Fast Iterative Shrinkage-Thresholding Algorithm

1 Sur la définition de FB, ISTA, FISTA

On souhaite résoudre un problème de la forme

$$\widehat{x} = \underset{x \in \mathcal{H}}{\arg\min} \, f(x) + g(x) \tag{1}$$

où f est une fonction différentiable de gradient L-Lipschitzien et $g \in \Gamma_0(\mathcal{H})$ est "proximable" (au sens où son opérateur proximal est facile à calculer).

La résolution de ce problème peut se faire par un schéma *forward-backward* qui s'écrit

for $t \in \mathbb{N}$ **do**
 $y_{n+1} = x_n - \gamma \nabla f(x_n)$
 $x_{n+1} = \mathrm{prox}_{\gamma g}(y_{n+1})$
end

Algorithm 1: Algorithme *forward-backward* (FB) pour résoudre le problème (1) (dans le cas $\rho_n = 1$)

Le cas d'une suite ρ_n non constante fait partie de la grande classe des algorithmes forward-backward.

Cas particulier : $g(x) = \|x\|_1$ L'opérateur proximal de la norme $\|\cdot\|_1$ réalise un seuillage (ou *soft-thresholding* en anglais) et s'écrit

$$p = \mathrm{prox}_{\lambda \|\cdot\|_1}(x) \Leftrightarrow p^{(i)} = \begin{cases} 0 & \text{si } |x^{(i)}| \leq \lambda \\ x^{(i)} - \lambda \frac{x^{(i)}}{|x^{(i)}|} & \text{sinon.} \end{cases}$$

Dans ce cas on parle de l'algorithme ISTA (pour **I**terative **S**hrinkage-**T**hresholding **A**lgorithm) qui est un cas particulier d'algorithme *forward-backward*. Cet algorithme a été accéléré, ce qui a donné une version **F**ast ISTA (FISTA) et cette accélération étant en fait valable dans un cadre plus général on a gardé le même nom lorsque g est une fonction quelconque de $\Gamma_0(\mathcal{H})$. Cela a donné lieu à quelques abus de langage : considérons ici que

— ISTA est un cas particulier de FB.
— on appelle FISTA l'algorithme forward-backward accéléré de minimisation du problème *général* (1) faisant intervenir la suite t_n. On pourra donc préciser FISTA de Beck (et Teboulle) lorsque la suite est $t_{n+1} = \sqrt{4t_n^2} \dots$ et FISTA de Chambolle (et Dossal) lorsque $t_n = (n + a - 1)/a$.

2 Si la suite converge, elle converge vers un minimiseur

Le chapitre 5 du cours qui présente les algorithmes et les résultats de convergence vous montre que l'algorithme *forward-backward* est construit sur un modèle de type $x_{n+1} = T(x_n)$ pour un certain opérateur non-expansif T qui dépend de f et g.

Dans ce cas si la suite $(x_n)_n$ converge (faiblement) c'est vers un point fixe de T. Tout l'enjeu de ce chapitre (extrêmement dense par ailleurs) est de vous montrer comment on construit un opérateur T dont les points fixes sont les minimiseurs de F.

3 Les lemmes de convergence

Les lemmes préliminaires sont assez techniques et nécessitent de connaître des propriétés *ad hoc* des fonctions fortement convexes (et/ou de gradient L-Lipschitzien), je ne vois pas d'intérêt à ce que vous les redémontriez. Après si vous le souhaitez je peux vous indiquer des sources. Mais ce n'est absolument pas prioritaire ! De même la co-coercivité de ∇f est un résultat d'analyse convexe, notamment recensé sur ce blog (https ://xingyuzhou.org/blog/notes/Lipschitz-gradient), que vous pouvez prendre tel quel.

4 La convergence de Φ_n est décisive

Cet article, comme beaucoup d'autres, ne développe pas un raisonnement purement linéaire et la difficulté c'est de remettre dans l'ordre les éléments de preuves, de comprendre ce qui est démontré, ce qui est simplement cité, ...

Plus que les calculs, ce qui m'intéresse c'est que vous vous soyez appropriés les étapes et ingrédients du raisonnement. Pour vous aider à y voir plus clair, voici l'architecture de la preuve. L'idéal c'est que la partie 2 de votre rapport présente avec vos mots les grandes étapes de cette preuve.

