Gradient Projeté

Alexandre Gramfort

alexandre.gramfort@telecom-paristech.fr

Telecom ParisTech

SIGMA 201b

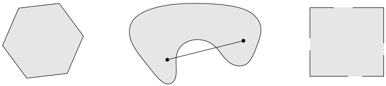
Plan

Ensemble convexe

Définition (ensemble convexe)

- Un ensemble $\mathcal{C} \subset \mathbb{H}$ est convexe ssi
  $\forall x, y \in \mathcal{C}$, $\forall t \in [0\,1], (1-t)x + ty \in \mathcal{C}$

1 ensemble convexe et 2 non-convexes :



Exemple convexe

Exemples

- Intervalle [a, b] sur $\mathbb{R}$.
- hyperplan $\{x | a^T x = b\}$ avec $(a \neq 0)$
- hemiplan $\{x | a^T x \leq b\}$ avec $(a \neq 0)$
- boule euclidienne $\mathcal{B}(x_c, b) = \{x | \|x - x_c\|_2 \leq b\}$
- ellipsoide $\{x | (x - x_c)^T P^{-1} (x - x_c) \leq b\}$ avec $P \in \mathcal{S}_n^{++}$

Comment établir la convexité d'un espace ?

- Appliquer la définition :
  $x, y \in \mathcal{C}, t \in [0, 1] \Rightarrow tx + (1-t)y \in \mathcal{C}$
- Montrer que $\mathcal{C}$ est obtenu à partir d'ensembles convexes simples par opérations qui préservent la convexité :
  - intersection
  - application d'une fonction affine $f : x \rightarrow Ax + b$

Fonction convexe

Définition (Fonction convexe)

- Une fonction $J : \mathcal{C} \to \mathbb{R} \cup \{+\infty\}$ est convexe si $\mathcal{C}$ est convexe et si $\forall x, y \in \mathcal{C}$, $\forall t \in [0\,1]$,
  $J((1-t)x + ty) \leq (1-t)J(x) + tJ(y)$

Remarque: Si J est convexe on dit de $-J$ est **concave**.

Remarque: J est convexe ssi toutes les fonctions $g(t) = J(x + th)$ sont convexes.

Définition (stricte convexité)

- Une fonction $J : \mathcal{C} \to \mathbb{R} \cup \{+\infty\}$ est **strictement** convexe si $\mathcal{C}$ est convexe et si $\forall x, y \in \mathcal{C}$ tels que $x \neq y$, $\forall t \in]0\,1[$,

$$J((1-t)x + ty) < (1-t)J(x) + tJ(y)$$

- Exemples : norme, fonctions affines

Exemples de fonctions convexes

Exemples

- $J(x) = \|x\|$, $x \in \mathbb{H}$ est strictement convexe
- $J(x) = a^T x + b$, $a \in \mathbb{H}$ et $b \in \mathbb{R}$ est convexe (et aussi concave).
- $J(x) = e^{ax}$, $a \in \mathbb{R}$ et $x \in \mathbb{R}$ est convexe.
- $J(x) = x^p, x \in \mathbb{R}_{++}$ avec $p \geq 1$ ou $p \leq 0$ est convexe.
- $J(x) = |x|^p, x \in \mathbb{R}$ avec $p \geq 1$ est convexe.
- $J(x) = x \ln(x), x \in \mathbb{R}_{++}$ est convexe.
- $J(x) = x^T A x - b^T x$, $A \in \mathcal{S}_n^+$ et $b \in \mathbb{H}$ est convexe.

Convexité

Définition (Domaine d'une fonction convexe)

Le domaine de J $\mathrm{dom}(J) = \{x / J(x) < +\infty\}$.

Remarque: Si J est convexe, $\mathrm{dom}(J)$ est convexe

Définition (Fonction propre)

Si $\mathrm{dom}(J)$ est non vide, on dit que J est **propre**.

