Optimisation sans contraintes

Gabriel STOLTZ

stoltz@cermics.enpc.fr

(CERMICS, Ecole des Ponts & Equipe-projet MATHERIALS, INRIA Rocquencourt)

Calcul scientifique, Ecole des Ponts, 19 février 2015

Deux exemples tirés du cours d'Économie

Consommation : maximisation d'utilité sous contrainte de budget

$$\sup_{x \in K} U(x_1 \dots x_N), \qquad K = \left\{ x \in \mathbb{R}^N;\ x_i \geqslant 0,\ \sum_{i=1}^N p_i x_i \leqslant R \right\}$$

U : fonction d'utilité, x_i : quantité de bien i consommée,
p_i : prix du bien i, R : budget disponible

Production : minimisation du coût à contrainte de production fixée

$$\inf_{x \in K} \left(\sum_{i=1}^N p_i x_i \right), \qquad K = \left\{ x \in \mathbb{R}^N;\ x_i \geqslant 0,\ f(x_1 \dots x_N) = y \right\}$$

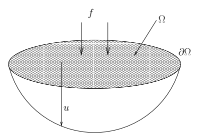
x_i : facteur de production, p_i : prix du facteur de production i,
y : quantité de bien à produire, f : fonction de production

Exemples tiré des cours de mécanique

Membrane élastique à l'équilibre sous un chargement vertical f

$$\inf_{v \in V} \mathcal{E}(v), \qquad V = H_0^1(\Omega), \qquad \mathcal{E}(v) = \frac{1}{2} \int_\Omega |\nabla v|^2 - \int_\Omega f v$$

u : déplacement (vertical), Ω : configuration de référence



Minimisation de l'énergie électronique de l'atome d'hydrogène

$$\inf_{\psi \in K} \left(\int_{\mathbb{R}^3} \frac{1}{2} |\nabla \psi(x)|^2 - \frac{|\psi(x)|^2}{|x|}\, dx \right), \qquad K = \left\{ \psi \in H^1(\mathbb{R}^3, \mathbb{C}) \,\middle|\, \|\psi\|_{L^2} = 1 \right\}$$

Formalisation mathématique

Un peu de vocabulaire (1)

- On considère...
  - un espace de Hilbert V, de dimension finie ou infinie
  - une fonctionnelle $J : V \to \mathbb{R}$: critère ou fonction coût

Problème d'optimisation sans contrainte

Chercher u tel que

$$u \in V \qquad \text{et} \qquad J(u) \leqslant J(v) \quad \forall v \in V$$

- On a donc $J(u) = \inf_{v \in V} J(v) = \min_{v \in V} J(v)$
- On dit que u est un minimiseur global de J sur V

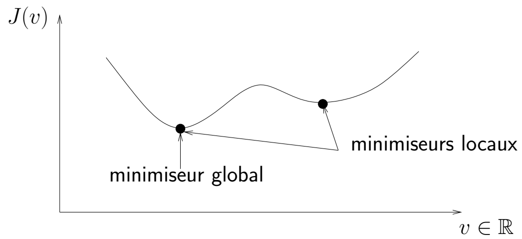
Un peu de vocabulaire (2)

- On dit que u est un minimiseur local de J sur V si

$$u \in V \quad \text{et} \quad \exists \delta > 0, \; J(u) \leqslant J(v) \; \forall v \in \mathcal{B}_V(u, \delta)$$

où $\mathcal{B}_V(u, \delta) = \{v \in V; \|u - v\|_V \leqslant \delta\}$

- Exemple 1D: $V = \mathbb{R}$ si bien que $J : \mathbb{R} \to \mathbb{R}$



- On ne considère que des problèmes de minimisation
$\to$ maximiser J revient à minimiser $-J$!

