Travaux dirigés analyse convexe

Exercice 1 *(Inégalité de Taylor-Lagrange)*
Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction dérivable dont le gradient est L-lipschitzien, c'est-à-dire que $\|\nabla f(y) - \nabla f(x)\| \leq L\|y - x\|$ pour tous x, y. Ici, nous considérerons la norme Euclidienne.

1. *Montrer que pour tous x, y, $\langle \nabla f(y) - \nabla f(x), y - x\rangle \leq L\|y - x\|^2$.*

   ▶ On utilise l'inégalité de Cauchy-Schwartz :

$$\begin{aligned}\langle \nabla f(y) - \nabla f(x), y - x\rangle &\leq |\langle \nabla f(y) - \nabla f(x), y - x\rangle| \leq \|\nabla f(y) - \nabla f(x)\| . \|y - x\| \\ &\leq L \|y - x\|^2\end{aligned}$$

2. *On définit $\varphi(t) = f(x + t(y - x))$ pour tout $t \in [0, 1]$. Montrer que*

$$f(y) - f(x) - \langle \nabla f(x), y - x\rangle = \varphi(1) - \varphi(0) - \varphi'(0) .$$

   ▶ Il est clair que $\varphi(0) = f(x)$ et $\varphi(1) = f(y)$. Notez que $\varphi(t) = f(g(t))$ où $g(t) = x + t(y - x)$. Par le théorème de dérivation des fonctions composées,

$$\varphi'(t) = \langle \nabla f(g(t)), g'(t)\rangle = \langle \nabla f(x + t(y - x)), y - x\rangle \tag{1}$$

   Ainsi $\varphi'(0) = \langle \nabla f(x), y - x\rangle$. En combinant ces trois inégalités, on obtient

$$\varphi(1) - \varphi(0) - \varphi'(0) = f(y) - f(x) - \langle \nabla f(x), y - x\rangle .$$

3. *En déduire que*

$$f(y) - f(x) - \langle \nabla f(x), y - x\rangle = \int_0^1 \langle \nabla f(x + t(y - x)) - \nabla f(x), y - x\rangle dt$$

   ▶ Comme φ est une primitive de φ',

$$\varphi(1) = \varphi(0) + \int_0^1 \varphi'(t)dt.$$

   Ainsi, en utilisant (1)

$$\begin{aligned}f(y) - f(x) - \langle \nabla f(x), y - x\rangle &= \varphi(1) - \varphi(0) - \varphi'(0) = \int_0^1 \varphi'(t)dt - \varphi'(0) \\ &= \int_0^1 \langle \nabla f(x + t(y - x)), y - x\rangle dt - \int_0^1 \langle \nabla f(x), y - x\rangle dt \\ &= \int_0^1 \langle \nabla f(x + t(y - x)) - \nabla f(x), y - x\rangle dt\end{aligned}$$

4. *En utilisant la première question, montrer que*

$$f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{L}{2} \|y - x\|^2 .$$

▶ On sait que pour tous x, z, $\langle \nabla f(z) - \nabla f(x), z - x \rangle \leq L\|z - x\|^2$. On utilise cette inégalité avec $z = x + t(y - x)$: $\langle \nabla f(x + t(y - x)) - \nabla f(x), t(y - x) \rangle \leq L\|t(y - x)\|^2$. En divisant par $t > 0$ et en intégrant entre 0 et 1, on obtient

$$\int_0^1 \langle \nabla f(x+t(y-x)) - \nabla f(x), y-x \rangle dt \leq \int_0^1 tL \|y - x\|^2 dt = \left[\frac{t^2}{2}\right]_0^1 L \|y - x\|^2 = \frac{L}{2} \|y - x\|^2$$

On conclut grâce à la question 3.

5. *On définit la fonction*

$$g_x : y \mapsto f(x) + \langle \nabla f(x), y - x \rangle + \frac{L}{2} \|y - x\|^2$$

Montrer que g_x a un unique minimiseur $\hat{x}$ et calculer ce minimiseur.

▶ g_x est fortement convexe donc a un unique minimiseur.
Il vérifie $\nabla g_x(\hat{x}) = 0$, c'est-à-dire

$$\nabla f(x) + L(\hat{x} - x) = 0.$$

On trouve $\hat{x} = x - \frac{1}{L} \nabla f(x)$.

6. *Montrer que $f(\hat{x}) \leq f(x) - \frac{1}{2L} \|\nabla f(x)\|^2$.*

▶ L'inégalité de Taylor-Lagrange nous donne

$$\begin{aligned} f(\hat{x}) = f(x - \frac{1}{L}\nabla f(x)) &\leq f(x) + \langle \nabla f(x), -\frac{1}{L}\nabla f(x) \rangle + \frac{L}{2} \left\| -\frac{1}{L} \nabla f(x) \right\|^2 \\ &\leq f(x) - \frac{1}{2L} \|\nabla f(x)\|^2 \end{aligned}$$