Le raisonnement proposé dans l'article de Chambolle et Dossal est le suivant :

i) **Le théorème 2 et son corollaire permettent d'étendre le théorème 1 et d'avoir plus d'informations sur la convergence de $F(x_n)$ vers $F(x^*)$.**

ii) **En particulier on peut en déduire que la suite $(x_n)_n$ est *faiblement compacte* :** En effet à partir du corollaire 2, la suite $(n(x_{n+1}-x_n))_n$ est bornée, donc a fortiori la suite $(x_{n+1}-x_n)_n$ est bornée. De plus, le Théorème 2 montre que la suite $(v_n)_n$ est bornée. Or

$$v_n := \frac{1}{2}\|u_n - x^*\|_2^2,$$

donc cela induit directement que $(u_n)_n$ est bornée et comme

$$\underbrace{u_n}_{\text{bornée}} = x_{n-1} + \underbrace{t_n(x_n - x_{n+1})}_{\text{bornée}}, \quad \text{avec } t_n = \frac{n+a-1}{a} \text{ (proposé dans cet article)},$$

la suite $(x_n)_n$ est bornée. Cela signifie donc qu'elle est *faiblement compacte* (i.e. compacte pour la topologie faible).

iii) **En combinant les points i) et ii) (et en traduisant ces résultats sur l'algorithme présenté en page 3), on obtient que toute valeur d'adhérence (pour la topologie faible) de la suite $(x_n)_n$ est un minimiseur de F :**
En effet, soit une sous-suite faiblement convergente $\left(x_{\varphi(n)}\right)_n$ telle que $x_{\varphi(n)} \underset{n\to\infty}{\rightharpoonup} \tilde{x}$ (pour l'instant on ne sait pas que $\tilde{x}$ est un minimiseur).
Comme $y_n = x_n + \alpha_n(x_n - x_{n-1})$ et $\delta_n = \frac{1}{2}\|x_n - x_{n-1}\|_2^2 \underset{n\to\infty}{\longrightarrow} 0$, on a $y_{\varphi(n)} \underset{n\to\infty}{\longrightarrow} \tilde{x}$. Or puisque $y_{\varphi(n)} = \left(1 - \frac{1}{t_{\varphi(n)+1}}\right) x_{\varphi(n)} + \frac{1}{t_{\varphi(n)}+1} u_{\varphi(n)}$ (avec $(u_n)_n$ bornée), on a donc $y_{\varphi(n)} \underset{n\to\infty}{\rightharpoonup} \tilde{x}$.

Par suite, on obtient, en passant à la limite $n \to \infty$, $\tilde{x} = T(\tilde{x})$. Or les points fixes de l'opérateur non-expansif T sont exactement les minimiseurs de F (T a été construit dans ce but, cf. le cours).

iv) **Les calculs des pages 9-10 aboutissent à la conclusion que $\Phi_n := \frac{1}{2}\|x_n - x^*\|_2^2$ admet une limite. Dans ce cas la suite x_n est convergente.**

En effet la suite étant faiblement compacte, elle est (faiblement) convergente si et seulement si elle admet une unique valeur d'adhérence (pour la topologie faible). Montrons que la convergence de $\Phi_n[x^*]$ (je note une dépendance en x^* car le résultats est démontré *quelque soit* le point fixe de T considéré) vers une limite notée ℓ implique bien l'unicité de la valeur d'adhérence.

Soit $(x_{\varphi(n)}) \rightharpoonup \tilde{x}$ et $(x_{\varphi'(n)}) \rightharpoonup \tilde{x}'$, on a alors $\tilde{x}$ et $\tilde{x}'$ qui sont des minimiseurs de F (ou de manière équivalente des points fixes de T). Par conséquent le résultats de convergence démontré pour $\Phi_n[\tilde{x}]$ et $\Phi_n[\tilde{x}']$ s'écrit :

$$\lim_{n\to\infty} \|x_{\varphi(n)} - \tilde{x}\| = \lim_{n\to\infty} \|x_{\varphi'(n)} - \tilde{x}\| \overset{\text{(déf)}}{=} \ell \qquad \text{car } \Phi_n[\tilde{x}] := \|x_n - \tilde{x}\|/2 \text{ admet une limite}$$

et

$$\lim_{n\to\infty} \|x_{\varphi'(n)} - \tilde{x}'\| = \lim_{n\to\infty} \|x_{\varphi(n)} - \tilde{x}'\| \overset{\text{(déf)}}{=} \ell' \qquad \text{car } \Phi_n[\tilde{x}'] := \|x_n - \tilde{x}'\|/2 \text{ admet une limite.}$$

Or le théorème d'Opial nous dit (en particulier car il est en fait plus général) que si une suite $(z_n)_n$ converge faiblement vers z_0 alors s'il existe $z \in \mathcal{H}$ tel que $\lim_{n\to\infty} \|z_n - z\| \leq \lim_{n\to\infty} \|z_n - z_0\|$ alors $\boxed{z = z_0}$.

i) si $\ell \leq \ell'$ alors

$$\lim_{n\to\infty} \|x_{\varphi'(n)} - \tilde{x}\| \leq \lim_{n\to\infty} \|x_{\varphi'(n)} - \tilde{x}'\|$$

et l'application du théorème d'Opial à la suite $\left(x_{\varphi'(n)}\right)_n$ donne $\tilde{x} = \tilde{x}'$.

ii) si $\ell' \leq \ell$ alors on applique le théorème d'Opial à la suite $\left(x_{\varphi(n)}\right)_n$ ce qui également donne $\tilde{x} = \tilde{x}'$.

Donc la suite $(x_n)_n$ possède une unique valeur d'adhérence (qui est un minimiseur de F) et elle converge (faiblement) vers ce minimiseur.