Chapitre 1

Méthodes de gradient projeté

1.1 Projection orthogonale sur un ensemble convexe fermé

Définition 1.1. Soit $\mathcal{C}$ un sous-ensemble convexe fermé non vide de $\mathbb{R}^n$. On appelle projection orthogonale d'un point $y \notin \mathcal{C}$ sur $\mathcal{C}$, tout point $x \in \mathcal{C}$ réalisant le minimum du problème suivant :

$$\begin{cases} \min \ \|x - y\|^2 \\ x \in \mathcal{C} \end{cases} \tag{1.1}$$

cette solution est notée par $\mathcal{P}_{\mathcal{C}}(y)$.

Théorème 1.1. *(Premier théorème de projection). Soit $\mathcal{C}$ un sous-ensemble convexe fermé non vide de $\mathbb{R}^n$. Alors le problème* (1.1) *a une solution optimale unique.*

Démonstration. On note $A = I$ matrice unitaire d'ordre 1. Alors, nous avons :

$$f(x) = \|x - y\|^2 = \|Ax - y\|^2 = x^T A x - 2y^T A x + \|y\|^2$$

est une fonction quadratique strictement convexe car A est définie positive. D'autre par, on a :

$$\lim_{\|x\| \to +\infty} f(x) = +\infty$$

D'où f est coercive. Donc, le problème (1.1) admet une solution unique. □

Exemple 1.1. Soit $\mathcal{C} = \mathbb{R}^n_+$. Pour calculer la projection orthogonale de $y \in \mathbb{R}^n$ sur C, nous devons résoudre le problème d'optimisation convexe suivant :

$$\begin{cases} \min \ \sum_{i=1}^{n} (x_i - y_i)^2 \\ x_1, x_2, \ldots, x_n \geq 0 \end{cases} \tag{1.2}$$

Pour tout $i = 1, 2, \ldots, n$, la solution optimale de problème $\min \left\{ (x_i - y_i)^2 : x_{i \geq 0} \right\}$ est donnée par :

$$x_i^* = [y_i]_+ = \begin{cases} y_i & \text{si} \ \ y_i \geq 0, \\ 0 & \text{si} \ \ y_i < 0. \end{cases}$$

Donc, la solution optimale du problème (1.2) est donnée par :

$$x^* = \mathcal{P}_{\mathcal{C}}(y) = [y]_+ = \left([y_1]_+, [y_2]_+, \ldots, [y_n]_+\right)^T.$$

Exemple 1.2. (Projection sur des boîtes fermés). Une boîte est un sous-ensemble de $\mathbb{R}^n$ de la forme

$$B = [\ell_1, u_1] \times [\ell_2, u_2] \times \ldots \times [\ell_n, u_n],$$

où $\ell_i \leq u_i$ pour tous $i = 1, 2, \ldots, n$. la projection orthogonale de $y \in \mathbb{R}^n$ sur B est donnée par :

$$x^* = \mathcal{P}_B\left(y\right), \tag{1.3}$$

où

$$x_i^* = \begin{cases} u_i & \text{si} \quad y_i \geq u_i, \\ y_i & \text{si} \quad \ell_i \leq y_i \leq u_i, \\ \ell_i & \text{si} \quad y_i \leq \ell_i. \end{cases}$$

pour tout $i = 1, 2, \ldots, n$.

Exemple 1.3. (Projection sur des boules fermés). Soit $\mathcal{C} = \bar{B}\left(x_0, r\right) = \{y \in \mathbb{R}^n : \|y - x_0\| \leq r\}$. La projection orthogonale de $y \in \mathbb{R}^n$ sur la boule fermée $\mathcal{C}$ est une solution unique du problème d'optimisation convexe suivant :

$$\min_x \left\{ \|x - y\|^2 : x \in \mathcal{C} \right\}.$$

Nous avons :

$$x^* = \mathcal{P}_{\mathcal{C}}\left(y\right) = \begin{cases} y \;\; \text{si} \quad y \in \mathcal{C}, \\ x_0 + r\frac{y - x_0}{\|y - x_0\|} \quad \text{si} \quad y \notin \mathcal{C}. \end{cases}$$