7. *Proposer un algorithme pour minimiser la fonction f.*

▶ L'algorithme du gradient défini par récurrence par $x_0 \in \mathbb{R}^n$ et

$$x_{k+1} = x_k - \frac{1}{L} \nabla f(x_k)$$

On vient de prouver que pour tout $k \geq 0$

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

Exercice 2 (Étude d'une fonction indicatrice convexe)
Soit $\iota_{\mathbb{R}_+}$ l'indicatrice convexe de $\mathbb{R}_+$ définie par

$$\iota_{\mathbb{R}_+}(x) = \begin{cases} 0 & \textit{si } x \geq 0 \\ +\infty & \textit{si } x < 0 \end{cases}$$

1. *Vérifier que $\iota_{\mathbb{R}_+}$ est bien une fonction convexe.*
   ▶ Soient $x, y \in \mathbb{R}$ et $\alpha \in]0, 1[$.
   Si $x \geq 0$ et $y \geq 0$, $\alpha x + (1-\alpha)y \geq 0$ et donc
   $\iota_{\mathbb{R}_+}(\alpha x + (1-\alpha)y) = 0 \leq 0 = \alpha\iota_{\mathbb{R}_+}(x) + (1-\alpha)\iota_{\mathbb{R}_+}(y)$.
   Si $x < 0$ ou $y < 0$, alors $\alpha\iota_{\mathbb{R}_+}(x) + (1-\alpha)\iota_{\mathbb{R}_+}(y) = +\infty$ et donc
   $\iota_{\mathbb{R}_+}(\alpha x + (1-\alpha)y) \leq +\infty = \alpha\iota_{\mathbb{R}_+}(x) + (1-\alpha)\iota_{\mathbb{R}_+}(y)$.
   Dans tout les cas, l'inégalité de convexité est vérifiée donc $\iota_{\mathbb{R}_+}$ est convexe.
2. *Montrer que si $x < 0$, alors $\partial\iota_{\mathbb{R}_+}(x) = \emptyset$.*
   ▶ Soit $x < 0$ et soit $q \in \partial\iota_{\mathbb{R}_+}(x)$.
   Alors $\forall y \in \mathbb{R}$, $\iota_{\mathbb{R}_+}(y) \geq \iota_{\mathbb{R}_+}(x) + q(y-x)$.
   En appliquant cette inégalité en $y = 0$, on trouve $0 \geq +\infty + qx$, ce qui est impossible.
   On en conclut grâce à un raisonnement par l'absurde qu'il n'existe pas de q dans $\partial\iota_{\mathbb{R}_+}(x)$.
3. *Montrer que si $x \geq 0$, alors*

   $$\partial\iota_{\mathbb{R}_+}(x) = \{q \in \mathbb{R} \ : \ \forall z \geq 0, q(z-x) \leq 0\}.$$

   Cela montre que $\partial\iota_{\mathbb{R}_+}(x)$ est le cône normal à $\mathbb{R}_+$ en x.
   ▶ Soit $x \geq 0$.

   $$\begin{aligned} q \in \partial\iota_{\mathbb{R}_+}(x) &\Leftrightarrow \forall y \in \mathbb{R}, \iota_{\mathbb{R}_+}(y) \geq \iota_{\mathbb{R}_+}(x) + q(y-x) \\ &\Leftrightarrow \forall y \in \mathbb{R}, \iota_{\mathbb{R}_+}(y) \geq q(y-x) \\ &\Leftrightarrow \forall y \in \mathbb{R}_+, 0 \geq q(y-x) \end{aligned}$$

   La dernière équivalence vient du fait que si $y \geq 0$, l'inégalité est triviale.
4. *En déduire que*

   $$\partial\iota_{\mathbb{R}_+}(x) = \begin{cases} \emptyset & \textit{si } x < 0 \\ \mathbb{R}_- & \textit{si } x = 0 \\ \{0\} & \textit{si } x > 0 \end{cases}$$

   ▶ Le cas $x < 0$ a déjà été traité.
   Supposons que $x = 0$.

   $$q \in \partial\iota_{\mathbb{R}_+}(0) \Leftrightarrow \forall y \in \mathbb{R}_+, 0 \geq qy$$

En prenant $y = 1$, on en déduit que $q \leq 0$. Réciproquement, si $q \leq 0$, alors pour tout $y \geq 0$, $qy \leq 0$. Ainsi $q \in \partial\iota_{\mathbb{R}_+}(0) \Leftrightarrow q \leq 0$.
Maintenant, soit $x > 0$.

$$q \in \partial\iota_{\mathbb{R}_+}(0) \Leftrightarrow \forall y \in \mathbb{R}_+, 0 \geq q(y - x)$$

En prenant $y = 0$, on en déduit que $0 \geq -qx$. En prenant $y = 2x$, on en déduit que $0 \geq qx$. Ainsi $qx = 0$ et donc $q = 0$. Réciproquement, $q = 0$ convient.