Un peu de vocabulaire (3)

- Ensemble des états admissibles : sous-ensemble non-vide $K \subset V$

Problème d'optimisation avec contraintes

Chercher u tel que

$$u \in K \qquad \text{et} \qquad J(u) \leqslant J(v) \quad \forall v \in K$$

- On a donc $J(u) = \inf_{v \in K} J(v) = \min_{v \in K} J(v)$
- On dit que u est un minimiseur global de J sur K
- On dit que u est un minimiseur local de J sur K si

$$u \in K \quad \text{et} \quad \exists \delta > 0, \; J(u) \leqslant J(v) \; \forall v \in \mathcal{B}_V(u, \delta) \cap K$$

- **Question :** les exemples du début du cours sont-ils en dimension finie ou infinie ? Avec ou sans contrainte ?

Objectifs du chapitre d'optimisation

- **Minimisation :** avec et sans contraintes, en dimension finie ou infinie
- **Etude théorique :**
  - connaitre des conditions suffisantes d'existence et unicité du minimiseur global
  - Établir des conditions nécessaires d'optimalité permettant de caractériser un minimiseur (local)
- **Méthodes numériques :** fondées sur ladite caractérisation
- **Contraintes :** deux situations génériques

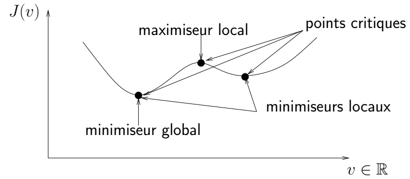
$$K = \begin{cases} \{v \in V; \; \Phi(v) = 0\} & \text{contraintes égalité} \\ \{v \in V; \; \Phi(v) \leqslant 0\} & \text{contraintes inégalité} \end{cases}$$

où $\Phi : V \to \mathbb{R}^m$ et $\Phi(v) \leqslant 0$ signifie $\Phi_i(v) \leqslant 0, \; \forall i \in \{1 \dots m\}$

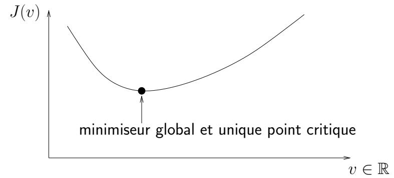
Outils techniques : différentiation et convexité

Différentiation et convexité : motivation 1D

- $V = \mathbb{R}$ et $J : \mathbb{R} \to \mathbb{R}$, supposée dérivable
- Si v est minimiseur local de J sur V, alors $J'(v) = 0$. On dit que v est point critique de J



- Être point critique est condition nécessaire d'optimalité locale
- Si J convexe : être point critique est une condition suffisante d'optimalité globale



Différentiation (1)

- Soit V un espace de Hilbert et $J : V \to \mathbb{R}$

Différentiabilité

On dit que J est différentiable en $v \in V$ s'il existe une forme linéaire continue $J'(v) \in V'$ telle que, pour tout $w \in V$,

$$J(v+w) = J(v) + \langle J'(v), w \rangle_{V',V} + \mathrm{o}(\|w\|_V)$$

- **En pratique** :
  - développer $J(v+w)$ autour de $J(v)$
  - regrouper les termes d'ordre 1 en w et les termes d'ordre supérieur

$$J(v+w) = J(v) + T_1(v,w) + T_2(v,w)$$

  - vérifier que $T_1(v,w)$ est linéaire et continue en $w \in V$
  - montrer que $T_2(v,w)/\|w\|_V \to 0$ lorsque $w \to 0$ dans V

Différentiation (2)

- J est différentiable sur V si J est différentiable en tout $v \in V$
- La différentielle de J est l'application $J' : V \to V'$
- **Exemple :** en dimension finie avec $V = \mathbb{R}^N$, on a

$$\forall w = (w_1 \dots w_N) \in \mathbb{R}^N, \quad \langle J'(v), w \rangle_{V',V} = \sum_{i=1}^{N} \frac{\partial J}{\partial v_i}(v) w_i = \Big(\nabla J(v), w\Big)_V$$

$\to$ on peut ainsi identifier différentielle et gradient

- Relation différentielle/gradient plus subtile en dimension infinie
$\to$ **utiliser plutôt la différentielle**

Convexité (1)

