COURS D'OPTIMISATION

ISIMA – F4 3ème année – Master Recherche Maths

Jonas Koko

ISIMA

Plan

Motivations

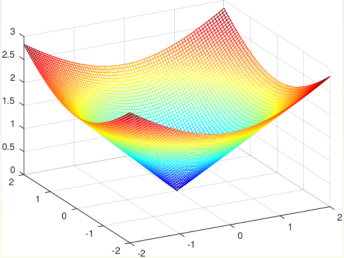
Régularisations de problèmes de type "moindres carrés"

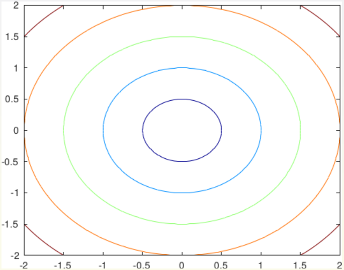
Applications :

- Programmation quadratique
- Restauration d'images
- least absolute deviations $\| Ax - b \|_1$
- ℓ_1 Loss minimization $\min l(x) + \lambda \| x \|_1$
- LASSO $\| Ax - b \|_2^2 + \lambda \| x \|_1$
- ...

Norme $\| \cdot \|_2$

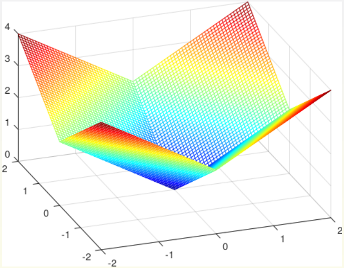
$$f(x) = \| x \|_2$$

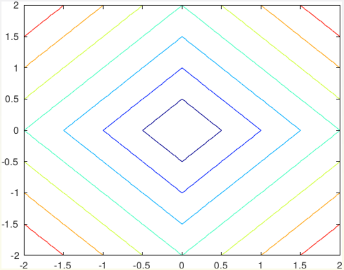




Norme $\| \cdot \|_1$

$$f(x) = \| x \|_1$$





Quelques Rappels

Définition (Boule unité)

$$B = \{x \in \mathbb{R}^n \mid \| x \| \leq 1\}$$

Définition

C un ensemble

- *Fermeture :* $cl(C) = \bigcap_{\epsilon > 0} (C + \epsilon B)$
- *Intérieur :* $int(C) = \{x \in C \mid \exists \varepsilon > 0;\ x + \epsilon B \subset C\}$
- *C est un cône si* $\forall x \in C,\ \lambda x \in C,\ \forall \lambda \geq 0.$

Rappels II

Définition (Hyperplan)

Dans $\mathbb{R}^n$, un sous-espace de dimension $n-1$. Par exemple,

$$x_i = \sum_{j \neq i} a_j x_j$$

Définition (Opérateur monotone)

A est monotone si

$$\langle A(u) - A(v), u - v \rangle \geq 0, \quad \forall u, v$$

Si A est linéaire

$$\langle A(u), u \rangle \geq 0, \quad \forall u$$

Si A est une matrice $\Longleftrightarrow$ A sémi définie positive

Ensemble convexe

Définition (Ensemble convexe)

C est convexe si $\lambda x + (1-\lambda)y \in C$, $\forall x, y \in C$ et $\forall \lambda \in (0,\ 1)$.

Exemples :

- Les boules $B(x_0, r) = \{x\ :\ \| x - x_0 \| \leq r\}$
- Cônes
- Sous-espaces vectoriels, sous-espaces affines

Propriétes :

- $C_1,\ C_2$ convexes, alors $C_1 \cap C_2$ convexe
- $C_1,\ C_2$ convexes, alors $\alpha_1 C_1 + \alpha_1 C_2$ convexe, $\forall \alpha_1, \alpha_2$

Définition (Combinaison convexe)

$$\sum^{m} \lambda_i x_i, \quad \lambda_i \geq 0, \quad \sum^{m} \lambda_i = 1.$$

Ensembles convexes II

Définition (Enveloppe convexe)

S ensemble convexe

$$conv(S) = \left\{ x \in \mathbb{R}^n, \ : \ x = \sum_{i=1}^{m} \lambda_i a_i, \ \lambda_i \geq 0, \ \sum_{i=1}^{m} \lambda_i = 1, \ a_i \in S \right\}$$

$conv(S)$ est convexe.

