Méthode du gradient

Hamza Ennaji

22 mars 2024

❈ Directions de descente

Définition 1. Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction de classe C^1. On dit qu'un vecteur non nul $\mathbf{d}$ de $\mathbb{R}^n$ est une direction de descente de f en x si

$$f'(x; \mathbf{d}) = \nabla f(x)^T \mathbf{d} < 0.$$

Il se trouve que suivre des directions de descentes avec un pas suffisamment petit, on arrive à décroitre la fonction objective. On a le résultat suivant :

Lemme 1. Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction de classe C^1 et $\mathbf{d}$ une direction de descente de f en $x \in \mathbb{R}^n$. Alors, il existe $\epsilon > 0$ tel que

$$f(x + t\mathbf{d}) < f(x),$$

pour tout $t \in (0, \epsilon]$.

Preuve. Laissée en exercice. □

D'une façon générale, une méthode de descente s'écrit comme suit

Méthode de descente

- Initialisation : On se donne un $x_0 \in \mathbb{R}^n$.
- Pour $k \geq 0$
  - Choisir une direction de descente $\mathbf{d}_k$.
  - Choisir un pas t_k tel que : $f(x + t_k\mathbf{d}) < f(x)$.
  - Prendre $x_{k+1} = x_k + t_k\mathbf{d}_k$.
  - Si un critère d'arrêt est vérifié, renvoyer x_{k+1}.

Remarque. — Le choix de la direction de descente donne lieu à des méthodes différentes de descentes (e.g., gradient, gradient conjugué, Newton, quasi-Newton etc).

— Il existe plusieurs critères d'arrêt, le plus utilisé est $\|\nabla f(x_{k+1})\| \leq \epsilon$ pour une tolérance $\epsilon > 0$ donnée.

— Le procédé permettant de trouver le pas t_k s'appelle recherche linéaire (line search en anglais). L'appellation vient du fait que ça revient à minimiser la fonction $g(t) = f(x_k + t\mathbf{d}_k)$. On discute ici quelques choix populaires.

1. **Pas constant** : dans ce cas $t_k = \bar{t}$ pour tout k. Bien que c'est un choix simple, un pas trop petit peut donner lieu à une convergence lente de l'algorithm tandis qu'un pas grand on peut perdre la décroissance de la fonction objective. (c.f. TP).
2. **Recherche linéaire exacte** : dans ce cas on cherche à minimiser exactement la fonction $g(t)$ définie ci-dessus. On choisit donc $t_k \in \operatorname{argmin}_{t\geq 0} f(x_k + t\mathbf{d}_k)$. Il se trouve qu'en général, on ne peut pas trouver le minimiseur exacte de g.
3. **Backtracking** : cette approche est permet d'éviter une recherche linéaire exacte de en choisissant convenablement un pas assurant la décroissance de la fonction objective. L'approche requiert trois paramètres $s > 0$ et $\alpha, \beta \in (0, 1)$. Dans un premier temps on prend $t_k = s$ et tant que

$$f(x_x) - f(x_k + t_k\mathbf{d}_k) < -\alpha t_k \nabla f(x_k)^T \mathbf{d}_k$$

on met à jour t_k par $t_k\beta$. Dans ce cas, le pas choisi serait $t_k = s\beta^{i_k}$ où i_k est le plus petit entier tel que

$$f(x_x) - f(x_k + t_k\mathbf{d}_k) \geq -\alpha s\beta^{i_k} \nabla f(x_k)^T \mathbf{d}_k$$

L'exemple suivant présente le cas d'une recherche linéaire exacte pour une fonction quadratique.

Exemple. On considère $f(x) = x^T Ax + 2b^T x + c$ avec $A \in \mathcal{S}_n^{++}, b \in \mathbb{R}^n, c \in R$. Soit $x \in \mathbb{R}^n$ et $\mathbf{d} \in \mathbb{R}^n \setminus \{0\}$ une direction de descente de f en x. Effectuer une recherche linéaire exacte revient à résoudre

$$\min_{t\geq 0} f(x+t\mathbf{d}) := g(t).$$

On a $g'(t) = t\mathbf{d}^T\nabla f(x+td)$ et comme $\nabla f(x) = 2(Ax+b)$, on trouve que $g'(t) = 2(\mathbf{d}^T A\mathbf{d})t + 2\mathbf{d}^T\nabla f(x)$. Donc

$$g^{'}(t) = 0 \Leftrightarrow t = -\frac{\mathbf{d}^T\nabla f(x)}{2\mathbf{d}^T A\mathbf{d}} > 0.$$

Comme $g^{''}(t) = 2(\mathbf{d}^T A\mathbf{d}) > 0$ alors le pas donné par la méthode de recherche linéaire exacte est $t^* = -\frac{\mathbf{d}^T\nabla f(x)}{2\mathbf{d}^T A\mathbf{d}}$.