Théorème (continuité)

Toute fonction J convexe propre sur un espace de dimension finie est continue sur l'intérieur de $\mathrm{dom}(J)$

Fonctions elliptiques

Définition (fonction elliptique)

Soit $J : \mathcal{C} \subset \mathbb{H} \to \mathbb{R}$; J est *elliptique*, ou α-convexe, ssi il existe une constante d'ellipticité $\alpha > 0$ telle que $\forall x, y$

a) $J((1-t)x + ty) \leq (1-t)J(x) + tJ(y) - \frac{\alpha}{2} t(1-t)\|x-y\|^2$

Proposition (critères d'ellipticité à l'ordre 1)

Si J est Gâteaux-différentiable, J est α-convexe ssi

b) $\forall x, y \in \mathcal{C}, J(y) \geq J(x) + (\nabla J(x), y - x) + \frac{\alpha}{2}\|x-y\|^2$

ce qui équivalent à

c) $\forall x, y \in \mathcal{C}, (\nabla J(y) - \nabla J(x), y - x) \geq \alpha\|x-y\|^2$

Fonctions elliptiques

Proposition (critère d'ellipticité à l'ordre 2)

Si J est différentiable à l'ordre 2, J est elliptique ssi $\exists\alpha > 0$ telle que

d) $\forall x, h \in \mathcal{C}, (\nabla^2 J(x)h,\ h) \geq \alpha \|h\|^2$

Exemple

- formes quadratiques $J(x) = x^T A x - b^T x$. A doit être définie positive. α est la plus petite valeure propre.

Plan

Résultats d'existence et d'unicité

Définition (Problème d'optimisation)

$$(\mathcal{P}) : \min J(x), x \in \mathcal{C} \subset \mathbb{R}^n$$

Théorème (Existence)

- *On suppose que*
  - *J est continue et $\mathcal{C}$ est un fermé non vide*
  - *Soit $\mathcal{C}$ est borné, soit J est coercitive*
- *Alors le problème $(\mathcal{P})$ admet au moins une solution*

Théorème (Existence et unicité)

- *Si de plus J est strictement convexe et $\mathcal{C}$ est convexe, alors $(\mathcal{P})$ admet une solution unique*

Résultats d'existence et d'unicité

Exemples

$\mathcal{C} = \{x / h(x) = 0, g(x) \leq 0\}$, où

1. $h = [h_1 \dots h_p]$, $g = [g_1 \dots g_q]$, avec h_i et g_j continues. On a alors $\mathcal{C}$ fermé.
2. h_i affines et g_j convexes (Définition d'un problème convexe).

Plan

Condition générale d'optimalité

Théorème (condition nécessaire)

Si J est Gâteaux-différentiable et si $\mathcal{C}$ est un convexe fermé, alors toute solution x^* de $(\mathcal{P})$ vérifie $\forall x \in \mathcal{C}, (\nabla J(x^*), x - x^*) \geq 0$

Théorème (condition nécessaire et suffisante)

Si de plus J est convexe, cette condition est nécessaire et suffisante

Remarque

- Si $\mathcal{C} = \mathbb{R}^n$, ou si $x^* \in$ intérieur$(\mathcal{C})$, alors $\nabla J(x^*) = 0$
- Exemples : contour d'un disque dans $\mathbb{R}^2$, intérieur d'un disque dans $\mathbb{R}^3$

Conditions d'optimalité qualifiées

Théorème (Conditions nécessaires qualifiées du 1er ordre (Karush, Kuhn et Tucker))

- *On suppose que*
  - J, h et g sont $\mathcal{C}^1$,
  - x^* est solution de $(\mathcal{P})$
  - x^* est *régulier* pour les contraintes h et g.
- *Alors les conditions de KKT sont vérifiées :* $\exists \lambda^* = (\lambda_1^* \dots \lambda_p^*)$ et $\mu^* = (\mu_1^* \dots \mu_q^*)$ *tels que*
  - $h(x^*) = 0$ et $g(x^*) \leq 0$
  - $\forall j \in \{1 \dots q\}$, $\mu_j^* \geq 0$ et $\mu_j^* g_j(x^*) = 0$
  - $\nabla J(x^*) + \sum_{i=1}^{p} \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^{q} \mu_j^* \nabla g_j(x^*) = 0$

