C12, algorithmes pour l'optimisation avec contrainte

$f \in C(\mathbb{R}^n, \mathbb{R})$, K un sous ensemble fermé non vide de $\mathbb{R}^n$.
On suppose qu'il existe $\bar{x} \in K$ tel que $f(\bar{x}) \leq f(x)$ pour tout $x \in K$ (c'est-à-dire $\bar{x} \in argmin_K f$).

Objectif : calculer un tel point $\bar{x}$.

Nous allons distinguer plusieurs méthodes possibles, selon la nature de l'ensemble K

1. K convexe. GPF ou GPO avec projection sur K
2. K défini par des contraintes "égalités". Méthodes de Newton ou quasi-Newton.
3. K défini par des contraintes "inégalités". Méthodes de dualité.

Projection sur un ensemble convexe fermé non vide

K sous ensemble convexe fermé non vide de $\mathbb{R}^n$, $x \in \mathbb{R}^n$.

Pour $y \in \mathbb{R}^n$ on pose $h(y) = |y - x|^2$.

h est strictement convexe, $h(y) \to +\infty$ quand $|y| \to +\infty$.

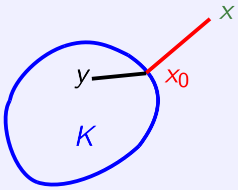
Donc,

Il existe un et un seul $x_0 \in K$ tel que $h(x_0) \leq h(y)$ pour tout $y \in K$. On note $x_0 = P_K(x)$.

Caractérisation de x_0 :

x_0 est l'unique point de K tel que $\nabla h(x_0) \cdot (y - x_0) \geq 0$ pour tout $y \in K$, c'est-à-dire

$$(x_0 - x) \cdot (y - x_0) \geq 0 \textit{ pour tout } y \in K.$$



L'opérateur P_K est contractant

K sous ensemble convexe fermé non vide de $\mathbb{R}^n$, P_K l'opérateur de projection sur K.

$x, y \in \mathbb{R}^n$,

$$|P_K(x) - P_K(y)|^2 = (P_K(x) - x + x - y + y - P_K(y)) \cdot (P_K(x) - P_K(y)) =$$
$$(P_K(x) - x) \cdot (P_K(x) - P_K(y)) + (x - y) \cdot (P_K(x) - P_K(y)) + (y - P_K(y)) \cdot (P_K(x) - P_K(y))$$

La caractérisation de P_K donne

$$(P_K(x) - x) \cdot (P_K(x) - P_K(y)) \leq 0, \ (y - P_K(y)) \cdot (P_K(x) - P_K(y)) \leq 0,$$

et donc, avec l'inégalité de Cauchy-Schwarz,

$$|P_K(x) - P_K(y)|^2 \leq (x - y) \cdot (P_K(x) - P_K(y)) \leq |x - y||P_K(x) - P_K(y)|.$$

Finalement, $|P_K(x) - P_K(y)| \leq |x - y|$.

Gradient à Pas Fixe avec projection sur K (GPFK)

$f \in C^1(\mathbb{R}^n, \mathbb{R})$, convexe et K convexe fermé non vide
On choisit $\rho > 0$

Initialisation $x^{(0)} \in K$.
Itération pour $k \geq 0$, on choisit $w^{(k)} = -\nabla f(x^{(k)})$.

$$x^{(k+1)} = P_K(x^{(k)} + \rho w^{(k)})$$

Si $x^{(k)} \to x$ quand $k \to +\infty$, alors $x \in argmin_K f$.
En effet, $x = P_K(x - \rho \nabla f(x))$ et donc, pour tout $y \in K$, la caractérisation de P_K donne, pour tout $y \in K$,

$$(x - (x - \rho \nabla f(x))) \cdot (y - x) \geq 0,$$

et donc $\nabla f(x)(y - x) \geq 0$.
Finalement par convexité de f, $f(y) \geq f(x) + \nabla f(x)(y - x) \geq f(x)$.
donc, $x \in argmin_K f$.
Question restante, la suite $(x^{(k)})_{k \geq 0}$ est-elle convergente ?