※ Descente de gradient

※ La méthode

Comme son nom l'indique, le choix de la direction de descente dans la méthode du gradient est $d_k = -\nabla f(x_k)$. Ce choix est bien une direction de descente au sens de la Définition-1. En effet, pour $\nabla f(x_k) \neq 0$:

$$f^{'}(x_k; -\nabla f(x_k)) = -\nabla f(x_k)^T\nabla f(x_k) = -\|\nabla f(x_k)\|^2 < 0.$$

On présente ci-dessous les étapes de la méthode, avec comme critère d'arrêt $\|\nabla f(x_{k+1})\| \leq \epsilon$ pour une tolérance ϵ donnée.

Méthode du gradient

— Initialisation : On se donne un $x_0 \in \mathbb{R}^n$.

— Pour $k \geq 0$

  — Choisir un pas t_k par recherche linéaire sur $g(t) = f(x_k - t\nabla f(x_k))$.

  — Prendre $x_{k+1} = x_k - t_k\nabla f(x_k)$.

  — Si $\|\nabla f(x_{k+1})\| \leq \epsilon$ s'arrêter et renvoyer x_{k+1}.

※ Analyse de convergence

Avons de commencer notre analyse, rappelons qu'une fonction $f : \mathbb{R}^n \to \mathbb{R}$ est dans $C_L^{1,1}(\mathbb{R}^n)$ si elle est de classe C^1 est que son gradient $\nabla f : R^n \to \mathbb{R}^n$ est L-Lipschitz, i.e.,

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x-y\| \text{ pour tout } x, y \in \mathbb{R}^n.$$

Exemple. — On vérifie facilement que $f(x) = a^T x + b \in C_0^{1,1}(\mathbb{R}^n)$.

— Un fonction quadratique $f(x) = x^T Ax + 2b^T x + c$ avec $A \in \mathcal{S}_n, b \in \mathbb{R}^n$ et $c \in R$ est $2\|A\|$−Lipschitz. En effect, comme $\nabla f(x) = 2Ax + b$ on a

$$\|\nabla f(x) - \nabla f(y)\| = 2\|(Ax + b) - (Ay + b)\| \le 2\|A\|\|x - y\|.$$

Si la fonction est C^2 alors régularité Lipschitz du gradient est équivalente au fait que la Hessienne est bornée :

Théorème 1. Soit $f \in C^2(\mathbb{R}^n)$. Alors

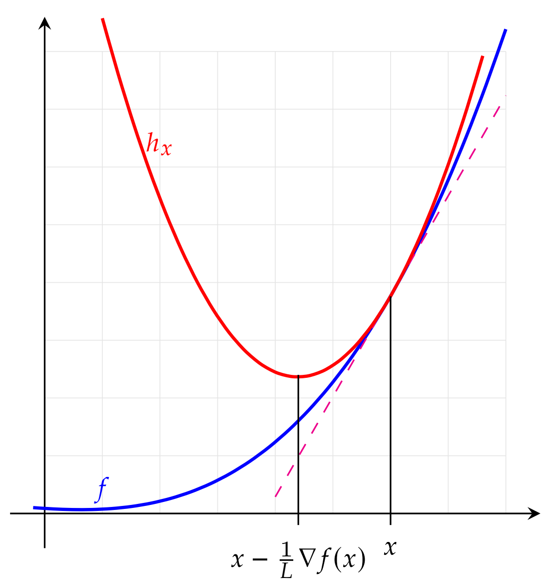
$$f \in C_L^{1,1}(\mathbb{R}^n) \Leftrightarrow \|\nabla^2 f(x)\| \le L \text{ pour tout } x \in \mathbb{R}^n.$$

Le résultat suivant joue un rôle important dans la suite.

