APPROXIMATION NUMERIQUE ET OPTIMISATION

G. ALLAIRE

13 novembre 2018

CHAPITRE IV (fin)

☞ Algorithmes d'optimisation avec contraintes.

☞ Méthodes d'approximations successives.

☞ Notions de programmation linéaire.

Algorithmes d'optimisation avec contraintes.

- ☞ Gradient projeté.
- ☞ Algorithme d'Uzawa.
- ☞ Pénalisation des contraintes.
- ☞ Méthodes d'approximations successives.

Gradient projeté

Soit K convexe fermé non vide de V. On veut résoudre

$$\inf_{v \in K} J(v) \ .$$

Initialisation: choisir u^0 dans K.

Itérations: pour $n \geq 0$

$$u^{n+1} = P_K\big(u^n - \mu\, J'(u^n)\big) \ ,$$

avec P_K projection orthogonale sur K.

Théorème. On suppose que J est α-convexe différentiable pour $\alpha > 0$,

$$J(v) \geq J(u) + \langle J'(u), v - u \rangle + \frac{\alpha}{2} \|v - u\|^2 \qquad \forall\, u, v \in V$$

et que J' est Lipschitzien sur V: il existe $C > 0$ tel que

$$\|J'(v) - J'(u)\| \leq C \|v - u\| \qquad \forall\, u, v \in V \ .$$

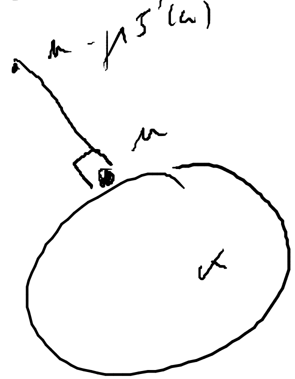
Alors, si $0 < \mu < 2\alpha/C^2$, l'algorithme de gradient projeté converge.

Preuve

nécessaire et suffisante car J convexe

Condition d'optimalité (inéquation d'Euler)

$$\langle J'(u), v - u\rangle \geq 0 \quad \forall\, v \in K\ .$$



Pour $\mu > 0$, c'est équivalent à

$$\langle u - \big(u - \mu J'(u)\big), v - u\rangle \geq 0 \quad \forall\, v \in K\ .$$

Autrement dit, l'unique point de minimum $u \in K$ est caractérisé par

$u = P_K\big(u - \mu J'(u)\big)$ avec P_K l'opérateur de projection orthogonale sur K.

On va montrer que l'application

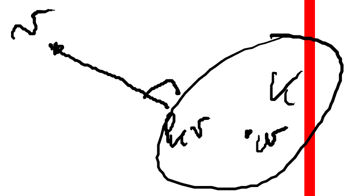
$$f(v) = P_K\big(v - \mu J'(v)\big)$$

est strictement contractante. Donc elle admet un unique point fixe $u = f(u)$ et la suite

$$u^{n+1} = P_K\big(u^n - \mu\, J'(u^n)\big) \quad \text{converge vers } u.$$

Définition de $P_K v$:

$$\langle P_K v - v, w - P_K v\rangle \geq 0$$



Preuve (suite)

L'opérateur de projection P_K est faiblement contractant

$$\|P_K v - P_K w\| \leq \|v - w\|.$$

On développe

$$\|P_K v - P_K w\|^2 = \underbrace{\langle P_K v - P_K w, v - w\rangle}_{\leq \|P_K v - P_K w\| \cdot \|v - w\|} + \underbrace{\langle P_K v - P_K w, P_K v - v\rangle}_{\substack{\leq 0 \\ \text{par déf. de } P_K v}} + \underbrace{\langle P_K v - P_K w, w - P_K w\rangle}_{\substack{\leq 0 \\ \text{par déf de } P_K w}}$$

$$\leq \|P_K v - P_K w\| \cdot \|v - w\|$$

Preuve (fin)

La fonction $v \to v - \mu J'(v)$ est strictement contractante car

$$\|v-\mu J'(v)-w+\mu J'(w)\|^2 = \|v-w\|^2-2\mu\langle v-w, J'(v)-J'(w)\rangle+\mu^2\|J'(v)-J'(w)\|^2$$

Grâce à l'α-convexité

$$\langle v-w, J'(v)-J'(w)\rangle \geq \alpha\|v-w\|^2$$

La Lipschitziannité de J' donne

$$\|J'(v)-J'(w)\|^2 \leq C^2\|v-w\|^2$$

Donc

$$\|v-\mu J'(v)-w+\mu J'(w)\|^2 \leq (1-2\mu\alpha+\mu^2C^2)\|v-w\|^2$$

Si $0 < \mu < \frac{2\alpha}{C^2}$ alors $(1-2\mu\alpha+\mu^2C^2) < 1$ et on conclut.