Lagrangien

Définition (Lagrangien)

- On appelle Lagrangien du problème $(\mathcal{P})$ la fonction $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^p \times (\mathbb{R}^+)^q \to \mathbb{R}$ définie par

$$\mathcal{L}(x, \lambda, \mu) = J(x) + \sum_{i=1}^{p} \lambda_i h_i(x) + \sum_{j=1}^{q} \mu_j g_j(x)$$

Remarques

- Conditions KKT $\Rightarrow \nabla_x \mathcal{L}(x^*, \lambda^*, \mu^*) = 0$
- Si J et g sont convexes et h est affine, alors $\forall \lambda \in \mathbb{R}^p$, $\mu \in (\mathbb{R}^+)^q$, $\mathcal{L}(x, \lambda, \mu)$ est convexe par rapport à x

Conditions d'optimalité qualifiées

Dans le cas convexe :

Théorème (Conditions nécessaires et suffisantes du 1er ordre (Karush, Kuhn et Tucker))

- *On suppose que*
  - J, h *et* g *sont* $\mathcal{C}^1$,
  - J *et* g *sont convexes et* h *est affine*,
  - x^* *est régulier pour les contraintes* h *et* g.
- *Alors* x^* *est solution de* $(\mathcal{P})$ *ssi les conditions de KKT sont vérifiées :* $\exists \lambda^*$ *et* μ^* *tels que*
  - $h(x^*) = 0$ *et* $g(x^*) \leq 0$
  - $\forall j \in \{1 \dots q\}$, $\mu_j^* \geq 0$ *et* $\mu_j^* g_j(x^*) = 0$
  - $\nabla J(x^*) + \sum_{i=1}^{p} \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^{q} \mu_j^* \nabla g_j(x^*) = 0$

Plan

Rappel : projection sur un convexe fermé

Théorème (projection sur un convexe fermé)

Soit $\mathcal{C}$ *convexe, fermé, non vide, et* $x \in \mathbb{R}^n$*. Alors il existe un unique* $x^* \in \mathcal{C}$ *qui minimise* $\|x - x^*\|^2$*, caractérisé par*
$\forall y \in \mathcal{C}, (x - x^*, y - x^*) \leq 0$

Proposition (continuité)

On pose $\pi_{\mathcal{C}}(x) = x^*$*. Alors la projection* $\pi_{\mathcal{C}}$ *est continue. De plus,* $\pi_{\mathcal{C}}$ *est une contraction :* $\forall x, y \in \mathbb{R}^n, \|\pi_{\mathcal{C}}(x) - \pi_{\mathcal{C}}(y)\| \leq \|x - y\|$

Méthode de gradient avec projection

- Hypothèse : $\mathcal{C}$ convexe, fermé, non vide
- Algorithme
  1. Initialisation
     - $k = 0$: choix de x_0 et de $\beta_0 > 0$
  2. Itération k
     - $x^{k+1} = \pi_{\mathcal{C}}(x^k - \beta_k \nabla J(x^k))$
  3. Critère d'arrêt
     - Si $\|x^{k+1} - x^k\| < \varepsilon$, stop
     - Sinon, on pose $k = k + 1$, et on retourne à 2

Méthode de gradient avec projection

Théorème (Convergence)

- *Soit J $\mathcal{C}^1$, α-convexe et de dérivée M-lipschitzienne*
- *Si* $\beta_k \in [\beta_{\min}, \beta_{\max}]$ *tel que* $0 < \beta_{\min} < \beta_{\max} < 2\alpha/M^2$,
- *Alors l'algorithme converge vers la solution de* $(\mathcal{P})$

La dérivée de J est M-lipschitzienne si on a

$$\forall x, y \in \mathcal{C}, \|\nabla J(x) - \nabla J(y)\| \leq M\|x - y\|$$

Éléments de preuve

- On remarque que $x^* = \pi_{\mathcal{C}}(x^* - \beta_k \nabla J(x^*))$
- On majore $\|x^{k+1} - x^*\|^2$ par une récurrence en utilsant l'optimalité de x^* et le fait que la projection est contractante.