Projection

Définition (Projection)

C ensemble convexe de $\mathbb{R}^n$, $u \in \mathbb{R}^n$. Le vecteur p est meilleure approximation de u sur C si

$$\| u - p \|_2 = \min_{x \in C} \| u - x \|_2$$

Théorème

C convex fermé non vide. $u \in \mathbb{R}^n$ possède un unique p tel que

$$\begin{aligned} \| u - p \|_2 &= \min_{x \in C} \| u - x \|_2 \\ \langle p - u, p - x \rangle &\leq 0, \quad \forall x \in C. \end{aligned}$$

Projection II

Théorème (monotonie)

φ_C projection sur C alors

$$
\begin{aligned}
\langle \varphi_C(u) - \varphi_C(v), u - v \rangle &\geq 0, \ \forall u, v \in C \textit{(monotone)} \\
\| \varphi_C(u) - \varphi_C(v) \| &\leq \| u - v \|, \ \forall u, v \in C \textit{(non-expansive)}
\end{aligned}
$$

Fonctions convexes

Définition

- *Domaine :* $Dom(f) = \{x \mid f(x) < +\infty\}$
- f *propre si* $Dom(f) \neq \emptyset$ *et* $f(x) > -\infty$
- *Epigraphe* $epi(f) = \{(x,z) \in \mathbb{R}^{n+1} \mid f(x) \leq z\}$

Définition (Fonction convexe)

$f \ : \ C \to \mathbb{R}$ *convexe si* $\forall\lambda \in (0,1)$

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y), \quad \forall x, y \in C.$$

f *fortement convexe si* $\exists\alpha > 0$

$$f(\lambda x+(1-\lambda)y) \leq \lambda f(x)+(1-\lambda)f(y)-\frac{\alpha}{2}\lambda(1-\lambda) \parallel x-y \parallel^2, \quad \forall x, y \in C.$$

Exemples de fonctions convexes

Remarque

Si f convexe, la fonction $g = -f$ est concave

- Fonctions affines (linéaires)
- $f(x) = (1/2x^\top Qx - b^\top x$, Q semi-définie positive
- $f(x) = x^2$
- $f(x) = e^x$
- $f(x) = -\log(x)$, $x > 0$
- $f(x) = -\sqrt{x}$, $x \geq 0$
- $f(x) = 1/x$, $x > 0$
- $f(x) = |x|$

Propriétés

Théorème (caractérisation fondamentale)

f est convexe si, et seulement si, $epi(f)$ est convexe.

f_1, f_2 convexes et $\mathrm{Dom}(f_1) \cup \mathrm{Dom}(f_2) \neq \emptyset$

- $f_1 + f_2$ convexe
- af_1 convexe, $\forall a \geq 0$
- $\sup\{f_1, f_2\}$ convexe
- $f_1 \Box f_2(x) := \inf_z f_1(z) + f_2(x - z)$ (inf-convolution) est convexe
- f convexe et A affine, $(f \circ A)(x) = f(Ax)$ convexe
- f_i $(i \in I)$ affines, $f(x) = \sup_{i \in I} f_i(x)$ convexe.

Continuité

Définition (Continuité)

C convexe ouvert de $\mathbb{R}^n$, $f : C \to \mathbb{R}$. f semi-continue inférieurement (sci) en $x \in C$ si $\forall \{x_k\} \subset C$, $x_k \to x$

$$f(x) \leq \lim_{x_k \to x} \inf f(x_k).$$

Définition (Dérivée directionnelle)

$f : C \to \mathbb{R}$, La dérivée directionnell de f en $x \in C$ dans la direction $d \in \mathbb{R}^n$ (si elle existe) est

$$f'(x; d) = \lim_{t \to 0} \frac{1}{t} (f(x + td) - f(x)))$$

f différentiable en x si $f'(x; d)$ existe dans toutes les directions.

Propriétés

Théorème

C ouvert convexe et f diférentiable sur C, alors les affirmations suivantes sont équivalentes

$$f \quad \text{convexe sur} \quad C$$

$$\forall x, y \in C, \langle \nabla f(x), y - x \rangle \leq f(y) - f(x)$$

$$\forall x, y \in C, \langle \nabla f(x) - \nabla f(y), y - x \rangle \geq 0.$$

Théorème

f fortement convexe sur C convexe ouvert de $\mathbb{R}^n$, alors

$$\forall x, y \in C, \langle \nabla f(x) - \nabla f(y), y - x \rangle \geq \alpha \parallel x - y \parallel_2^2 .$$

$\nabla^2 f(x)$ semi définie positive

Fonction conjuguée

Définition (Fonction support)

C sous-ensemble de $\mathbb{R}^n$,

$$\sigma_C(y) = \sup_{x \in C} \langle y, x \rangle$$

La fonction σ_C est convexe sci.