Exemples d'opérateurs de projection P_K

☞ Si $V = \mathbb{R}^n$ et $K = \prod_{i=1}^{n} [a_i, b_i]$, alors pour $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$

$$P_K(x) = y \quad \text{avec} \quad y_i = \min\left(\max\left(a_i, x_i\right), b_i\right) \quad \text{pour} \quad 1 \leq i \leq n \, .$$

☞ Si $V = \mathbb{R}^n$, B matrice $m \times n$ avec $\mathrm{rg}(B) = m \leq n$ et
$K = \{x \in \mathbb{R}^n \;\; Bx = c\}$, alors

$$P_K(x) = x + B^*(BB^*)^{-1}(c - Bx).$$

$\mathrm{Ker}\, B \perp \mathrm{Im}\, B^*$

$x - P_K x \in \mathrm{Im}\, B^*$

$P_K x = x + B^* p$

on calcule p en écrivant $B(P_K x) = c$

Remarque. Pour des convexes fermés K plus généraux, P_K peut être très difficile à déterminer en pratique.

Algorithme d'Uzawa

L'algorithme d'Uzawa recherche un point selle du Lagrangien $\mathcal{L}(v,q) = J(v) + q \cdot F(v)$ pour le problème de minimisation

$$\min_{F(v)\leq 0} J(v) \ ,$$

Rappel (Théorème de Kuhn et Tucker). Si J et F sont convexes et différentiables, alors il existe un point de minimum u avec multiplicateur de Lagrange p si et seulement si (u,p) est un point-selle du Lagrangien

$$\forall\, q \in \mathbb{R}_+^M \qquad \mathcal{L}(u,q) \leq \mathcal{L}(u,p) \leq \mathcal{L}(v,p) \qquad \forall\, v \in V \ .$$

Idée de l'algorithme: alternativement on minimise en v à q fixé, et on maximise en q à v fixé.

Algorithme d'Uzawa (suite)

Initialisation: soit $p^0 \in (\mathbb{R}_+)^M$.

Itérations: on construit les suites (u^n) et (p^n) par

$$\mathcal{L}(u^n, p^n) = \inf_{v \in V} \mathcal{L}(v, p^n) \ ,$$

$$p^{n+1} = P_{\mathbb{R}_+^M}(p^n + \mu F(u^n)) \ ,$$

avec $\mu > 0$ fixé (c'est une itération de l'algorithme du gradient projeté pour maximiser en q).

Remarques.

1. Variante: algorithme d'Arrow-Hurwicz (un seul pas de gradient pour la minimisation en v).
2. Peut être mis en oeuvre pour des contraintes d'égalité.

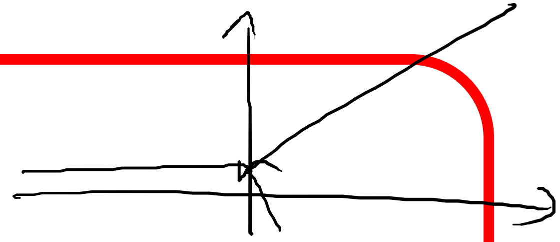
Algorithme d'Uzawa (fin)

Théorème. On suppose que J est α-convexe différentiable, que F est convexe et Lipschitzienne de V dans $\mathbb{R}^M$, $\exists\, C > 0$ tel que

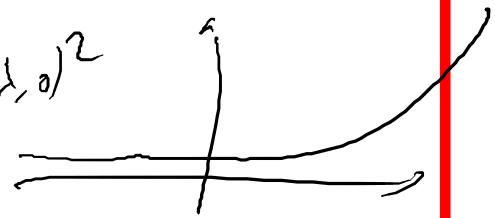
$$\|F(v) - F(w)\| \leq C\|v - w\| \quad \forall\, v, w \in V \ ,$$

et qu'il existe un point-selle (u, p) du Lagrangien sur $V \times (\mathbb{R}_+)^M$. Alors, si $0 < \mu < 2\alpha/C^2$, l'algorithme d'Uzawa converge : $\forall p^0 \in (\mathbb{R}_+)^M$, la suite (u^n) converge vers le point de minimum u.