Convergence de l'algorithme GPFK

Théorème

Soient $f \in C^1(\mathbb{R}^n, \mathbb{R})$ et K convexe fermé non vide.
On suppose qu'il existe $\alpha > 0$ et $M > 0$ tels que

$$\text{(mon)} \quad (\nabla f(x) - \nabla f(y)) \cdot (x - y) \geq \alpha |x - y|^2 \text{ pour tout } x, y \in \mathbb{R}^n$$

$$\text{(lip)} \quad |\nabla f(x) - \nabla f(y)| \leq M |x - y| \text{ pour tout } x, y \in \mathbb{R}^n$$

Alors :

1. *Il existe un unique $\bar{x} \in K$ tel que $f(\bar{x}) \leq f(x)$ pour tout $x \in K$*
2. *Si $0 < \rho < \frac{2\alpha}{M^2}$, la suite donnée par l'algorithme (GPFK) converge vers $\bar{x}$*

Démonstration du premier item : l'hypothèse (mon) implique que f est strictement convexe et $f(x) \to +\infty$ quand $|x| \to +\infty$ et donc, comme K est convexe fermé non vide, il existe un unique $\bar{x} \in K$ tel que $f(\bar{x}) \leq f(x)$ pour tout $x \in X$.

Démonstration du deuxième item, convergence de l'algorithme (GPFK)

On pose $h(x) = x - \rho \nabla f(x)$ et $\bar{h}(x) = P_K(h(x))$.

L'algorithme (GPFK) est l'algorithme du point fixe pour $\bar{h}$:

$$x^{(k+1)} = P_K(x^{(k)} - \rho \nabla f(x^{(k)})) = P_K(h(x^{(k)}) = \bar{h}(x^{(k)})$$

On a déjà vu (Cours 10) que sous les hypothèses (mon)-(lip), si $0 < \rho < \frac{2\alpha}{M^2}$, la fonction h est strictement contractante.
Comme P_K est contractante, la fonction $\bar{h}$ est strictement contractante.
La suite donnée par l'algorithme (GPFK) converge donc vers l'unique point fixe de $\bar{h}$ et on déjà vu que ce point fixe est $\bar{x} = \mathrm{argmin}_K f$.

Gradient à Pas Optimal avec projection sur K (GPOK)

$f \in C^1(\mathbb{R}^n, \mathbb{R})$ et K convexe fermé non vide

Initialisation $x^{(0)} \in K$.

Itération pour $k \geq 0$,

1. on choisit $w^{(k)} = -\nabla f(x^{(k)})$,
2. on choisit (si c'est possible) $\rho_k > 0$ tel que
   $f(x^{(k)} + \rho_k w^{(k)}) \leq f(x^{(k)} + \rho w^{(k)})$ pour tout $\rho \geq 0$,
3. $x^{(k+1)} = P_K(x^{(k)} + \rho_k w^{(k)})$.

Questions :

1. Pour (GPOK). Existence de ρ_k ? Calcul de ρ_k ?
   A t'on $\lim_{k\to+\infty} x^{(k)} = \bar{x}$?
2. Pour (GPOK) et (GPFK), calcul de P_K ?
3. Que faire si K non convexe ?

Calcul de P_K, cas simples

Premier cas, $K = C^+ = \{y \in \mathbb{R}^n;\ y \geq 0\}$.
Soit $x \in \mathbb{R}^n$. On cherche le point de C^+ minimisant (sur C^+) la fonction $h(y) = |x - y|^2$, c'est-à-dire

$$h(y) = \sum_{i=1}^{n} (x_i - y_i)^2.$$

La solution consiste à prendre $y_i = x_i^+$, c'est-à-dire $P_K(x) = x^+$.

Deuxième cas, $K = \{y \in \mathbb{R}^n;\ \alpha_i \leq y_i \leq \beta_i, i = 1, \ldots, n\}$.
Soit $x \in \mathbb{R}^n$. On cherche le point de K minimisant (sur K) la fonction $h(y) = |x - y|^2$.
La solution consiste à prendre $y_i = \max\{\min\{x_i, \beta_i\}, \alpha_i\}$.

K défini par de contraintes "égalités"

$f \in C^1(\mathbb{R}^n, \mathbb{R})$, $g \in C^1(\mathbb{R}^n, \mathbb{R}^p)$.
$K = \{x \in \mathbb{R}^n;\ g(x) = 0\} = \{x \in \mathbb{R}^n;\ g_i(x) = 0,\ i = 1, \ldots, p\}$.