Théorème 1.2. *(Deuxième théorème de projection). Soit $\mathcal{C}$ un sous-ensemble convexe fermé non vide de $\mathbb{R}^n$. Alors, $z = \mathcal{P}_{\mathcal{C}}\left(x\right)$ si et seulement si*

$$\left(x - z\right)^T \left(y - z\right) \leq 0, \;\; \textit{pour tout } y \in \mathcal{C}. \tag{1.4}$$

Démonstration. $z = \mathcal{P}_{\mathcal{C}}\left(x\right)$ si et seulement si est une solution optimale du problème

$$\begin{cases} \min \;\; g\left(y\right) = \|y - x\|^2 \\ y \in \mathcal{C}. \end{cases}$$

Par convexité de $\mathcal{C}$, on a

$$\forall \lambda \in [0, 1], (1 - \lambda) z + \lambda y = z + \lambda\left(y - z\right) \in \mathcal{C},$$

donc

$$g\left(z + \lambda\left(y - z\right)\right) - g\left(z\right) \geq 0$$

On divise ensuite par $\lambda > 0$ et on fait tendre λ tend vers 0^+, on obtient (1.4). □

Théorème 1.3. *Soit $\mathcal{C}$ un sous-ensemble convexe fermé non vide de $\mathbb{R}^n$. Alors, nous avons :*

1. *Pour tout $v, w \in \mathbb{R}^n$, on a :*

$$\left(\mathcal{P}_{\mathcal{C}}\left(v\right) - \mathcal{P}_{\mathcal{C}}\left(w\right)\right)^T \left(v - w\right) \geq \|\mathcal{P}_{\mathcal{C}}\left(v\right) - \mathcal{P}_{\mathcal{C}}\left(w\right)\|^2. \tag{1.5}$$

2. *Pour tout $v, w \in \mathbb{R}^n$, on a :*

$$\|\mathcal{P}_{\mathcal{C}}\left(v\right) - \mathcal{P}_{\mathcal{C}}\left(w\right)\| \leq \|v - w\|. \tag{1.6}$$

Démonstration. 1. Avec Théorème 1.2, nous avons pour tout $x \in \mathbb{R}^n$ et $y \in \mathcal{C}$, on a :

$$(x - \mathcal{P}_{\mathcal{C}}(x))^T (y - \mathcal{P}_{\mathcal{C}}(x)) \leq 0. \tag{1.7}$$

On pose dans (1.7) $x = v$ et $y = \mathcal{P}_{\mathcal{C}}(w)$, nous avons :

$$(v - \mathcal{P}_{\mathcal{C}}(v))^T (\mathcal{P}_{\mathcal{C}}(w) - \mathcal{P}_{\mathcal{C}}(v)) \leq 0. \tag{1.8}$$

D'autre part, on pose dans (1.7) $x = w$ et $y = \mathcal{P}_{\mathcal{C}}(v)$, nous avons :

$$(w - \mathcal{P}_{\mathcal{C}}(w))^T (\mathcal{P}_{\mathcal{C}}(v) - \mathcal{P}_{\mathcal{C}}(w)) \leq 0. \tag{1.9}$$

En additionnant (1.8) et (1.9), on obtient :

$$(\mathcal{P}_{\mathcal{C}}(w) - \mathcal{P}_{\mathcal{C}}(v))^T (v - w + \mathcal{P}_{\mathcal{C}}(w) - \mathcal{P}_{\mathcal{C}}(v)) \leq 0,$$

et ainsi,

$$(\mathcal{P}_{\mathcal{C}}(v) - \mathcal{P}_{\mathcal{C}}(w))^T (v - w) \geq \|\mathcal{P}_{\mathcal{C}}(v) - \mathcal{P}_{\mathcal{C}}(w)\|^2.$$

2. Si $\mathcal{P}_{\mathcal{C}}(v) = \mathcal{P}_{\mathcal{C}}(w)$, l'inégalité est trivial. Si $\mathcal{P}_{\mathcal{C}}(v) \neq \mathcal{P}_{\mathcal{C}}(w)$, avec l'inégalité de Cauchy-Schwarz, on obtient :

$$(\mathcal{P}_{\mathcal{C}}(v) - \mathcal{P}_{\mathcal{C}}(w))^T (v - w) \leq \|\mathcal{P}_{\mathcal{C}}(v) - \mathcal{P}_{\mathcal{C}}(w)\| \|v - w\|$$

En utilisant (1.5), on obtient :

$$\|\mathcal{P}_{\mathcal{C}}(v) - \mathcal{P}_{\mathcal{C}}(w)\|^2 \leq \|\mathcal{P}_{\mathcal{C}}(v) - \mathcal{P}_{\mathcal{C}}(w)\| \|v - w\|$$

Divisant par $\|\mathcal{P}_{\mathcal{C}}(v) - \mathcal{P}_{\mathcal{C}}(w)\|$ on trouve (1.6).