$\lambda \to \max(0,\lambda)$



$\max(\lambda,0)^2$



Pénalisation des contraintes

Par exemple, on remplace

$$\inf_{v\in V,\ F(v)\leq 0} J(v)$$

par

$$\inf_{v\in V} \left(J(v) + \frac{1}{\epsilon} \sum_{i=1}^{M} \left[\max\left(F_i(v), 0\right)\right]^2 \right).$$

où on a **pénalisé** les contraintes avec un petit paramètre $\epsilon > 0$.

On utilise ensuite un algorithme sans contrainte.

Méthodes d'approximations successives.

Que faire pour résoudre

$$\inf_{v \in V,\ F(v)=0} J(v)$$

quand J et F n'ont pas de propriétés particulières ?

Idée: remplacer J et F par leurs développements de Taylor d'ordre 1 ou 2.

Initialisation: $v^0 \in V$. Itérations:

1. Soit v^n la solution approchée à l'itération n. On définit les approximations $J^n(v)$ et $F^n(v)$ au voisinage de v^n.
2. On calcule la solution v^{n+1} de

$$\inf_{v \in V,\ F^n(v)=0} J^n(v)$$

Remarque. On peut aussi imposer une condition de région de confiance $\|v - v^n\| \leq \epsilon$ pour être certain que l'approximation est correcte.

Programmation linéaire séquentielle (SLP)

On linéarise (Taylor d'ordre 1):

$$\inf_{F(v^{n-1})+F'(v^{n-1})\cdot(v-v^{n-1})=0} \left\{ J(v^{n-1}) + J'(v^{n-1}) \cdot (v - v^{n-1}) \right\},$$

Il existe des algorithmes très efficace pour ce problème linéaire.

On note v^n un point de minimum et on itère.

Remarque. Si l'infimum vaut $-\infty$, on rajoute la condition de région de confiance (linéaire) $\|v - v^{n-1}\|_\infty \leq \epsilon$.

Programmation quadratique séquentielle (SQP)

Approximation quadratique pour J, linéaire pour F:

$$\inf_{F^n(v)=0} J^n(v),$$

avec

$$F^n(v) = F(v^{n-1}) + F'(v^{n-1}) \cdot (v - v^{n-1})$$

$$J^n(v) = J(v^{n-1}) + J'(v^{n-1}) \cdot (v - v^{n-1}) + \frac{1}{2} Q^{n-1}(v - v^{n-1}) \cdot (v - v^{n-1})$$

mais Q^{n-1} n'est pas la Hessienne de J mais du Lagrangien

$$Q^{n-1} = J''(v^{n-1}) + \lambda^{n-1} \cdot F''(v^{n-1})$$

avec λ^{n-1} le multiplicateur de Lagrange pour l'itération $n-1$.

En effet, il faut rester sur la variété définie par la contrainte !

Autre vision méthode de Newton pour résoudre $J'(u) + \lambda \cdot F'(u) = 0$

$$J(v) \approx J(v_{n-1}) + \langle J'(v_{n-1}), v - v_{n-1} \rangle + \frac{1}{2} \langle J''(v_{n-1})(v - v_{n-1}), v - v_{n-1} \rangle$$

$$F(v) = 0 \approx F(v_{n-1}) + \langle F'(v_{n-1}), v - v_{n-1} \rangle + \frac{1}{2} \langle F''(v_{n-1})(v - v_{n-1}), v - v_{n-1} \rangle$$

$$J(v) = J(v) + \lambda F(v) \quad \text{si } F(v) = 0 \qquad \approx 0 \text{ si } \lambda = \lambda_{n-1}$$

$$J(v) + \lambda F(v) \approx J(v_{n-1}) + \lambda . F(v_{n-1}) + \langle J'(v_{n-1}) + \lambda F'(v_{n-1}), v - v_{n-1} \rangle$$

$$+ \frac{1}{2} \langle (J'' + \lambda . F'')(v_{n-1})(v - v_{n-1}), v - v_{n-1} \rangle$$

Notions de programmation linéaire.

On appelle **programme linéaire** sous forme standard

$$\inf_{x \in \mathbb{R}^n \text{ tel que } Ax=b,\ x \geq 0} c \cdot x,$$

où A est une matrice $m \times n$ (avec $\mathrm{rg}(A) = m$), $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$.

