Optimisation Non linèaire

Par

Professeur Abdellatif El Afia

Chapitre 8

Méthodes duals

Professor Abdellatif El Afia

Méthodes duals

Soit le problème primal

$$(P)\begin{cases} Min & f(x) \\ sujet\ à & f_i(x) \leq 0 \quad\quad i = 1, \dots, m \\ & x \in \mathbb{R}^n \end{cases}$$

- $L(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i f_i(x)$
- $X = \{x \in \mathbb{R}^n, f_i(x) \leq 0 \ \ \forall i = 1, \dots, m\}$ avec $f_1(x), \dots, f_m(x): \mathbb{R}^n \to \mathbb{R}$.
- $O = \{x \in \mathbb{R}^n, F(x) = (f_1(x), \dots, f_m(x)): \leq 0\}$
- $\Lambda = \{\lambda \in \mathbb{R}^m, \lambda = (\lambda_1, \dots, \lambda_m)^T, \lambda_i \geq 0, i = 1, \dots, m\}$

Approche de résolution: méthode itérative

- **Itération initiale** : Déterminer une solution initiale $\lambda^0 \in \Lambda$.
- **Itération générale $(\boldsymbol{k+1})$:** Etant donné λ^k
  - Déterminer x_k qui minimise $L(x, \lambda_k)$ (en utilisant l'un des algorithmes sans contraintes)
  - Déterminer λ_{k+1}
- Répéter l'itération générale jusqu'à ce qu'un **critère d'arrêt** soit satisfait.

Méthode de projection

Théorème : Projection sur un convexe fermé

Soit $X \subset \mathbb{R}^n$ un ensemble convexe, fermé, non vide et soit $v \in \mathbb{R}^n$. Alors $\exists!\, P_X(v) \in X$ tel que

$$\|v - P_X(v)\| = \inf_{x \in X} \|v - x\| = \min_{x \in X} \|v - x\|.$$

L'élément $P_X(v) \in X$ s'appelle **la projection de $\boldsymbol{v}$ sur $\boldsymbol{X}$ en $\mathbb{R}^{\boldsymbol{n}}$.**

En plus on a :

- $< v - P_X(v), x - P_X(v) > \leq 0, \qquad \forall u \in X$
- Si $w \in X$ est tel que $< v - w, x - w > \leq 0, \quad \forall x \in X$ alors $w = P_X(v)$
- On a $\|P_X(v_1) - P_X(v_2)\| \leq \|v_1 - v_2\| \qquad \forall v_1, v_2 \in \mathbb{R}^n$
  P_X est une fonction lipschitzienne, donc continue.
- $v = P_X(v) \Leftrightarrow v \in X$
- Si U est le sous-espace affine de $\mathbb{R}^n$ donné par $X = a + X_0$ avec $a \in \mathbb{R}^n$ et X_0 un sous espace vectoriel de $\mathbb{R}^n$ alors $v - P_X(v) \perp X_0$, c'est-à-dire

$$< v - P_X(v), a + x - P_X(v) > = < v - P_X(v), x > = 0 \qquad \forall u \in X_0$$

Méthode de projection

Cas particulier : Si $X = \mathbb{R}_+^n$ alors $P_X(v) = (v_1^+, v_2^+, \dots, v_n^+)^T$, avec $v_i^+ = \max(0, v_i)$.

Supposons que $x^* \in X$ est tel que $f(x^*) = \min_{x \in X} f(x)$ Alors $\forall x \in X$

- $< \nabla f(x^*), x - x^* > \geq 0 \quad \Leftrightarrow \forall \rho > 0, \rho < -\nabla f(x^*), x - x^* > \leq 0$
- $\Rightarrow < [x^* - \rho \nabla f(x^*)] - x^*, x - x^* > \leq 0$

En utilisant le Théorème de projection cette dernière inégalité est équivalente à l'égalité

$$x^* = P_X\big(x^* - \rho \nabla f(x^*)\big)$$

Une idée pour approcher numériquement ce point fixe est d'utiliser une méthode itérative, qui consiste à construire une suite : $x^{(k+1)} = P_X(x^{(k)} - \rho \nabla f\big(x^{(k)}\big) \in X, \rho > 0, k \in \mathbb{N}$

Il est aussi possible de faire varier ρ à chaque pas. Donc l'algorithme général sera dans ce cas :

$$\begin{cases} x^{(k+1)} = P_X(x^{(k)} - \rho_k \nabla f\big(x^{(k)}\big), k \in \mathbb{N} \\ x^{(0)} \in X \end{cases}$$

Avec $\rho_k > 0$ données. C'est la **méthode de gradient avec projection à pas variable**. On l'appelle **méthode de gradient avec projection à pas fixe** si ρ_k est indépendant de k.