$\square$

Théorème 1.4. *Soient $\mathcal{C}$ un sous-ensemble convexe fermé non vide de $\mathbb{R}^n$, $f : \mathcal{C} \to \mathbb{R}$ une fonction de classe $\mathcal{C}^1$ et $\alpha > 0$. Alors, $x^* \in \mathcal{C}$ est un point critique de problème suivant :*

$$(P) : \begin{cases} \min \ f(x) \\ x \in \mathcal{C}. \end{cases}$$

si et seulement si

$$x^* = \mathcal{P}_{\mathcal{C}}(x^* - \alpha \nabla f(x^*)) \tag{1.10}$$

Démonstration. Avec Théorème 1.2, nous avons $x^* = \mathcal{P}_{\mathcal{C}}(x^* - \alpha \nabla f(x^*))$ si et seulement si

$$(x^* - \alpha \nabla f(x^*) - x^*)^T (x - x^*) \leq 0, \ \text{ pour tout } x \in \mathcal{C}.$$

Cependant, cette dernière relation est équivalente à

$$\nabla f(x^*)^T (x - x^*) \geq 0, \ \text{ pour tout } x \in \mathcal{C}.$$

D'où x^* est un point critique de f sur $\mathcal{C}$. $\square$

1.2 Méthode de gradient projeté

Définition 1.2. Soient $\mathcal{C}$ un sous-ensemble convexe fermé non vide de $\mathbb{R}^n$ et $f : \mathcal{C} \to \mathbb{R}$ une fonction de classe $\mathcal{C}^1$. On appelle méthode de gradient projeté pour résoudre le problème d'optimisation suivant :

$$\begin{cases} \min \ f(x) \\ x \in \mathcal{C}. \end{cases}$$

la suite de récurrence définie par :

$$\begin{cases} x_0 \in \mathcal{C} \text{ donné}, \\ x_{k+1} = \mathcal{P}_{\mathcal{C}}\left(x_k - \alpha_k \nabla f(x_k)\right), \ \ \text{pour tout } k \in \mathbb{N}. \end{cases} \tag{1.11}$$

De plus,

1. Si $\alpha_k = \alpha > 0$ pour tout $k \in \mathbb{N}$, on dit (1.11) méthode de gradient projeté à pas fixe.
2. Si α_k est le point de minimum global de la fonction g définie par : $g(\alpha) = f(x_k - \alpha \nabla f(x_k))$, on dit (1.11) méthode de gradient projeté à pas optimale.

Pour étudier la convergence de la méthode du gradient projeté, nous avons le lemme suivant :

Lemme 1.1. *Soient $\mathcal{C}$ un sous-ensemble convexe fermé non vide de $\mathbb{R}^n$ et $f : \mathcal{C} \to \mathbb{R}$ une fonction de classe $\mathcal{C}^1$. Si ∇f est de Lipschitz c'est à dire*

$$\exists L > 0, \ \forall x, y \in \mathcal{C} \ \ \|\nabla f(x) - \nabla f(y)\| \leq L \|x - y\|,$$

alors,

1. *Nous avons :*
$$f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{L}{2} \|y - x\|^2. \tag{1.12}$$
2. *Pour tout $x \in \mathcal{C}$ et $\alpha \in \,]0, 2/L[$, nous avons :*
$$f\left[\mathcal{P}_{\mathcal{C}}(x - \alpha \nabla f(x))\right] - f(x) \leq \left(\frac{L}{2} - \frac{1}{\alpha}\right) \|\mathcal{P}_{\mathcal{C}}(x - \alpha \nabla f(x)) - x\|^2. \tag{1.13}$$