Lemme. Tout programme linéaire

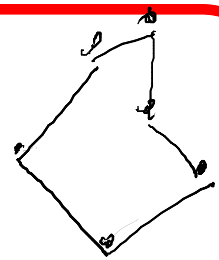
$$(*) \qquad \inf_{x \in \mathbb{R}^n \text{ tel que } \underbrace{Ax \geq b}_{m \times n \quad b \in \mathbb{R}^m},\ A'x=b'} c \cdot x.$$

peut se mettre sous la forme standard.

Preuve. On introduit des variables d'écart (slack variable) $y \in \mathbb{R}^m$ $y \geq 0$

$Ax = y + b \iff Ax \geq b$

$$(*) \iff \inf_{\substack{x \in \mathbb{R}^n\ y \in \mathbb{R}^m\ y \geq 0 \\ Ax = b + y\quad A'x = b'}} c \cdot x$$



Définitions

1) L'ensemble X_{ad} des vecteurs de $\mathbb{R}^n$ qui satisfont les contraintes

$$X_{ad} = \{x \in \mathbb{R}^n \text{ tel que } Ax = b,\ x \geq 0\},$$

est appelé ensemble des **solutions admissibles** (c'est un polyèdre).

2) Un point de minimum de

$$\inf_{x \in X_{ad}} c \cdot x$$

est appelé **solution optimale**.

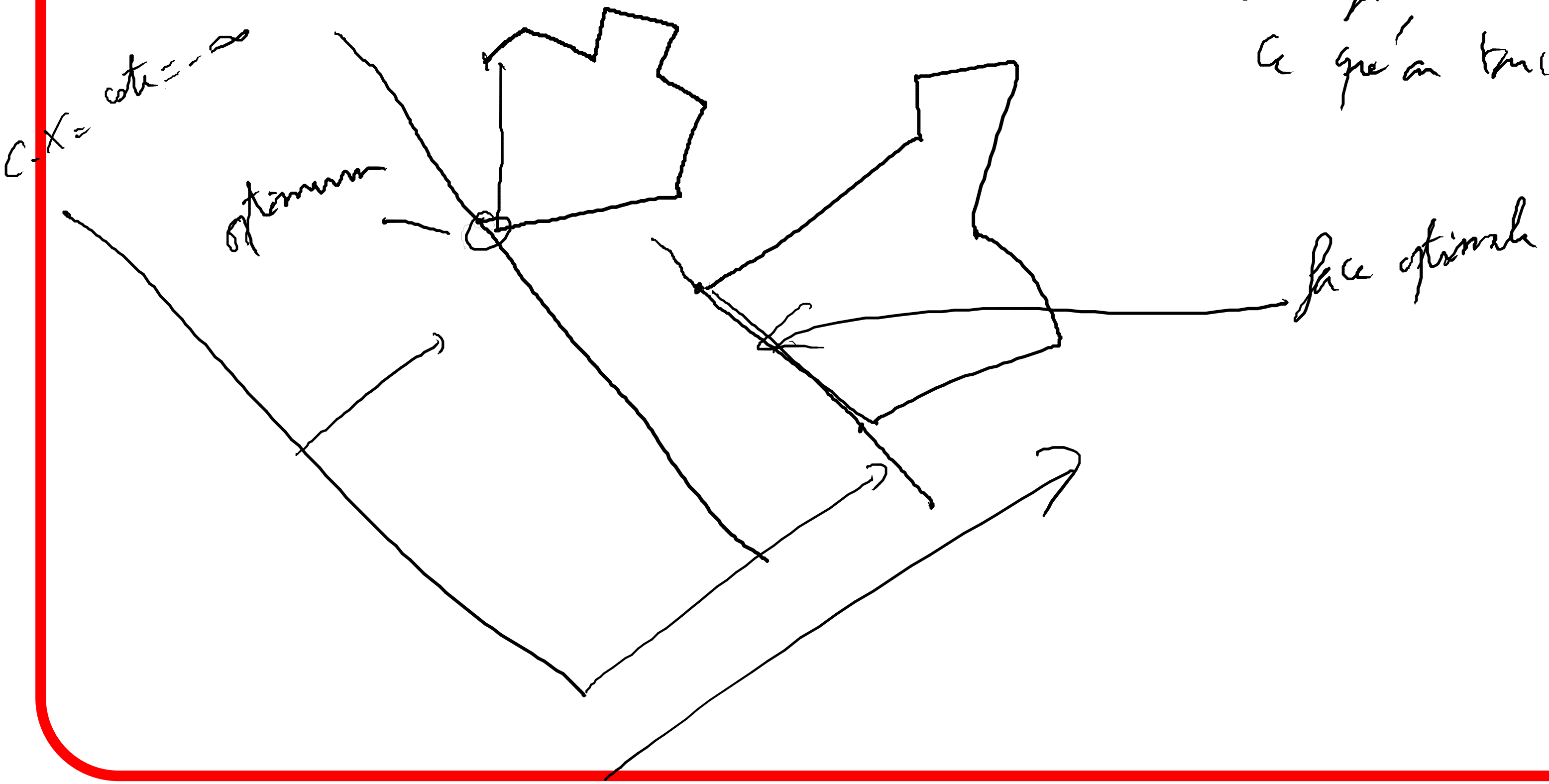
3) On appelle **sommet** ou point extrémal de X_{ad} tout point $\overline{x} \in X_{ad}$ qui ne peut pas se décomposer en une combinaison convexe de deux autres points de X_{ad}, c'est-à-dire que, s'il existe $y, z \in X_{ad}$ et $\theta \in]0, 1[$ tels que $\overline{x} = \theta y + (1 - \theta) z$, alors $y = z = \overline{x}$. Ce sont les sommets du polyèdre.