Lemme 2 (Lemme de descente)**.** Soit $f \in C_L^{1,1}(\mathbb{R}^n)$. Alors pour tout $x, y \in \mathbb{R}^n$

$$f(y) \le f(x) + \nabla f(x)^T (y - x) + \frac{L}{2}\|x - y\|^2.$$

Preuve. Voir TD. □



Fixons $x \in \mathbb{R}^n$ et considérons $h_x(y) = f(x) + \nabla f(x)^T (y - x) + \frac{L}{2}\|x - y\|^2$ qui est quadratique en y et majore f : $f(y) \le h_x(y)$ pour tout $y \in \mathbb{R}^n$. En plus, le minimum de h_x est atteint en $x^* = x - \frac{1}{L}\nabla f(x)$ (see Figure-??).

Comme conséquence du Lemme-2, on a

Lemme 3. Soit $f \in C_L^{1,1}(\mathbb{R}^n)$. Alors pour tout $x \in \mathbb{R}^n$ et $t > 0$

$$f(x) - f(x - t\nabla f(x)) \geq t\left(1 - \frac{tL}{2}\right)\|\nabla f(x)\|^2. \tag{1}$$

Preuve. On applique le Lemme de descente avec $y = x - t\nabla f(x)$ pour un $x \in \mathbb{R}^n$. On a donc

$$f(x - t\nabla f(x)) \leq f(x) + t\nabla f(x)^T \nabla f(x) + \frac{Lt^2}{2}\|\nabla f(x)\|^2.$$

□

Si nous optons pour une stratégie avec un pas constant, i.e., $t_k = t_* \in (0, 2/L)$ pour tout k, on a alors d'après Lemme-3

$$f(x_k) - f(x_{k+1}) \geq t^*\left(1 - \frac{t^*L}{2}\right)\|\nabla f(x_k)\|^2.$$

La fonction $t \mapsto t\left(1 - \frac{tL}{2}\right)$ sur $(0, 2/L)$ atteint un maximum en $t^* = 1/L$. Pour ce choix on a $x_{k+1} = x_k - \frac{1}{L}\nabla f(x_k))$ et

$$f(x_k) - f(x_{k+1}) \geq \frac{1}{2L}\|\nabla f(x_k)\|^2.$$

Pour une recherche linéaire exacte, i.e., $x_{k+1} = x_k - t_k\nabla f(x_k)$ avec

$$t_k \in \text{argmin}_{t\geq 0} f\Big(x_k - t\nabla f(x_k)\Big),$$

on remarque que $f(x_k - t_k\nabla f(x_k)) \leq f\left(x_k - \frac{1}{L}\nabla f(x_k)\right)$ soit

$$f(x_k) - f(x_{k+1}) \geq f(x_k) - f\left(x_k - \frac{1}{L}\nabla f(x_k)\right) \geq \frac{1}{2L}\|f(x_k)\|^2.$$

Dans le cade de beckstracking, on cherche pour $\alpha \in (0, 1)$, un pas t_k suffisamment petit tel que

$$f(x_k) - f(x_{k+1}) \geq \alpha t_k\|f(x_k)\|^2.$$

Dans ce cas deux options se présentent : prendre $t_k = s$ où $s > 0$ est la valeur initiale du pas, soit prendre t_k avec la méthode de backtracking comme décrit dans la remarque-, et dans ce cas, le choix $t_k = t_k/\beta$ ne serait pas admissible, i.e.,

$$f(x_k) - f\left(x_k - \frac{t_k}{\beta}\nabla f(x_k)\right) < \frac{\alpha t_k}{\beta}\|f(x_k)\|^2. \tag{2}$$

En appliquant (15) avec $x = x_k$ et $t = \frac{t_k}{\beta}$ on obtient

$$f(x_k) - f\left(x_k - \frac{t_k}{\beta}\nabla f(x_k)\right) \geq \frac{t_k}{\beta}\left(1 - \frac{t_k L}{\beta}\right)\|\nabla f(x)\|^2. \tag{3}$$

Les équations (2)-(3) impliquent que $\frac{t_k}{\beta}\left(1 - \frac{t_k L}{\beta}\right) < \frac{\alpha t_k}{\beta}$, soit $t_k > \frac{2(\alpha-1)\beta}{L}$. Finalement, pour la méthode de backtracking, on a

$$f(x_k) - f\left((x_k - t_k \nabla f(x_k)\right) \geq \alpha \min\left(s, \frac{2(\alpha-1)\beta}{L}\right)\|f(x_k)\|^2.$$