Définition (Fonction conjuguée)

La conjuguée d'une fonction f de $\mathbb{R}^n$ dans $\mathbb{R}$ est

$$f^*(y) = \sup_{x \in Dom(f)} \langle x, y \rangle - f(x)$$

Inégalité de Fenchel

$$f(x) \geq \langle x, y \rangle - f^*(y), \quad \forall y \in \mathrm{Dom}(f^*) \text{ et } \forall x \in \mathrm{Dom}(f)$$

Fonction conjuquée : exemple de calcul

- $f(x) = |x|$

$$xy \leq |x||y| \Rightarrow yx - |x| \leq (|y| - 1)|| \Rightarrow f^*(y) = \begin{cases} 0 & \text{si } |y| \leq 1 \\ +\infty & \text{sinon} \end{cases}$$

- $f(x) = \chi_C$, C convexe

$$f^*(y) = \sup \langle x, y \rangle - \chi_C(x) = \sup_{x \in C} \langle x, y \rangle = \sigma_C(y)$$

Sous-différentiabilité

Définition (Sous-gradient)

$\gamma \in \mathbb{R}^n$ *sous-gradient de* f *en* x_0 *si*

$$f(x) \geq f(x_0) + \langle \gamma, x - x_0 \rangle, \quad \forall x \in Dom(f).$$

$\partial f(x_0)$ *ensemble des sous-gradient en* x_0 *(sous-différentiel).*

Conditions équivalentes

$$y_0 \in \partial f(x_0)$$
$$x_0 \in \partial f^*(y_0)$$
$$f(x_0) + f^*(y_0) = \langle x_0, y_0 \rangle$$

f différentiable en $x_0 \Leftrightarrow \partial f(x_0) = \{\nabla f(x_0)\}$

Dualité de Fenchel

X, Y $\mathbb{R}$-espaces vectoriels normés complets de dual X^*, Y^*

$\Lambda \in \mathscr{L}(X, Y)$, $\Lambda^* \in \mathscr{L}(Y^*, X^*)$ (adjoint)

$F \ : \ X \ \to \bar{\mathbb{R}}$ convexe, $G \ : \ Y \ \to \bar{\mathbb{R}}$ convexe

F^*, G^* conjuguées

Définition (Problème primal)

$$(P) \qquad \inf_{x \in X} F(x) + G(\Lambda x)$$

Définition (Problème dual)

$$(P^*) \qquad \sup_{y^* \in Y*} -F^*(-\Lambda^* y^*) - G^*(y^*)$$

Conditions d'optimalités

Theorem (Fenchel)

F, G convexes sci propres

$\exists x_0 \in X$, $F(x_0 < +\infty$, $G(\lambda x_0) < +\infty$, G continue en Λx_0.

Alors (P) et (P^*) admettent au moins une solution $\bar{x}$ et $\bar{y}^*$

$$\alpha = \inf_{x \in X} F(x) + G(\Lambda x) = \sup_{y^* \in Y*} -F^*(-\Lambda^* y^*) - G^*(y^*)$$

Theorem

Si $|\alpha| < +\infty$ alors

$$\begin{aligned} -\Lambda^* \bar{y}^* &\in \partial F(\bar{x}) \\ \bar{y}^* &\in \partial G(\Lambda \bar{x}) \end{aligned}$$

$\bar{x}$ solution de (P), $\bar{y}^*$ solution de (P^*)

Descente du gradient

Gradient Descent (GD)

1. For $k = 0, \ldots, K-1$

$$g_k \in \partial f(x_k)$$

$$x_{k+1} = x_k - t_k g_k$$

2. Return x_K or $x_m = (x_0 + \cdots x_{K-1})/K$

Marche aussi pour le cas avec contrainte, $x \in C$:

$$x_{k+1} = \mathrm{Proj}_C(x_k - t_k g_k)$$

Descente du Gradient 2

Hypothèse (HL) (Lipschitz)

$$\exists L > 0, \quad \forall x, \quad g \in \partial f(x) \text{ tel que } \| g \| \leq L$$

Théorème

Si (HL) et $t_k = R/(L\sqrt{K})$ *alors*

$$f\left(\frac{1}{K}\sum_{k=0}^{K-1} x_k\right) - f(x^*) \leq \frac{RL}{\sqrt{K}}$$

