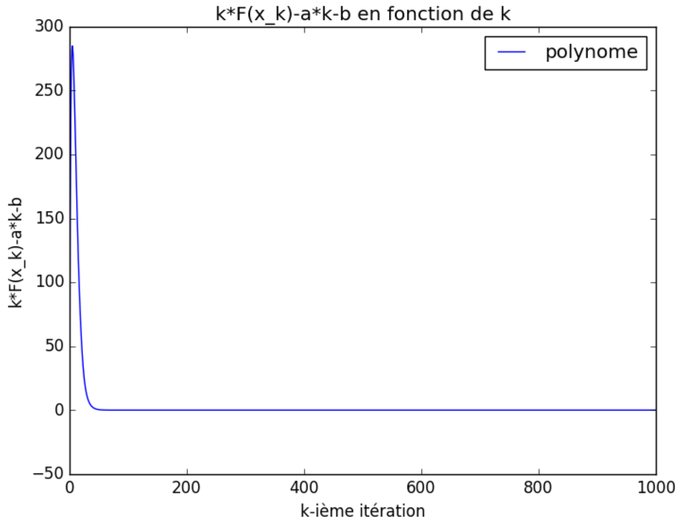

FIGURE 2 – Valeur de $|kF(x_k)|$ en fonction de k

La courbe semble tendre vers $o(1)$, cela confirme numériquement le résultat de notre cours.

Essayons de trouver une valeur de λ_{l1} permettant de débruité efficacement notre image. Afin de trouver ce λ_{l1} nous commençons par afficher les images pour des λ_{l1} d'ordre de grandeur bien distincts. Nous obtenous les figures suivantes :

(a) $\lambda_{l1} = 0.4$

(b) $\lambda_{l1} = 0.1$

(c) $\lambda_{l1} = 0.03$ (d) $\lambda_{l1} = 0.005$

FIGURE 3 – Image pour $n_{it} = 1000$ en fonction de λ_{l1}

Nous constatons qu'une valeur de $\lambda_{l1} = 0.03$ donne un meilleur rendu, bien que les différences entre $\lambda_{l1} = 0.1$ et $\lambda_{l1} = 0.03$ soient minimes. Nous pouvons supposé que la valeur optimale de λ recherchée se situe dans l'intervalle $[0.03, 0.1]$. Dans le cadre de notre problème, nous cherchons une solution x_{min} du problème de minimisation :

$$\underset{\mathbb{R}^d}{\operatorname{argmin}}\{\|y - Ax\|^2 + \lambda g(x)\}$$

L'objectif du terme $\lambda g(y)$ est de régulariser les solution afin d'obtenir un problème bien posé au sens d'Hadamard. Intuitivement, on souhaite que la solution ne varie pas brusquement lors d'une légère perturbation de y. Cependant λ ne doit pas être trop grand pour ne pas complètement modifier le problème initial.