Démonstration. 1. Soient $x, y \in \mathcal{C}$. On définit la fonction $g(t) = f(x + t(y - x))$ pour tout $t \in [0, 1]$. Nous avons f est de classe $\mathcal{C}^1$ sur $\mathcal{C}$, alors g est de classe $\mathcal{C}^1$ sur $[0, 1]$. Alors, nous avons :

$$g(1) - g(0) = \int_0^1 g'(t)\, dt$$

Ainsi, $g(1) = f(y)$, $g(0) = f(x)$ et $g'(t) = \langle \nabla f(x + t(y - x)), y - x \rangle$, nous obtenons

$$\begin{aligned} f(y) - f(x) - \langle \nabla f(x), y - x \rangle &= \int_0^1 \langle \nabla f(x + t(y - x)) - \nabla f(x), y - x \rangle\, dt \\ &\leq \|y - x\| \int_0^1 \|\nabla f(x + t(y - x)) - \nabla f(x)\|\, dt \\ &\leq L \|y - x\|^2 \int_0^1 t dt \\ &= \frac{L}{2} \|y - x\|^2. \end{aligned}$$

D'où (1.12).

2. On pose dans (1.12) $x^+ = \mathcal{P}_{\mathcal{C}}(x - \alpha \nabla f(x))$, on obtient :

$$f\left(x^+\right) \le f(x) + \left\langle \nabla f(x), x^+ - x \right\rangle + \frac{L}{2}\left\|x^+ - x\right\|^2. \tag{1.14}$$

En utilisant Théorème 1.2, nous avons :

$$\left(x - \alpha \nabla f(x) - x^+\right)^T \left(x - x^+\right) \le 0.$$

Donc, on a :

$$\nabla f(x)^T \left(x^+ - x\right) \le -\frac{1}{\alpha}\left\|x^+ - x\right\|^2. \tag{1.15}$$

De (1.14) et (1.15), on obtient :

$$f\left(x^+\right) \le f(x) + \left(\frac{L}{2} - \frac{1}{\alpha}\right)\left\|x^+ - x\right\|^2.$$

D'où (1.13).

□

On peut donc montrer un premier résultat de convergence de la méthode de gradient projeté à pas fixe.

Théorème 1.5. *Soient $\mathcal{C}$ un sous-ensemble convexe fermé non vide de $\mathbb{R}^n$, $f : \mathcal{C} \to \mathbb{R}$ une fonction de classe $\mathcal{C}^1$ et bornée inférieurement. On suppose que ∇f est de Lipschitz sur $\mathcal{C}$ c'est à dire*

$$\exists L > 0,\ \forall x, y \in \mathcal{C}\ \ \|\nabla f(x,y) - \nabla f(x,y)\| \le L\|x - y\|.$$

Soit (x_k) la suite de la méthode de gradient projeté définie par :

$$\begin{cases} x_0 \in \mathcal{C} \text{ donné}, \\ x_{k+1} = \mathcal{P}_{\mathcal{C}}(x_k - \alpha \nabla f(x_k)), \ \text{ pour tout } k \in \mathbb{N}. \end{cases}$$

Alors, si on choisit un pas $0 < \alpha < 2/L$, nous avons les résultats suivants :

1. *Si (x_k) converge vers x^*, alors x^* est un point critique de la fonction f sur $\mathcal{C}$.*
2. *Si f est convexe sur $\mathcal{C}$ et (x_k) converge vers x^*, alors x^* est un point de minimum global de f sur $\mathcal{C}$.*
3. *Si f est concave sur $\mathcal{C}$ et (x_k) converge vers x^*, alors x^* est un point de maximum global de f sur $\mathcal{C}$.*

Démonstration. 1. On pose dans (1.13), $x = x_k$, on obtient :

$$f(x_{k+1}) - f(x_k) \le \left(\frac{L}{2} - \frac{1}{\alpha}\right)\|\mathcal{P}_{\mathcal{C}}(x_k - \alpha \nabla f(x_k)) - x_k\|^2 \le 0. \tag{1.16}$$

Donc, la suite $(f(x_k))$ et décroissante et bornée inférieurement, alors est convergente vers une limite finie. D'autre parte, si $x_k \xrightarrow[k\to+\infty]{} x^*$, alors par passage à la limite dans (1.16), on obtient :

$$\mathcal{P}_{\mathcal{C}}(x^* - \alpha \nabla f(x^*)) = x^*$$

Donc, d'après Théorème 1.4, x^* est un point critique de f sur $\mathcal{C}$.