On récapitule cette discussion dans le résultat suivant :

Proposition 1. Soit $f \in C_L^{1,1}(\mathbb{R}^n)$ et $(x_k)_k$ la suite générée par la méthode du gradient pour résoudre

$$\min_{x \in \mathbb{R}^n} f(x),$$

avec pas contant, recherche linéaire exacte et becktracking de paramètres $s > 0, \alpha, \beta \in (0, 1)$. Alors

$$f(x_k) - f\left(x_k - \frac{t_k}{\beta}\nabla f(x_k)\right) \geq M\|\nabla f(x)\|^2, \tag{4}$$

avec

$$M = \begin{cases} t^*\left(1 - \frac{t^* L}{2}\right) & \text{pas constant,} \\ \frac{1}{2L} & \text{recherche linéaire exacte,} \\ \alpha \min\left(s, \frac{2(\alpha-1)\beta}{L}\right) & \text{backtracking.} \end{cases}$$

On démontre maintenant que $\|\nabla f(x_k)\| \to 0$ quand $k \to \infty$.

Théorème 2. Soit $f \in C_L^{1,1}(\mathbb{R}^n)$ et $(x_k)_k$ la suite générée par la méthode du gradient pour résoudre

$$\min_{x \in \mathbb{R}^n} f(x),$$

avec pas contant, recherche linéaire exacte et becktracking de paramètres $s > 0, \alpha, \beta \in (0, 1)$. Supposons que f est minorée, i.e., il existe $c > 0$ tel que $f(x) > c$ pour tout $x \in \mathbb{R}^n$. Alors

1. La suite $(f(x_k))_k$ est décroissante, avec $f(x_{k+1}) < f(x_k)$ sauf si $\|\nabla f(x_k)\| = 0$.
2. $\|f(x_k)\| \to 0$ quand $k \to \infty$.

Preuve. 1. On a grâce à (4)

$$f(x_k) - f\left(x_k - \frac{t_k}{\beta}\nabla f(x_k)\right) \geq M\|\nabla f(x)\|^2 \geq 0, \tag{5}$$

pour $M > 0$. Donc $(f(x_k))_k$ est décroissante et $f(x_k) = f(x_{k+1})$ ne peut avoir lieu que si $\nabla f(x_k) = 0$.

2. Comme la suite $(f(x_k))_k$ est décroissante et f minorée, alors elle est convergente, i.e., $f(x_k) - f(x_{k+1}) \underset{k\to\infty}{\longrightarrow} 0$. Cela implique avec (5) que $\|\nabla f(x_k)\| \to 0$ quand $k \to \infty$.

□

On termine cette section avec le résultat suivant permettant d'avoir des taux de convergence de la norme du gradient.

Théorème 3. Sous les même hypothèses, on a pour tout $n \geq 0$

$$\min_{k=0,\ldots,n} \|\nabla f(x_k)\| \leq \sqrt{\frac{f(x_0) - f^*}{M(n+1)}}, \tag{6}$$

où f^* est la limite de la suite $(f(x_k))_k$ et la constante M donnée dans Proposition-1.

Preuve. En sommant de 0 à n dans l'inégalité (4), on obtient

$$f(x_0) - f(x_{n+1}) = \sum_{k=0}^{n} f(x_k) - f(x_{k+1}) \geq M \sum_{k=0}^{n} \|\nabla f(x_k)\|^2.$$

Par définition, $f^* \leq f(x_{k+1})$, et par la suite

$$f(x_0) - f^* \geq M \sum_{k=0}^{n} \|\nabla f(x_k)\|^2 \geq M \min_{k=0,\ldots,n} \|\nabla f(x_k)\|^2 \sum_{k=0}^{n} 1\,,$$

soit

$$f(x_0) - f^* \geq M(n+1) \min_{k=0,\ldots,n} \|\nabla f(x_k)\|^2.$$

Finalement $\min_{k=0,\ldots,n} \|\nabla f(x_k)\| \leq \sqrt{\frac{f(x_0)-f^*}{M(n+1)}}$, comme voulu. □