On cherche à calculer $\bar{x}$ tel que
$\bar{x} \in K$, $f(\bar{x}) \leq f(x)$ pour tout $x \in K$.
Une possibilité est de résoudre le système de $(n+p)$ équations à $(n+p)$ inconnues ($\bar{x}$ dans $\mathbb{R}^n$ et $\lambda_1, \ldots, \lambda_p$) :

$$\nabla f(x) + \sum_{i=1}^{p} \lambda_i \nabla g_i(x) = 0 \ (n \text{ équations}),$$
$$g(x) = 0 \ (p \text{ équations}).$$

Ceci peut se faire avec la méthode de Newton ou une méthode de quasi-Newton

Rappel : Si $rang(J_g(\bar{x})) = p$, on sait qu'il existe λ tel que $(\bar{x}, \lambda)$ est solution de ce système

K défini par de contraintes "inégalités"

$f \in C^1(\mathbb{R}^n, \mathbb{R})$, $g \in C^1(\mathbb{R}^n, \mathbb{R}^p)$.
$K = \{x \in \mathbb{R}^n;\ g(x) \leq 0\} = \{x \in \mathbb{R}^n;\ g_i(x) \leq 0,\ i = 1, \ldots, p\}$.

On cherche à calculer $\bar{x}$ tel que
Problème primal : $\bar{x} \in K$, $f(\bar{x}) \leq f(x)$ pour tout $x \in K$.

Une possibilité est de résoudre le système suivant :

$$
\begin{aligned}
&\nabla f(x) + \sum_{i=1}^{p} \lambda_i \nabla g_i(x) = 0,\\
&\lambda \cdot g(x) = 0,\\
&g(x) \leq 0,\\
&\lambda \geq 0.
\end{aligned}
$$

Rappel : Si $\{\nabla g_i(\bar{x}),\ i \in I(\bar{x})\}$ est une famille libre avec $I(\bar{x}) = \{i \in \{1, \ldots, p\};\ g_i(\bar{x}) = 0\}$, on sait qu'il existe λ tel que $(\bar{x}, \lambda)$ est solution de ce système

Définition du Lagrangien

$f \in C^1(\mathbb{R}^n, \mathbb{R})$, $g \in C^1(\mathbb{R}^n, \mathbb{R}^p)$, $K = \{x \in \mathbb{R}^n;\ g(x) \leq 0\}$.

Pour $x \in \mathbb{R}^n$ et $\lambda \in C^+$, on pose
$L(x, \lambda) = f(x) + \lambda \cdot g(x) = f(x) + \sum_{i=1}^p \lambda_i g_i(x)$

Et on introduit le problème de minimisation sans contrainte :

$$M(\lambda) = \inf_{x \in \mathbb{R}^n} L(x, \lambda)$$

Lien avec le théorème de Kuhn-Tucker :
Si $L(x_\lambda, \lambda) = M(\lambda)$ alors $\nabla f(x) + \sum_{i=1}^p \lambda_i \nabla g_i(x) = 0$

Problème dual

Pour $x \in \mathbb{R}^n$ et $\lambda \in C^+$,
$L(x, \lambda) = f(x) + \lambda \cdot g(x) = f(x) + \sum_{i=1}^{p} \lambda_i g_i(x)$

$$M(\lambda) = \inf_{x \in \mathbb{R}^n} L(x, \lambda)$$

La fonction M est concave car pour tout $x \in \mathbb{R}^n$ et $t \in [0, 1]$
$L(x, t\lambda + (1-t)\mu) = tL(x, \lambda) + (1-t)L(x, \mu) \geq tM(\lambda) + (1-t)M(\mu)$
et donc, en prenant en passant à la borne inférieure sur x,
$M(t\lambda + (1-t)\mu) \geq tM(\lambda) + (1-t)M(\mu)$

Ceci suggère le problème dit "dual"

Problème dual : $\bar{\lambda} \in C^+$, $M(\bar{\lambda}) \geq M(\lambda)$ pour tout $\lambda \in C^+$.

Point selle du Lagrangien

Théorème

$f \in C^1(\mathbb{R}^n, \mathbb{R})$, $g \in C^1(\mathbb{R}^n, \mathbb{R}^p)$.