2. Si f est convexe sur $\mathcal{C}$ et $x_k \xrightarrow[k\to+\infty]{} x^*$, alors d'après la condition nécessaire et suffisante d'optimalité du premier ordre x^* est un point de minimum global de f sur $\mathcal{C}$.

3. Si f est concave sur $\mathcal{C}$ et $x_k \xrightarrow[k\to+\infty]{} x^*$, alors d'après la condition nécessaire et suffisante d'optimalité du premier ordre x^* est un point de maximum global de f sur $\mathcal{C}$.

□

Exemple 1.4. On considère le problème suivant :

$$\begin{cases} \min \ f(x) = \frac{1}{2}x^T Ax, \\ \|x\|^2 \leq 1. \end{cases} \tag{1.17}$$

Où A est une matrice symétrique et définie positive. Supposons que nous appliquons l'algorithme de gradient projeté à pas fixe sur le problème (1.17).

1. Écrire x_{k+1} en fonction de x_k, A et α.
2. Est-il possible que l'algorithme ne converge pas vers une solution optimale ? même si le pas $\alpha > 0$ est arbitrairement petit.
3. Montrer que pour $0 < \alpha < 1/\lambda_{max}$ (où λ_{max} est la plus grande valeur propre de la matrice A), l'algorithme de gradient projeté à pas fixe est converge vers une solution optimale, à condition que x_0 ne soit pas orthogonal aux vecteurs propres de A correspondant à la plus petite valeur propre.

Solution :

1. On pose $\mathcal{C} = \left\{x \in \mathbb{R}^n : \|x\|^2 \leq 1\right\}$. L'opérateur de projection est donné par :

$$\mathcal{P}_{\mathcal{C}}(x) = \frac{x}{\|x\|} \quad \text{pour tout } x \in \mathbb{R}^n \backslash \{0\}.$$

Donc, nous avons :

$$x_{k+1} = \mathcal{P}_{\mathcal{C}}[x - \alpha\nabla f(x_k)] = \frac{x_k - \alpha Ax_k}{\|x_k - \alpha Ax_k\|} = \beta_k (I - \alpha A) x_k \ \text{ où } \ \beta_k = 1/\|(I - \alpha A) x_k\|.$$

2. Si nous commençons par x_0 étant un vecteur propre de A, alors on a :

$$x_1 = \beta_0 (I - \alpha A) x_0 = \frac{(1 - \alpha\lambda_0) x_0}{|1 - \alpha\lambda_0| \|x_0\|} = x_0.$$

Donc, $x_k = x_0$ pour tout $k \in \mathbb{N}$. Par conséquent, il est clair que l'algorithme est bloqué à un point qui n'est pas optimal.

3. Nous avons :

$$\begin{aligned} x_{k+1} &= \beta_k (I - \alpha A) x_k \\ &= \beta_k (I - \alpha A) \left(y_1^k v_1 + \ldots + y_n^k v_n\right) \\ &= \beta_k \left[y_1^k (I - \alpha A) v_1 + \ldots + y_n^k (I - \alpha A) v_n\right]. \end{aligned}$$

Car $(I - \alpha A) v_i = (1 - \alpha\lambda_i) v_i$ où λ_i est la valeur propre correspondante de vecteur propre v_i. Ainsi,

$$x_{k+1} = \beta_k \left[y_1^k (I - \alpha\lambda_1) v_1 + \ldots + y_n^k (I - \alpha\lambda_n) v_n\right],$$

ce qui signifie que $y_i^{k+1} = \beta_k y_i^k (1 - \alpha\lambda_i)$. D'autre part, nous avons :

$$y_i^k = \left(\prod_{i=0}^{k-1} \beta_i\right) y_i^0 (1 - \alpha\lambda_i)^k.$$

Donc, on a :

$$x_k = \sum_{i=1}^{n} y_i^k v_i = y_1^k \left[v_1 + \sum_{i=2}^{n} \frac{y_i^k}{y_1^k} v_i \right].$$