❊ Méthode de Newton

Avant de présenter la méthode de Newton comme une méthode de décente, rappelons la méthode de Newton pour trouver les zéros d'une fonction réelle. Soit donc $f : \mathbb{R} \to \mathbb{R}$, on cherche à résoudre l'équation $f(t) = 0$. L'idée est de remplacer f par son approximation de premier ordre en un point t_0, i.e., résoudre $\tilde{f}(t) = f(t_0) + (t - t_0)f'(t_0) = 0$ au lieu de $f(t) = 0$. Si $f'(t_0) \neq 0$, on obtient $t = t_0 - \frac{f(t_0)}{f'(t_0)}$. Les itérations de Newton s'écrivent

$$t_{k+1} = t_k - \frac{f(t_k)}{f'(t_k)}.$$

Maintenant, soit $g : \mathbb{R}^n \to \mathbb{R}^n$. On s'intéresse au système linéaire $g(x) = 0$. De la même manière, en approchant g à l'ordre un autour d'un point $x_0 \in \mathbb{R}^n$, i.e., $g(x_0) + \nabla g(x_0)(x - x_0) = 0$. On obtient

$$\tilde{g}(x) := g(x_0) + \nabla g(x_0)(x - x_0) = 0 \Rightarrow x = x_0 - \Big(\nabla g(x_0)\Big)^{-1} g(x_0).$$

L'équation ci-dessus suppose évidemment que $\det\Big(\nabla g(x_0)\Big) \neq 0$. De même, on peut construire une suite $(x_k)_k$ qui approchera, à priori le zéro de g, en définissant

$$x_{k+1} = x_k - \Big(\nabla g(x_k)\Big)^{-1} g(x_k). \tag{7}$$

Maintenant, revenons à notre problème d'optimisation

$$\min_{x \in \mathbb{R}^n} f(x). \tag{8}$$

On s'intéresse ici au système linéaire $\nabla f(x) = 0$. Rappelons que lorsque f est convexe, la conditions $\nabla f(x^*) = 0$ est nécessaire et suffisante pour dire que x^* minimise f. On applique (7) avec $g = \nabla f$, pour trouver

$$x_{k+1} = x_k - \Big(\nabla^2 f(x_k)\Big)^{-1} \nabla f(x_k). \tag{9}$$

Ceci est une façon pour faire le lien entre Newton et Newton...

Une autre façon pour retrouver (12) est la suivante. Supposons que f est C^2 et que $\nabla^2 f(x) \in \mathcal{S}_n^{++}$ pour tout $x \in \mathbb{R}^n$. On cherchera donc à minimiser l'approximation d'ordre deux de f autour d'un certain $x_0 \in \mathbb{R}^n$, i.e., remplacer f

$$\tilde{f}(x) = f(x_0) + (x - x_0)^T \nabla f(x_0) + \frac{1}{2}(x - x_0)\nabla^2 f(x)(x - x_0) \tag{10}$$

dans le problème (8). Donc étant donnée une itérée x_k, x_{k+1} est obtenue comme suit

$$x_{k+1} = \underset{x\in\mathbb{R}^n}{\operatorname{argmin}} \left\{ f(x_k) + (x - x_k)^T \nabla f(x_k) + \frac{1}{2}(x - x_k)\nabla^2 f(x)(x - x_k) \right\}. \tag{11}$$

L'unique minimiseur de (11) est en fait l'unique point stationnaire, et on a

$$\nabla \tilde{f}(x_{k+1}) = 0 \Leftrightarrow \nabla f(x_k) + \nabla^2 f(x_k)(x_{k+1} - x_k) = 0,$$

ce qui donne

$$x_{k+1} = x_k - \left(\nabla^2 f(x_k)\right)^{-1} \nabla f(x_k). \tag{12}$$

La méthode de Newton est donc une méthode de descente avec comme direction $d_k = -\left(\nabla^2 f(x_k)\right)^{-1} \nabla f(x_k)$.

Il se trouve que sous certaines conditions de régularité sur la Hessienne, on peut obtenir autour de la solution optimale un taux de convergence quadratique.

Théorème 4. Soit $f \in C^2(\mathbb{R}^n)$ et supposons que :

1. il existe $m > 0$ tel que $\nabla^2 f(x) \geq mI_n$ pour tout $x \in \mathbb{R}^n$,
2. il existe $L > 0$ tel que

$$\|\nabla^2 f(x) - \nabla^2 f(y)\| \leq L\|x - y\|$$

pour tout $x, y \in \mathbb{R}^n$.