Soit $(\bar{x}, \bar{\lambda}) \in \mathbb{R}^n \times C^+$, *point "selle" du Lagrangien, c'est-à-dire*

$$L(\bar{x}, \lambda) \leq L(\bar{x}, \bar{\lambda}) \leq L(x, \bar{\lambda}) \text{ pour tout } (x, \lambda) \in \mathbb{R}^n \times C^+$$

Alors $\bar{x}$ *est solution du problème primal*

$\bar{x} \in K$, $f(\bar{x}) \leq f(x)$ *pour tout* $x \in K$

et $\bar{\lambda}$ *est solution du problème dual*

$\bar{\lambda} \in C^+$, $M(\bar{\lambda}) \geq M(\lambda)$ *pour tout* $\lambda \in C^+$

Il faut des hypothèses supplémentaires sur f et g pour la réciproque

Intérêt : au lieu de chercher à résoudre directement la problème primal, difficile à cause de la contrainte d'appartenir à K, on va résoudre le problème dual, facile car P_{C^+} est facile à calculer

Démonstration du théorème

$L(x,\lambda) = f(x) + \lambda \cdot g(x)$. Soit $(\bar{x}, \bar{\lambda}) \in \mathbb{R}^n \times C^+$

$$L(\bar{x}, \lambda) \leq L(\bar{x}, \bar{\lambda}) \leq L(x, \bar{\lambda}) \textit{ pour tout } (x, \lambda) \in \mathbb{R}^n \times C^+$$

Etape 1: $L(\bar{x}, \lambda) \leq L(\bar{x}, \bar{\lambda})$ pour tout $\lambda \in C^+$ donne

$$(\lambda - \bar{\lambda}) \cdot g(\bar{x}) = \sum_{i=1}^{p} (\lambda_i - \bar{\lambda}_i) g_i(\bar{x}) \leq 0, \textit{ pour tout } \lambda \in C^+$$

Ceci donne $g_i(\bar{x}) \leq 0$ si $\bar{\lambda}_i = 0$ (en prenant $\lambda_i = 1$, $\lambda_j = \bar{\lambda}_j$ si $j \neq i$)
et $g_i(\bar{x}) = 0$ si $\bar{\lambda}_i > 0$ (en prenant $\lambda_i = 0$ et $\lambda_i = 2\bar{\lambda}_i$, $\lambda_j = \bar{\lambda}_j$ si $j \neq i$)
Donc $g(\bar{x}) \leq 0$ (et donc $\bar{x} \in K$) et $\bar{\lambda} \cdot g(\bar{x}) = 0$.

Etape 2: $\bar{x}$ est solution du problème primal car, si $x \in K$,

$$f(\bar{x}) = f(\bar{x}) + \bar{\lambda} \cdot g(\bar{x}) = L(\bar{x}, \bar{\lambda}) \leq L(x, \bar{\lambda}) = f(x) + \bar{\lambda} \cdot g(x) \leq f(x)$$

On peut aussi remarquer que $L(\bar{x}, \bar{\lambda}) = M(\bar{\lambda})$

Etape 3: $\bar{\lambda}$ est solution du problème dual car, si $\lambda \in C^+$,

$$M(\lambda) \leq L(\bar{x}, \lambda) \leq L(\bar{x}, \bar{\lambda}) = M(\bar{\lambda})$$

Algorithme d'Uzawa

L'algorithme d'Uzawa est l'algorithme (GPFK) pour le probème dual.
On se donne $\rho > 0$,
Initialisation $\lambda^{(0)} \in C^+$.
Itération pour $k \geq 0$,

$$\lambda^{(k+1)} = P_{C^+}(\lambda^{(k)} + \rho \nabla M(\lambda^{(k)}))$$

Deux questions :

1. Calcul de P_{C^+}. Facile (déjà vu). $P_{C^+}\lambda = \lambda^+$
2. Calcul de $\nabla M(\lambda)$ si $\lambda \in C^+$