- Soit V un espace vectoriel réel

Définition de la convexité

La fonctionnelle $J : V \to \mathbb{R}$ est convexe sur V si, $\forall (v,w) \in V \times V$, $\forall \theta \in [0,1]$,

$$J(\theta v + (1-\theta) w) \leqslant \theta J(v) + (1-\theta) J(w)$$

- Soit V un espace vectoriel réel normé

Définition de la forte convexité

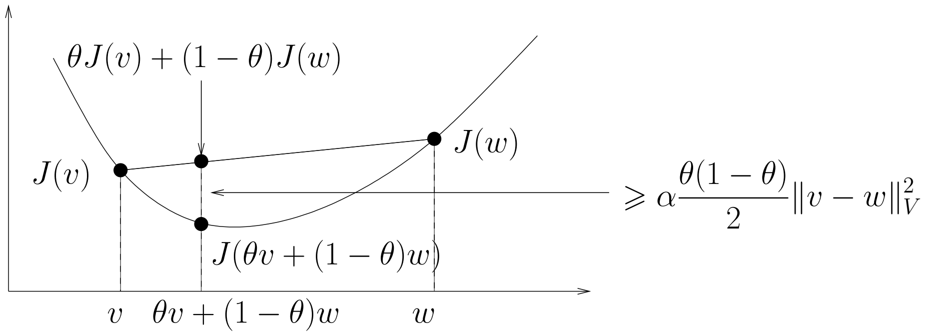
La fonctionnelle $J : V \to \mathbb{R}$ est fortement convexe de paramètre $\alpha > 0$ sur V si, $\forall (v,w) \in V \times V$, $\forall \theta \in [0,1]$,

$$J(\theta v + (1-\theta) w) \leqslant \theta J(v) + (1-\theta) J(w) - \alpha \frac{\theta(1-\theta)}{2} \|v - w\|_V^2$$

On dit aussi que J est α-convexe

Convexité (2)

- Illustration 1D de la forte convexité : $V = \mathbb{R}$, $J : \mathbb{R} \to \mathbb{R}$

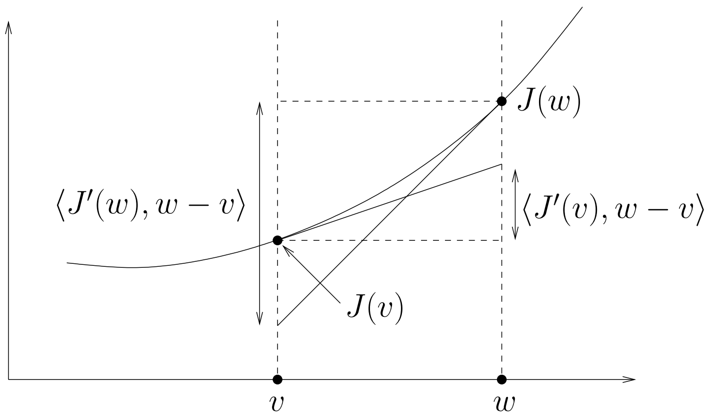


Convexité : caractérisations alternatives (1)

- Fonctionnelles $J : V \to \mathbb{R}$ différentiables
- La fonctionnelle J est convexe sur V si et seulement si
  - $\forall (v,w) \in V \times V,\ J(w) \geqslant J(v) + \langle J'(v), w - v \rangle_{V',V}$
  - $\forall (v,w) \in V \times V,\ \langle J'(w) - J'(v), w - v \rangle_{V',V} \geqslant 0$
- La fonctionnelle J est α-convexe sur V si et seulement si
  - $\forall (v,w) \in V \times V,\ J(w) \geqslant J(v) + \langle J'(v), w - v \rangle_{V',V} + \frac{\alpha}{2} \|w - v\|_V^2$
  - $\forall (v,w) \in V \times V,\ \langle J'(w) - J'(v), w - v \rangle_{V',V} \geqslant \alpha \|w - v\|_V^2$
- **En pratique, les caractérisations "symétriques" sont souvent les plus utiles !**

Convexité : caractérisations alternatives (2)