Méthode d'Uzawa

Le problème primal du début peut donc s'écrire de la façon suivante $\min_{x \in X} f(x)$

On définit l'application $g(\lambda) = \min_{x \in \mathbb{R}^n} L(x, \lambda)$, est concave ou encore $-g$ est convexe

Maintenant on considère le problème d'optimisation dit « dual » suivant :

$$\begin{cases} \lambda^* \in \Lambda \\ g(\lambda^*) \geq g(\lambda) \quad \forall \lambda \in \Lambda \end{cases}$$

La méthode d'Uzawa est basée sur la formulation duale et la recherche des points selle.

Définition :

On dit que les contraintes de O sont **qualifiées** en $v \in O$ si

1) Soit $I(v) = \emptyset$
2) Soit $I(v) \neq \emptyset$ et il existe $w \in \mathbb{R}^n$ tel qu'on a $\forall i \in I(v) < \nabla f_i(v), w > \leq 0$ Avec en plus : $< \nabla f_i(v), w > < 0$ si f_i est non-affine

- $I(v) = \{i = 1, \dots m, f_i(v) = 0\}$ est un ensemble des contraintes actives

Méthode d'Uzawa

Nous faisons les hypothèses suivantes :

1) f est elliptique, c'est-à-dire qu'elle est de classe C^1 et il existe $\alpha > 0$ tel que :
   $< f(x) - f(y), x - y > \geq \alpha \|x - y\|^2 \quad \forall x, y \in \mathbb{R}^n$
2) Pour tout $i = 1, \dots, m$ on a que f_i est une fonction convexe et de classe C^1. En plus F est lipschitzienne, donc il existe $M > 0$ tel que :
   $$\|F(x) - F(y)\| \leq M\|x - y\| \;\; \forall x, y, \in \mathbb{R}^n$$
3) Les contraintes de O sont qualifiés en tout point $x \in O$.

Corollaire 1 :

Supposons que les fonctions $f, f_1, \dots, f_m$ sont convexes. Soit x^* un point de minimum de f sur O tel que les contraintes de O sont qualifiés en x^*. Alors il existe $\lambda^* \in \mathbb{R}_+^m$ tel que (x^*, λ^*) est un point selle du Lagrangien L

Méthode d'Uzawa

Principe:

Le prince de la méthode d'Uzawa est qu'à chaque itération pour λ^k connu, on cherche à minimiser sur $\mathbb{R}^n$ la fonctionnelle $x \to L(x, \lambda^k)$

Ceci permet de construire x^k, avec x^k connu on cherche à maximiser sur $\mathbb{R}_+^p$ la fonctionnelle $\lambda \to L(x^k, \lambda)$

Pour calculer l'itération λ_{k+1}, on peut utiliser un algorithme de gradient avec projection à pas variable

$$\lambda_{k+1} = P_{\mathbb{R}_+^p}(\lambda i_k + \rho_k \nabla g(\lambda_k))$$

Ce qui a pour expression sur $\mathbb{R}_+^p$

$$\lambda_{k+1} = \max\left\{\lambda i_k + \rho_k \frac{\partial g}{\partial \lambda_i}(\lambda_k), 0\right\} \qquad i = 1, \dots, p$$

$$\rho_k > 0$$

$$\frac{\partial g}{\partial \lambda_i}(\lambda_k) = f_i(x_k)$$

Méthode d'Uzawa

Algorithme d'Uzawa :

- **Etape1 :** on choisit $\lambda_0 > 0$, et ϵ $assez\ petit$
- **Etape 2 :** On calcule x_k qui minimise $L(x, \lambda_k)$ (en utilisant l'un des algorithmes sans contraintes)
  - Si $(k \geq 1)$ alors
  - Si $\|x_k - x_{k-1}\| < \epsilon$ On arrête, sinon en va à l'**étape3**
- **Etape3**
  - $\lambda_{k+1} = \max\{\lambda_{i_k} + \rho_k f_i(x_k), 0\}$
  - On pose ; $k = k + 1$ $et\ en\ retourne$ à l'é$\boldsymbol{tape2}$