Proposition. S'il existe une solution optimale du programme linéaire, alors il existe une solution optimale qui est un sommet du polyèdre.

Remarque. Il y a un nombre fini de sommets. L'algorithme du simplexe de Dantzig énumère l'ensemble des sommets en faisant décroitre la fonction $c \cdot x$.

Preuve (géométrique).

en augmentant la cote de $-\infty$ jusqu'à ce qu'on touche le polyèdre



Rappels de dualité

Définition. Soit un Lagrangien $\mathcal{L}(v,q)$ défini sur $V \times P$.

$$\text{Pour } v \in V \text{ posons } \mathcal{J}(v) = \sup_{q \in P} \mathcal{L}(v,q).$$

$$\text{Pour } q \in P \text{ posons } \mathcal{G}(q) = \inf_{v \in V} \mathcal{L}(v,q).$$

On appelle problème primal

$$\inf_{v \in V} \mathcal{J}(v) \ ,$$

et problème dual

$$\sup_{q \in P} \mathcal{G}(q) \ .$$

Rappels de dualité

Lemme (dualité faible). On a toujours

$$\inf_{v \in U} \mathcal{J}(v) \geq \sup_{q \in P} \mathcal{G}(q).$$

Preuve: $\inf \sup \mathcal{L} \geq \sup \inf \mathcal{L}$.

Théorème (dualité forte). Le couple (u, p) est un point-selle de $\mathcal{L}$ sur $U \times P$ si et seulement si

$$\mathcal{J}(u) = \min_{v \in U} \mathcal{J}(v) = \max_{q \in P} \mathcal{G}(q) = \mathcal{G}(p) \ .$$

Remarque. Le problème dual est souvent plus simple que le primal. Si on peut résoudre le dual, on obtient la solution du primal grâce à une minimisation sans contrainte.

Application à la programmation linéaire

Soit le programme linéaire standard ou problème primal

$$\inf_{x\in\mathbb{R}^n \text{tel que } Ax=b,\ x\geq 0} c\cdot x \qquad (P)$$

avec $x, c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$ et A matrice $m \times n$. Pour $p \in \mathbb{R}^m$, le Lagrangien est

$$L(x,p) = c\cdot x + p\cdot(b - Ax),$$

où on ne "dualise" que les contraintes d'égalité. On introduit la fonction duale

$$G(p) = \min_{x\geq 0} L(x,p) = \begin{cases} p\cdot b & \text{si } A^*p - c \leq 0 \\ -\infty & \text{sinon.} \end{cases}$$

Le problème dual est donc

$$\sup_{p\in\mathbb{R}^m \text{tel que } A^*p-c\leq 0} p\cdot b \qquad (D)$$

$$\inf_{x \in \mathbb{R}^n \text{ tel que } Ax=b,\ x \geq 0} \quad c \cdot x \qquad (P)$$

$$\sup_{p \in \mathbb{R}^m \text{ tel que } A^* p - c \leq 0} \quad p \cdot b \qquad (D)$$

Théorème. Si (P) ou (D) a une valeur finie, alors il existe un point selle (x^*, p^*) du Lagrangien et x^* est une solution optimale de (P) tandis que x^* est une solution optimale de (D).

Le problème primal a n inconnues et m contraintes.

Le problème dual a m inconnues et n contraintes.

Une conséquence. Si $m << n$, alors le dual est plus facile que le primal.