On suppose que $y_1^0 \neq 0$, on obtient :

$$\frac{y_i^k}{y_1^k} = \frac{y_i^0 \left(1 - \alpha\lambda_i\right)^k}{y_1^0 \left(1 - \alpha\lambda_1\right)^k} = \frac{y_i^0}{y_1^0} \left(\frac{1 - \alpha\lambda_i}{1 - \alpha\lambda_1} \right)^k.$$

Nous avons $\lambda_i > \lambda_1$ et $0 < \alpha < \lambda_{max}$, alors $\left(1 - \alpha\lambda_i\right) / \left(1 - \alpha\lambda_1\right) < 1$, nous déduisons que

$$\frac{y_i^k}{y_1^k} \xrightarrow[k \to +\infty]{} 0,$$

ce qui implique que $x_k \xrightarrow[k \to +\infty]{} v_1$.

1.3 Méthodes de gradient projeté avec contraintes d'égalités linéaires

Dans cette section, nous considérons les problèmes d'optimisation de la forme

$$\begin{cases} \min \ f(x) \\ Ax = b. \end{cases} \tag{1.18}$$

Où $f : \mathbb{R}^n \to \mathbb{R}$ est une fonction de classe $\mathcal{C}^1$, A est une matrice rectangulaire de taille $m \times n$ avec $m < n$ et $rang\,(A) = m$, $b \in \mathbb{R}^m$. Dans le problème ci-dessus, l'ensemble de contraintes est

$$\Omega = \{ x \in \mathbb{R}^n : Ax = b \}.$$

La structure spécifique de l'ensemble de contraintes nous permet de calculer l'opérateur de projection orthogonal $\mathcal{P}$. Plus précisément, $\mathcal{P}(x)$ peut être défini en utilisant la matrice de projection orthogonale P donné par :

$$P = I_n - A^T \left(AA^T\right)^{-1} A.$$

Propriété 1.1. La matrice de projection orthogonale est vérifiée les propriétés suivantes :

1. $P = P^T$.
2. $P^2 = P$.

Une autre propriété de la matrice projection orthogonal dont nous avons besoin dans notre discussion est donnée dans le lemme suivant.

Lemme 1.2. *Soit $v \in \mathbb{R}^n$. Alors, nous avons :*

1. *$Pv = 0$ si et seulement si $v \in Im\left(A^T\right)$; c'est dire $Ker\,(P) = Im\left(A^T\right)$.*
2. *$Av = 0$ si et seulement si $v \in Im\,(P)$; c'est à dire $Ker\,(A) = Im\,(P)$.*

Démonstration. 1. Nous avons :

$$Pv = \left(I_n - A^T \left(AA^T\right)^{-1} A\right) v = v - A^T \left(AA^T\right)^{-1} Av.$$

Si $Pv = 0$, alors

$$v = A^T \left(AA^T\right)^{-1} Av$$

et donc $v \in Im\left(A^T\right)$.
D'autre part, Supposons qu'il existe $u \in \mathbb{R}^m$ tel que $v = A^T u$. Alors,

$$Pv = \left(I_n - A^T \left(AA^T\right)^{-1} A\right) A^T u = A^T u - A^T AA^T AA^T u = 0.$$

Par conséquent, nous avons prouvé que $Ker\left(P\right) = Im\left(A^T\right)$.

2. En utilisant un argument similaire à celui ci-dessus, nous pouvons montrer que $Ker\left(A\right) = Im\left(P\right)$.

□

Proposition 1.1. *Soit $x^* \in \Omega$. Alors, $P\nabla f\left(x^*\right) = 0$ si et seulement si x^* satisfait à la condition de Lagrange.*

Démonstration. Avec Lemme 1.2, nous avons :

$$\begin{aligned} P\nabla f\left(x^*\right) = 0 &\Leftrightarrow \nabla f\left(x^*\right) \in Im\left(A^T\right) \\ &\Leftrightarrow \exists \lambda^* \in \mathbb{R}^m,\ \nabla f\left(x^*\right) + A^T\lambda^* = 0 \\ &\Leftrightarrow x^* \text{ satisfait la condition de Lagrange du problème (1.18).} \end{aligned}$$

□