Théorème :

Soient $a_1, a_2 \in \mathbb{R}$ tels que $0 < a_1 \leq a_2 < \frac{2\alpha}{M^2}$

Supposons que les facteurs ρ_k sont tels que :$a_1 \leq \rho_k \leq a_2 \quad \forall k \in \mathbb{N}$

Alors la suite x_k générée par l'algorithme d'Uzawa converge vers x^* solution du problème

Méthode d'Arrow-Hurwicz

Comme l'algorithme d'Uzawa le but de cet algorithme est de trouver le point selle, comme la condition de KKT nous garantit que la solution du problème d'optimisation (x^*, λ^*) est un point selle

En plus de l'expression de $\boldsymbol{\lambda_{k+1}}$ on cherche un $\boldsymbol{x_k}$ de la forme :

$$x_k = P_O\left(x_{k-1} - \epsilon_k\left(\nabla_x f(x_k) + \lambda_k \nabla_x \theta(x_k)\right)\right)$$

P_o est la projection de l'itération sur l'ensemble O des contraintes comme les contraintes sont convexes cet ensemble l'est aussi, on définit dont ce qu'est une projection sur un ensemble convexe fermé :

Définition :

Soit E un espace de Hilbert, muni d'une norme $\|.\|$ induite par un produit scalaire $<.,.>$ et soit O un convexe fermé non vide de E. Alors, $\forall x \in E$, il existe un unique $x_0 \in O$ tel que

$$\|x - x_0\| \leq \|x - y\| \;\; \forall y \in O$$

On note $x_0 = P_O(x)$ la projection orthogonale de x sur O.

Méthode d'Arrow-Hurwicz: Algorithme

- **Etape1 :**
  - Poser $k = 0$,
  - on choisit $\lambda^{(0)} \geq 0$ ($\lambda^{(0)} = 0$) et on choisit $k_{max} \in \mathbb{N}$, ($k_{\max} = 1000$) et $\epsilon > 0$ ($\epsilon = 10^{-6}$).
- **Etape2 :**
  - Calculer $x^k = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} L\left(x, \lambda^{(k)}\right) \rightarrow L\left(x, \lambda^{(k)}\right) = f(x) + \sum_{j=1}^{m} \lambda_i^{(k)} f_j(x)$
  - Si ($k \geq 1$) alors :
    - Si $\left\|x^k - x^{k-1}\right\| \leq \epsilon$ alors x^k est le point de minimum recherché (l'algorithme a convergé)
    - Sinon on va à l'Etape 3
- **Etape 3 :**
  - Si ($k = k_{\max}$) alors l'algorithme diverge
  - Sinon, faire :
    - $\lambda_i^{(k+1)} = \max\left\{\lambda_i^{(k)} + \rho_k f_i\left(x^k\right), 0\right\} \quad i = 1, \dots, m$
    - $x_{k+1} = P_O\left(x_k - \epsilon_k\left(\nabla_x f(x_k) + \lambda_k \nabla_x \theta(x_k)\right)\right)$
  - Ensuite $k = k + 1$, retour à l'Etape 2

Méthode d'Arrow-Hurwicz

Théorème :

Sous les conditions de KKT, le point x^* est la solution du problème (1) si et seulement s'il existe un réel λ^* tel que le point (x^*, λ^*) est un point selle du lagrangien $L(x, \lambda)$. De plus, si les suites (ϵ_k) et (ρ_k) respectent les propriétés suivantes :

- $\sum_{k=0}^{+\infty} \rho_k = +\infty;$
- $\sum_{k=0}^{+\infty} \epsilon_k = +\infty;$
- $\sum_{k=0}^{+\infty} \rho_k^2 < +\infty;$
- $\sum_{k=0}^{+\infty} \epsilon_k^2 < +\infty,$

Alors la suite x_k de l'algorithme d'Arrow-Hurwicz converge vers le pont x^*.

Remarque :

Ici nous avons considéré que les contraintes d'inégalité puisque d'un ensemble de contraintes d'égalité $f_i = 0$ on peut se remmener à un ensemble d'inégalité en considérant : $-f_i \leq 0$ et $f_i \geq 0$