

Optimisation non-linéaire avec contraintes

Alexandre Gramfort

alexandre.gramfort@telecom-paristech.fr

Telecom ParisTech



MDI 210

Plan

- 1 Théorie de l'optimisation sous contrainte
- 2 Algorithmes

Problèmes d'optimisation

Définition (Problème d'optimisation (\mathcal{P}))

- $\min f(x)$, $x \in X$, où $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ est appelée **fonction de coût, objectif, ou critère**
- $X = \{x \in \mathbb{R}^n / g(x) \leq 0 \text{ et } h(x) = 0\}$ est l'**ensemble des contraintes**
- $g(x) \leq 0$ représente les **contraintes en inégalité**.
 $g(x) = (g_1(x), \dots, g_m(x))$ soit m contraintes d'inégalité.
- $h(x) = 0$ représente les **contraintes en égalité**.
 $h(x) = (h_1(x), \dots, h_p(x))$ soit p contraintes d'égalité.
- un élément $x \in X$ est dit **réalisable**
- l'ensemble des points réalisables X est appelé **domaine réalisable**

Si $g_i(x) = 0$ on dit que la contrainte g_i est saturée.

Gradient et matrice Hessienne

On note $\nabla f(x)$ le gradient et $\nabla^2 f(x)$ la matrice hessienne. On a :

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix} \quad \nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

Et la formule de Taylor qui s'écrit :

$$\forall h \in \mathbb{R}^n, f(x+h) = f(x) + \nabla f(x)^T h + \frac{1}{2} h^T \nabla^2 f(x) h + o(\|h\|^2)$$

Exemples

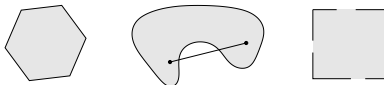
- Si $f(x) = x^T A x - b^T x$, $\nabla f(x) = (A + A^T)x - b$ et $\nabla^2 f(x) = A + A^T$

Ensemble convexe

Définition (ensemble convexe)

- Un ensemble $\mathcal{C} \subset \mathbb{R}^n$ 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 $(a \neq 0)$
- hémipplan $\{x | a^T x \leq b\}$ avec $(a \neq 0)$
- boule euclidienne $\mathcal{B}(x_c, b) = \{x | \|x - x_c\|_2 \leq b\}$
- ellipsoïde $\{x | (x - x_c)^T P^{-1} (x - x_c) \leq b\}$ avec $P \in \mathcal{S}_n^{++}$

Fonction convexe

Définition (Fonction convexe)

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

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

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

Définition (stricte convexité)

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

$$f((1-t)x + ty) < (1-t)f(x) + tf(y)$$

- Exemples : norme, fonctions affines

Programme convexe

Définition (Programme convexe)

- (\mathcal{P}) est un problème de programmation convexe (ou programme convexe) quand les fonctions f et $g_i, i = 1 \dots m$, sont convexes, et les fonctions $h_j, j = 1 \dots p$, sont affines.

Remarque

Montrer que sous ces hypothèses le domaine réalisable de (\mathcal{P}) est un ensemble convexe.

Résultats d'existence et d'unicité

Définition (Problème d'optimisation)

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

Théorème (Existence)

- Soit f est continue et X est un fermé non vide. Si
 - X est borné
 - ou bien f est coercitive
- Alors le problème (\mathcal{P}) admet au moins une solution

Théorème (Existence et unicité)

- Si de plus f est strictement convexe et X est convexe, alors (\mathcal{P}) admet une solution unique

Condition générale d'optimalité

Théorème (condition nécessaire)

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

Théorème (condition nécessaire et suffisante)

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

Remarque

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

Direction admissible

Définition (Direction admissible)

On dit qu'une direction d est admissible en $x_0 \in X$ s'il existe une fonction ϕ de \mathbb{R} dans \mathbb{R}^n telle que :

- $\phi(0) = x_0$
- pour tout $t > 0$ assez petit, $\phi(t) \in X$
- la dérivée à droite de ϕ en 0 est d .