Rem : Pour une précision ε, $K = R^2 L^2/\varepsilon^2$

Descente du Gradient 3

Hypothèse (βL) (β-Lipschitz)

$$\forall x, y, \quad \| \nabla f(x) - \nabla f(y) \| \leq \beta \| x - y \|$$

Théorème

Si (βL) et $t_k = 1/\beta$ alors

$$f(x_K) - f(x^*) \leq \frac{2\beta \| x_0 - x^* \|^2}{K + 4}$$

Rem : Pour une précision ε, $K = 2\beta/\varepsilon$

Descente du Gradient 4

Hypothèse (αC) f fortement convexe

$$\forall x, y, \quad g_x \in \partial f(x), \qquad f(y) \geq f(x) + \langle g, y - x \rangle + \frac{\alpha}{2} \parallel y - x \parallel^2$$

Théorème

Si (βL), (αC) et $t_k = 1/(\alpha(k+1))$ alors

$$f\left(\frac{1}{K}\sum_{k=0}^{K-1} x_k\right) - f(x^*) \leq \frac{2L^2\log(K)}{\alpha K}$$

Rem : Pour une précision ε, $K = 2L^2/(\alpha\varepsilon)$

Descente du Gradient accélérée (Nesterov)

Nesterov

1. $k = 0,\ y_1 = x_0,\ t_0 = 1$
2. For $k \geq 0$

$$
\begin{aligned}
x_k &= y_k - (1/L)\nabla f(y_k) \\
t_{k+1} &= (1 + \sqrt{1 + 4t_k^2})/2 \\
y_{k+1} &= x_k + (t_k - 1)(x_k - x_{k-1})/t_{k+1}
\end{aligned}
$$

Théorème

$$
f(x_k) - f(x^*) \leq \frac{2L \parallel x_0 - x^* \parallel^2}{k^2}
$$

Rem : Pour une précision ε, $K = 2L/\sqrt{\varepsilon}$

ADMM

ADMM = Alternating Direction Method of Multiplier.
Problem modèle

$$\min_{x,z} f(x) + g(z)$$
$$Ax + Bz = c$$

$x \in \mathbb{R}^n$, $z \in \mathbb{R}^m$, A $p \times n$, B $p \times m$, $c \in \mathbb{R}^p$.

$$p^* = \inf_{(x,z)} \{f(x) + g(z) \mid Ax + Bz = c\}$$

Lagrangien augmenté

$$\mathscr{L}_r(x, z, \lambda) = f(x) + g(z) + \lambda^\top (Ax + Bz - c) + \frac{r}{2} \| Ax + Bz - c \|_2^2 .$$

ADMM Standard

$r > 0$ fixé ; λ^0, z^0 données

$$
\begin{aligned}
x^{k+1} &= \arg\min_x \mathscr{L}_r(x, z^k, \lambda^k) \\
z^{k+1} &= \arg\min_x \mathscr{L}_r(x^{k+1}, z, \lambda^k) \\
\lambda^{k+1} &= \lambda^k + r(Ax^{k+1} + Bz^{k+1} - c)
\end{aligned}
$$

ADMM Symétrique

$r > 0$ fixé ; λ^0, z^0 données

$$
\begin{aligned}
x^{k+1} &= \arg\min_x \mathscr{L}_r(x, z^k, \lambda^k) \\
\lambda^{k+1/2} &= \lambda^k + \frac{r}{2}(Ax^{k+1} + Bz^k - c) \\
z^{k+1} &= \arg\min_x \mathscr{L}_r(x^{k+1}, z, \lambda^{k+1/2}) \\
\lambda^{k+1} &= \lambda^{k+1/2} + \frac{r}{2}(Ax^{k+1} + Bz^{k+1} - c)
\end{aligned}
$$

ADMM Convergence

(H1) f, g sci convexes propres fermées (épigraphes de f et g convexes fermées non vides)

(H2) $\mathscr{L}_r$ admet un selle

$$\mathscr{L}_r(x^*, z^*, \lambda) \leq \mathscr{L}_r(x^*, z^*, \lambda^*) \leq \mathscr{L}_r(x, z, \lambda^*)$$

Théorème

Si (H1)+(H2), alors quand $k \to +\infty$

$$\begin{aligned} r^k = Ax^k + Bz^k - c &\to 0 && (1)\\ f(x^k) + g(z^k) &\to p^* && (2)\\ \lambda^k &\to \lambda^* && (3) \end{aligned}$$