- Illustration 1D : $V = \mathbb{R}$, $J : \mathbb{R} \to \mathbb{R}$



Convexité : un exemple

- V Hilbert, $u \in V$ fixé, $J : V \to \mathbb{R}$ avec $J(v) = \dfrac{1}{2}\|v-u\|_V^2$
- Montrons que J est 1-convexe (avec égalité)
- Calcul direct : considérer $(v,w) \in V \times V$, $\theta \in [0,1]$, et se lancer...

$$\begin{aligned}
J(\theta v + (1-\theta)w) &= \frac{1}{2}\|\theta v + (1-\theta)w - u\|_V^2 = \frac{1}{2}\|\theta(v-u) + (1-\theta)(w-u)\|_V^2 \\
&= \frac{1}{2}\theta^2\|v-u\|_V^2 + \frac{1}{2}(1-\theta)^2\|w-u\|_V^2 + \theta(1-\theta)(v-u,w-u)_V \\
&= \theta J(v) + (1-\theta)J(w) - \frac{1}{2}\theta(1-\theta)\left\{\|v-u\|_V^2 + \|w-u\|_V^2 - 2(v-u,w-u)_V\right\} \\
&= \theta J(v) + (1-\theta)J(w) - \frac{1}{2}\theta(1-\theta)\|v-w\|_V^2
\end{aligned}$$

- Utilisation de la différentielle : $\langle J'(v), w\rangle_{V',V} = (v-u,w)_V$

On a $\langle J'(w) - J'(v), w-v\rangle_{V',V} = \Big((w-u)-(v-u), w-v\Big)_V = \|w-v\|_V^2$

Minimisation sans contrainte : résultats théoriques

Existence d'un minimiseur global

- Problème $\inf\limits_{v\in V} J(v)$ pour V espace vectoriel

Conditions suffisantes d'existence d'un minimiseur global

- V est un espace de Hilbert
- J est continue $V \to \mathbb{R}$
- J est coercive (*i.e.* $J(v) \to +\infty$ lorsque $\|v\|_V \to +\infty$)
- J est convexe (inutile si V est de dimension finie)

- Quelques commentaires sur ces hypothèses...
  - Il peut y avoir plusieurs minima globaux/une infinité. **Exemple ?**
  - Si J est différentiable alors elle est continue (utile en pratique)
  - Convexité n'implique pas continuité en dimension infinie
  - J pas coercive : infimum (et pas minimum) en général. **Exemple ?**

Existence/unicité du minimiseur global

- Unicité du minimiseur si J strictement convexe

$\forall (v,w) \in V \times V, \quad \forall \theta \in]0,1[, \qquad J(\theta v + (1-\theta)w) < \theta J(v) + (1-\theta)J(w)$

Preuve : $u_1 \neq u_2$ deux minimiseurs, contradiction car

$$J\left(\frac{u_1+u_2}{2}\right) < \frac{1}{2}J(u_1) + \frac{1}{2}J(u_2) = \inf_{v\in V} J(v)$$

- Si J est fortement convexe, alors J est strictement convexe et il existe $\gamma > 0$, $\delta \in \mathbb{R}$ tels que

$$\forall v \in V, \qquad J(v) \geqslant \gamma\|v\|_V^2 + \delta$$

- Résultat très utile en pratique :

Conditions suffisantes d'existence/unicité du minimiseur global

- V est un espace de Hilbert
- J est continue $V \to \mathbb{R}$
- J est α-convexe

Conditions nécessaires voire suffisantes

Condition d'Euler (1)

- On dit que v est point critique de J si $J'(v) = 0$ $(\in V')$, *i.e.*

$$\forall w \in V, \qquad \langle J'(v), w\rangle_{V',V} = 0$$

Condition nécessaire pour être minimiseur local

V espace de Hilbert, $J : V \to \mathbb{R}$ fonctionnelle différentiable

$$u \text{ minimiseur local de } J \implies u \text{ point critique de } J : \ J'(u) = 0 \ (\in V')$$

- La condition $J'(u) = 0$ porte le nom de condition d'Euler
- **Preuve :** Soit $v \in V$ arbitraire. Pour $t \in \mathbb{R}$ assez petit,

$$J(u) \leqslant J(u + tv) = J(u) + t\langle J'(u), v\rangle_{V',V} + \mathrm{o}(t)$$