5. *Soit* $\operatorname{prox}_{\iota_{\mathbb{R}_+}}(x) = \arg\min_y \iota_{\mathbb{R}_+}(y) + \frac{1}{2}(x-y)^2$ *l'opérateur proximal de* $\iota_{\mathbb{R}_+}$.
Montrer que $\operatorname{prox}_{\iota_{\mathbb{R}_+}}(x) = \max(0, x)$.
▶ Dans ce cas-ci, l'opérateur proximal de $\iota_{\mathbb{R}_+}$ est la projection sur $\mathbb{R}_+$.
On peut vérifier que $\max(0, x)$ vérifie bien la règle de Fermat :

$$\partial\iota_{\mathbb{R}_+}\big(\max(0,x)\big) + \max(0,x) - x = \begin{cases} \mathbb{R}_- + 0 - x =]-\infty, -x] & \text{si } x \leq 0 \\ \{0\} + x - x = \{0\} & \text{si } x > 0 \end{cases}$$

et donc dans les deux cas, $0 \in \partial\iota_{\mathbb{R}_+}\big(\max(0,x)\big) + \max(0,x) - x$

Exercice 3 (Vitesse de convergence de la descente de gradient)
Soit fonction $f : \mathbb{R}^n \to \mathbb{R}$ *convexe et dérivable avec un gradient L-lipschitzien. On suppose que* f *a au moins un minimiseur* x_*. *On considère l'algorithme du gradient défini de manière récursive par*

$$x_{k+1} = x_k - \frac{1}{L}\nabla f(x_k).$$

1. *Montrer que pour tout* $z \in \mathbb{R}^n$,

$$\frac{L}{2}\|x_k - z\|^2 = \frac{L}{2}\|x_{k+1} - z\|^2 + \frac{L}{2}\|x_k - x_{k+1}\|^2 + \langle \nabla f(x_k), x_{k+1} - x_k\rangle - \langle \nabla f(x_k), z - x_k\rangle.$$

▶ Soit $z \in \mathbb{R}^n$.

$$\begin{aligned} \frac{L}{2}\|x_k - z\|^2 &= \frac{L}{2}\|x_k - x_{k+1} + x_{k+1} - z\|^2 \\ &= \frac{L}{2}\|x_{k+1} - x_k\|^2 + \frac{L}{2}\|x_{k+1} - z\|^2 + L\langle x_k - x_{k+1}, x_{k+1} - z\rangle \\ &= \frac{L}{2}\|x_{k+1} - x_k\|^2 + \frac{L}{2}\|x_{k+1} - z\|^2 + \langle \nabla f(x_k), x_{k+1} - z\rangle \end{aligned}$$

On a utilisé le fait que $L(x_k - x_{k+1}) = \nabla f(x_k)$.

2. *Montrer que*

$$f(x_{k+1}) \leq f(x_k) + \langle \nabla f(x_k), x_\star - x_k\rangle + \frac{L}{2}\|x_k - x_*\|^2 - \frac{L}{2}\|x_{k+1} - x_*\|^2.$$

▶ Comme f a un gradient L-lipschitzien,

$$f(x_{k+1}) \leq f(x_k) + \langle \nabla f(x_k), x_{k+1} - x_k \rangle + \frac{L}{2} \|x_{k+1} - x_k\|^2$$

D'après ce qu'on vient de montrer à la question précédente, on a pour tout z,

$$f(x_{k+1}) \leq f(x_k) + \langle \nabla f(x_k), z - x_k \rangle + \frac{L}{2} \|x_k - z\|^2 - \frac{L}{2} \|x_{k+1} - z\|^2$$

En appliquant ce résultat en $z = x_\star$, on a ce qui était demandé.

3. *En déduire que pour tout $k \geq 1$,*

$$\sum_{i=1}^{k} f(x_i) \leq k f(x_*) + \frac{L}{2} \|x_0 - x_*\|^2 .$$

▶ Comme f est convexe et dérivable, $f(x_\star) \geq f(x_k) + \langle \nabla f(x_k), x_\star - x_k \rangle$. On obtient que pour tout i

$$f(x_{i+1}) \leq f(x_\star) + \frac{L}{2} \big(\|x_i - x_\star\|^2 - \|x_{i+1} - x_\star\|^2 \big)$$

En sommant cette inégalité pour i allant de 0 à $k-1$, on obtient

$$\sum_{i=0}^{k-1} f(x_{i+1}) \leq k f(x_\star) + \sum_{i=0}^{k-1} \frac{L}{2} \big(\|x_i - x_\star\|^2 - \|x_{i+1} - x_\star\|^2 \big)$$

Mais la somme de droite est une somme téléscopique. On a donc

$$\sum_{i=1}^{k} f(x_i) \leq k f(x_\star) + \frac{L}{2} \|x_0 - x_\star\|^2 - \frac{L}{2} \|x_k - x_\star\|^2$$

On obtient le résultat demandé en remarquant que $-\frac{L}{2} \|x_k - x_\star\|^2 \leq 0$.

4. *Montrer que*

$$f(x_k) - f(x_*) \leq \frac{L\|x_0 - x_\star\|^2}{2k} .$$

▶ On a montré dans l'exercice 1 que pour tout k,

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

Ainsi, $\sum_{i=1}^{k} f(x_i) \geq k f(x_k)$. En divisant par k l'inégalité de la question 3, on obtient la vitesse de convergence de l'objectif lorsqu'on utilise la méthode du gradient sur une fonction convexe.