Calcul de $\nabla M(\lambda)$

$L(x, \lambda) = f(x) + \lambda \cdot g(x)$.
On suppose qu'il existe un unique $x_\lambda \in \mathbb{R}^n$ tel que

$$M(\lambda) = L(x_\lambda, \lambda) \leq L(x, \lambda) \textit{ pour tout } x \in \mathbb{R}^n,$$

et que $\Phi : \lambda \mapsto x_\lambda$ est dérivable.
On a alors

$$\nabla M(\lambda) = J_\Phi(\lambda)^t \nabla_1 L(x_\lambda, \lambda) + \nabla_2 L(x_\lambda, \lambda) = g(x_\lambda)$$

car $\nabla_1 L(x_\lambda, \lambda) = 0$ (minimisation sans contrainte) et $\nabla_2 L(x_\lambda, \lambda) = g(x_\lambda)$

Algorithme d'Uzawa en pratique

On se donne $\rho > 0$,
Initialisation $\lambda^{(0)} \in C^+$.
Itération pour $k \geq 0$,

1. On cherche $x^{(k)} \in \mathbb{R}^n$ tel $L(x^{(k)}, \lambda^{(k)}) \leq L(x, \lambda^{(k)})$ pour tout $x \in \mathbb{R}^n$
2. On calcule $g(x^{(k)})$ (qui est égal à $\nabla M(\lambda^{(k)})$)
3. $\lambda^{(k+1)} = P_{C^+}(\lambda^{(k)} + \rho g(x^{(k)}))$

Convergence de l'algorithme d'Uzawa

Théorème

$f \in C^1(\mathbb{R}^n, \mathbb{R})$, $g \in C^1(\mathbb{R}^n, \mathbb{R}^p)$, $K = \{x \in \mathbb{R}^n;\ g(x) \leq 0\}$, $K \neq \emptyset$.

On suppose qu'il existe $\alpha > 0$, C *et* d *tels que*

(mon) $(\nabla f(x) - \nabla f(y)) \cdot (x - y) \geq \alpha |x - y|^2$ *pour tout* $x, y \in \mathbb{R}^n$

(aff) $g(x) = Cx - d$ *avec* $C \in \mathcal{M}_{p,n}$ *et* $d \in \mathbb{R}^p$ *(contraintes affines)*

(KT) *Soit* $\bar{x} = \mathrm{argmin}_K f$. *Il existe* $\lambda \in C^+$ *tel que*
$\nabla f(\bar{x}) + \sum_{i=1}^p \lambda_i \nabla g_i(\bar{x}) = 0$, $\lambda \cdot g(\bar{x}) = 0$

Alors, si $0 < \rho < \frac{2\alpha}{\|C\|^2}$, *la suite* $(x^{(k)}, \lambda^{(k)})_{k \geq 0}$ *donnée par l'algorithme d'Uzawa vérifie*

1. $\lim_{k \to \infty} x^{(k)} = \bar{x}$, $\bar{x}$ *unique solution du problème primal*
2. *la suite* $(\lambda^{(k)})_{k \geq 0}$ *est bornée*

Démonstration dans le td12

Remarques sur les hypothèses du théorème

$f \in C^1(\mathbb{R}^n, \mathbb{R})$, $g \in C^1(\mathbb{R}^n, \mathbb{R}^p)$, $K = \{x \in \mathbb{R}^n;\ g(x) \leq 0\}$, $K \neq \emptyset$.

(mon) $(\nabla f(x) - \nabla f(y)) \cdot (x - y) \geq \alpha |x - y|^2$ *pour tout* $x, y \in \mathbb{R}^n$

(aff) $g(x) = Cx - d$ avec $C \in \mathcal{M}_{p,n}$ et $d \in \mathbb{R}^p$ (contraintes affines)

(KT) Soit $\bar{x} = argmin_K f$. Il existe $\lambda \in C^+$ tel que $\nabla f(\bar{x}) + \sum_{i=1}^p \lambda_i \nabla g_i(\bar{x}) = 0$, $\lambda \cdot g(\bar{x}) = 0$

1. Les hypothèses (mon) et (aff) permettent de montrer que le problème primal a une unique solution, notée $\bar{x}$, et que pour tout $\lambda \in C^+$, il existe un et un seul $x_\lambda \in \mathbb{R}^n$ tel que $L(x_\lambda, \lambda) \leq L(x, \lambda)$ pour tout $x \in \mathbb{R}^n$
2. l'hypothèse (KT) est vérifiée si $\{\nabla g_i(\bar{x}),\ i \in I(\bar{x})\}$ est une famille libre avec $I(\bar{x}) = \{i \in \{1, \ldots, p\};\ g_i(\bar{x}) = 0\}$