En pratique ça veut dire :

$$f(\phi(t)) = f(x_0) + td^\top \nabla f(x_0) + t\epsilon(t)$$

On note $A(x_0)$ l'ensemble des directions admissibles et $I_0(x^0) = \{i/g_i(x_0) = 0\}$, i.e., l'ensemble des contraintes saturées.

Direction admissible

Proposition

Si d est une direction admissible en $x_0 \in X$ alors

- $\forall i \in I_0(x_0), d^\top \nabla g_i(x_0) \leq 0$ (i)
- $\forall j \in J, d^\top \nabla h_j(x_0) = 0$ (ii)

On note $B(x_0)$ l'ensemble des j qui vérifie (i) et (ii). On a donc $A(x_0) \subseteq B(x_0)$.

Proposition

On suppose que, pour $j \in J$, les fonctions h_j sont affines. Soient $x_0 \in X$ et d une direction vérifiant :

- $\forall i \in I_0(x_0), d^\top \nabla g_i(x_0) < 0$
- $\forall j \in J, d^\top \nabla h_j(x_0) = 0$

Alors d est une direction admissible en x_0 .

Point régulier / Contraintes qualifiées

Définition

On dit que les contraintes sont qualifiées en $x_0 \in X$, ou bien que x_0 est un point régulier, si toute direction dans $B(x_0)$ est limite d'une suite de directions de $A(x_0)$, i.e., $A(x_0)$ est dense dans $B(x_0)$.

Proposition (Conditions de Slater)

Si :

- les fonctions g_i sont convexes
- les fonctions h_j sont affines,
- et il existe $\tilde{x} \in X$ avec,
 - pour tout $i \in I$, $g_i(\tilde{x}) < 0$
 - pour tout $j \in J$, $h_j(\tilde{x}) = 0$

alors les contraintes sont qualifiées en tout point de X .

Point régulier / Contraintes qualifiées (suite)

Proposition (Indépendance des gradients)

On suppose que, pour $j \in J$, les fonctions h_j sont affines. Si, en $x_0 \in X$, l'ensemble des gradients :

- $\nabla g_i(x_0)$ pour $i \in I_0(x_0)$
- $\nabla h_j(x_0)$ pour $j \in J$

sont linéairement indépendants, alors les contraintes sont qualifiées en x_0 (x_0 est régulier).

Remarque

- En réalité, n'est pas nécessaire que les h_j soient affines.
- C'est la propriété la plus utile en pratique si (\mathcal{P}) n'est pas convexe.

Point régulier / Contraintes qualifiées (suite)

Proposition

On suppose que, pour $j \in J$, les fonctions h_j sont affines. Soit $x_0 \in X$. S'il existe une direction \tilde{d} telle que :

- $\forall i \in I_0(x_0), \tilde{d}^\top \nabla g_i(x_0) < 0$
- $\forall j \in J, \tilde{d}^\top \nabla h_j(x_0) = 0$

alors les contraintes sont qualifiées en x_0 .

Condition générale d'optimalité

Théorème (condition nécessaire)

Si f est différentiable, si X est non-vide, si x^ est un minimum local où les contraintes sont qualifiées. Alors, si $d \in B(x^*)$:*

$$d^\top \nabla f(x^*) \geq 0$$

En d'autres termes, aucune direction admissible en x^* n'est une direction de descente.

Conditions d'optimalité qualifiées

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

- On suppose que
 - f , 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 : $\exists \lambda^* = (\lambda_1^* \dots \lambda_m^*)$ et $\mu^* = (\mu_1^* \dots \mu_p^*)$ tels que
 - $h(x^*) = 0$ et $g(x^*) \leq 0$ (satisfiabilité)
 - $\forall i \in \{1 \dots m\}$, $\lambda_i^* \geq 0$ et $\lambda_i^* g_i(x^*) = 0$ (complémentarité)
 - $\nabla f(x^*) + \sum_{j=1}^p \mu_j^* \nabla h_j(x^*) + \sum_{i=1}^m \lambda_i^* \nabla g_i(x^*) = 0$ (stationarité)