Alors

$$\|x_{k+1} - x^*\| \leq \frac{L}{2m}\|x_k - x^*\|^2,$$

où $(x_k)_k$ est la suite générée par la méthode de Newton et x^* l'unique minimiseur de f sur $\mathbb{R}^n$. De plus, si $\|x_0 - x^*\| \leq m/L$, alors

$$\|x_k - x^*\| \leq \frac{2m}{L}\left(\frac{1}{2}\right)^{2^k}, \quad k \geq 0.$$

Preuve. Laissée en exercice. □

※ Méthode de gradient projeté

6.1 Cas général

On s'intéresse ici à des problèmes du type

$$\min_{x\in C} f(x), \tag{13}$$

où $C \subset \mathbb{R}^n$ est un convexe fermé et $f : C \to \mathbb{R}$ de classe C^1. Il se trouve qu'on peut caractériser les points stationnaires de (16) en terme de l'opérateur de projection orthogonale.

Théorème 5. Soit $f : C \to \mathbb{R}$ de classe C^1 où C est un convexe fermé et $s > 0$. Alors x^* est un point stationnaire de (16) si et seulement si

$$x^* = P_C(x^* - s\nabla f(x^*)). \tag{14}$$

Preuve. Laissée en exercice. □

L'équation (14) affirme que x^* est point fixe de l'application $x \mapsto P_C(x - s\nabla f(x))$.. Cela suggère des algorithmes du type point fixe pour résoudre (14), et donc (16).

Méthode du gradient projeté

On se donne une tolérance $\epsilon > 0$.

- Initialisation : On se donne un $x_0 \in C$.
- Pour $k \geq 0$
  - Choisir un pas t_k par recherche linéaire.
  - Prendre $x_{k+1} = P_C(x_k - t_k\nabla f(x_k))$.
  - Si $\|x_k - x_{k+1}\| \leq \epsilon$, renvoyer x_{k+1}.

On démontre un résultat similaire au Lemme-3.

Lemme 4. Soit $f \in C_L^{1,1}(\mathbb{R}^n)$. Alors pour tout $x \in \mathbb{R}^n$ et $t > 0$

$$f(x) - f(P_C(x - t\nabla f(x))) \geq t\left(1 - \frac{tL}{2}\right)\left\|\nabla \frac{1}{t}(x - P_C(x - t\nabla f(x)))\right\|^2. \tag{15}$$

Preuve. Il suffit d'appliquer le Lemme-2 avec $y = P_C(x - t\nabla f(x))$ et utiliser le fait que

$$\langle x - t\nabla f(x) - y, x - y\rangle \leq 0.$$

□

Remarque. Quand $C = \mathbb{R}^n$, la méthode du gradient projeté n'est rien d'autre que la méthode de descente du gradient classique présentée en Section-3.

Notation. Par la suite, on note, pour $M > 0$

$$G_M(x) = M\left(x - P_C(x - \frac{1}{M}\nabla f(x)\right).$$

On note que pour $C = \mathbb{R}^n$, $G_M(x) = \nabla f(x)$.. L'opérateur servira comme une mesure d'optimalité. Avec cette notation, le Lemme-4 se réécrit

$$f(x) - f(P_C(x - t\nabla f(x))) \geq t\left(1 - \frac{tL}{2}\right)\|G_{1/t}(x)\|^2.$$

On peut démontrer sans difficultés un résultat similaire au Théorème-2 pour la méthode du gradient projeté, on changera $\nabla f(x)$ par $G_M(x)$.

6.2 Cas convexe

On s'intéresse ici à des problèmes du type

$$\min_{x \in C} f(x), \tag{16}$$

où $C \subset \mathbb{R}^n$ est un convexe fermé et $f \in C_L^{1,1}(C)$ est convexe. Dans ce cas, on peut démontrer à la fois la convergence de $(f(x_k))_k$ vers f^* ainsi que la suite $(x_k)_k$.

Théorème 6. Soit $f \in C_L^{1,1}(C)$ une fonction convexe avec C un convexe fermé de $\mathbb{R}^n$. Soit $(x_k)_k$ la suite générée par la méthode du gradient projeté avec un pas $t_k = t^* \in (0, 1/L]$. Supposons que $\operatorname{argmin}_C f := S \neq \emptyset$. Alors

1. Pour tout $k \geq 0$ et $x^* \in S$
$$f(x_k) - f^* \leq \frac{\|x_0 - x^*\|^2}{2t^* k}.$$
2. $x_k \to x^* \in S$ lorsque $k \to \infty$.