Remark : On peut écrire KKT comme

$$\nabla f(x^*) = \sum_{j=1}^p \mu_j^* \nabla h_j(x^*) - \sum_{i \in I_0(x^*)} \lambda_i^* \nabla g_i(x^*)$$

Contraintes en égalité

Corollaire (Conditions nécessaires qualifiées du 1er ordre (Conditions de Lagrange))

- On suppose que
 - $X = x/h(x) = 0$ et f et h sont \mathcal{C}^1 ,
 - x^* est solution de (\mathcal{P})
 - les p vecteurs $\nabla h_i(x^*)$ sont indépendants (condition de *qualification* des contraintes, équivalent ici à supposer x^* régulier)
- Alors $\exists p$ *multiplicateurs de Lagrange* $\mu_1^* \dots \mu_p^*$ tels que
 - x^* est réalisable : $h(x^*) = 0$
 - $\nabla f(x^*) + \sum_{j=1}^p \mu_j^* \nabla h_j(x^*) = 0$

Conditions d'optimalité qualifiées

Dans le cas convexe :

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

- On suppose que
 - f, h et g sont \mathcal{C}^1 ,
 - f 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 $g(x^*) \leq 0$ (satisfiabilité)
 - $\forall i \in \{1 \dots m\}$, $\lambda_i^* \geq 0$ et $\lambda_i^* g_i(x^*) = 0$ (complémentarité)
 - $\nabla f(x^*) + \sum_{i=1}^p \mu_i^* \nabla h_i(x^*) + \sum_{j=1}^m \lambda_j^* \nabla g_j(x^*) = 0$ (stationarité)

Contre exemple de qualification

Soit le domaine X défini par les contraintes

$$\begin{cases} x^2 + (y - 1)^2 \geq 1 \\ x \leq 1 \\ y \geq 0 \end{cases} \quad (1)$$

Montrer que le point $(0, 0)$ n'est pas régulier.

Plan

- 1 Théorie de l'optimisation sous contrainte
- 2 Algorithmes

Méthode de descente

$x^{k+1} = x^k + sd, s > 0$ avec d direction de descente.

Idée :

$$\min_{d \in \mathbb{R}^n} d^\top \nabla f(x^k)$$

avec $d^\top \nabla g_i(x^k) \leq 0$ pour tout $i \in I_0(x^k)$ et $d^\top d = 1$.

Si l'on remplace $d^\top d = 1$ par $\|d\|_\infty \leq 1$ on a programme linéaire.

Linéarisation

Idee : Construire une suite minimisante (x^k) en approchant $f(x^k)$ par Taylor à l'ordre 1.

Soit $x^0 \in X$ un point initial quelconque.

On itère :

$$\begin{aligned}x^{k+1} &= \arg \min_{x \in X} f(x^k) + (x - x^k)^\top \nabla f(x^k) \\ &= \arg \min_{x \in X} x^\top \nabla f(x^k)\end{aligned}\tag{2}$$

Remarque

Si X est un polyèdre (contraintes affines), cela revient à résoudre un programme linéaire à chaque itération.

Frank-Wolfe

Soit $x^0 \in X$ un point initial quelconque.

On itère :

$$\begin{aligned}\tilde{x}^{k+1} &= \arg \min_{x \in X} x^\top \nabla f(x^k) \\ x^{k+1} &= \arg \min_{s \in [0,1]} f(sx^k + (1-s)\tilde{x}^{k+1})\end{aligned}\tag{3}$$

Théorème (Convergence Frank-Wolfe)

Soit f une fonction de \mathbb{R}^n dans \mathbb{R} , de classe \mathcal{C}^1 , strictement convexe. Soit X un polyèdre convexe compact de \mathbb{R}^n . La méthode de Frank et Wolfe appliquée au problème (\mathcal{P}) de la minimisation de f sur X converge vers le minimum global.