  - limite $t \to 0^+$: $\langle J'(u), v\rangle_{V',V} \geqslant 0$
  - changer v en $-v$: $\langle J'(u), v\rangle_{V',V} = 0$
  - v étant arbitraire, on a $J'(u) = 0$ $(\in V')$

Condition d'Euler (2)

Condition nécessaire et suffisante

Si J est convexe, la condition d'Euler est une condition suffisante d'optimalité globale

$$J'(u) = 0 \ (\in V') \implies u \text{ est un minimiseur global de } J \text{ sur } V$$

- Conséquence immédiate du fait que, par convexité de J,

$$\begin{aligned}\forall v \in V, \qquad J(v) &\geqslant J(u) + \langle J'(u), v - u\rangle_{V',V} \\ &\geqslant J(u) + 0 \\ &= J(u)\end{aligned}$$

Cas des fonctionnelles quadratiques (1)

- En dimension finie : $V = \mathbb{R}^N$, $A \in \mathbb{R}^{N,N}$, $b \in \mathbb{R}^N$ et

$$J(v) = \frac{1}{2}(v, Av)_{\mathbb{R}^N} - (b, v)_{\mathbb{R}^N} = \frac{1}{2}\sum_{i,j=1}^N A_{ij} v_i v_j - \sum_{i=1}^N b_i v_i$$

- Par un calcul facile : $\displaystyle \frac{\partial J}{\partial v_i}(v) = \frac{1}{2}\sum_{j=1}^N (A_{ij} + A_{ji}) v_j - b_i$ $(1 \leqslant i \leqslant N)$, soit

$$\nabla J(v) = \frac{1}{2}\left(A + A^{\mathrm{T}}\right) v - b$$

Lorsque A est symétrique

Résoudre le système linéaire $Av = b$ revient à chercher un point critique de J (car $\nabla J(v) = Av - b$)

Cas des fonctionnelles quadratiques (2)

- Dimension infinie : $J(v) = \frac{1}{2}a(v,v) - \langle b, v\rangle_{V',V}$
- a forme bilinéaire continue sur $V \times V$, b forme linéaire continue sur V
- J est différentiable sur V avec

$$\langle J'(v), w\rangle_{V',V} = \frac{1}{2}\Big(a(v,w) + a(w,v)\Big) - \langle b, w\rangle_{V',V} \qquad \forall w \in V$$

Lorsque a est symétrique $(a(v,w) = a(w,v))$

Résoudre le problème $a(v,w) = \langle b, w\rangle_{V',V}$ $\forall w \in V$, revient à chercher un point critique de J

- Si a symétrique : a coercive (de paramètre α) ssi J est α-convexe

$$\begin{aligned}\langle J'(w) - J'(v), w - v\rangle_{V',V} &= a(w, w-v) - \langle b, w-v\rangle_{V',V} \\ &\quad - a(v, w-v) + \langle b, w-v\rangle_{V',V} \\ &= a(w-v, w-v) \geqslant \alpha\|w-v\|_V^2\end{aligned}$$

Méthodes numériques

Méthode de gradient (1)

- **Dimension finie :** discrétisation du problème sur une base idoine, fonctionnelle J différentiable

- **Objectif :** construire une suite $(v^k)_{k\in\mathbb{N}} := v^0, v^1, \dots$ telle que

$$v^k \to v \qquad \text{où} \qquad \nabla J(v) = 0$$

On construit donc un point critique de J de manière itérative

- **Principe :** pour $v^k \in V$ donné, chercher une direction de descente $d^k \in V$ telle que

$$\text{pour } t \text{ suffisamment petit,} \qquad J(v^k + td^k) \leqslant J(v^k)$$

Comme $J\left(v^k + td^k\right) = J(v^k) + t\left(\nabla J(v^k), d^k\right)_V + \mathrm{o}(t)$, on peut proposer

$$d^k = -\nabla J(v^k)$$

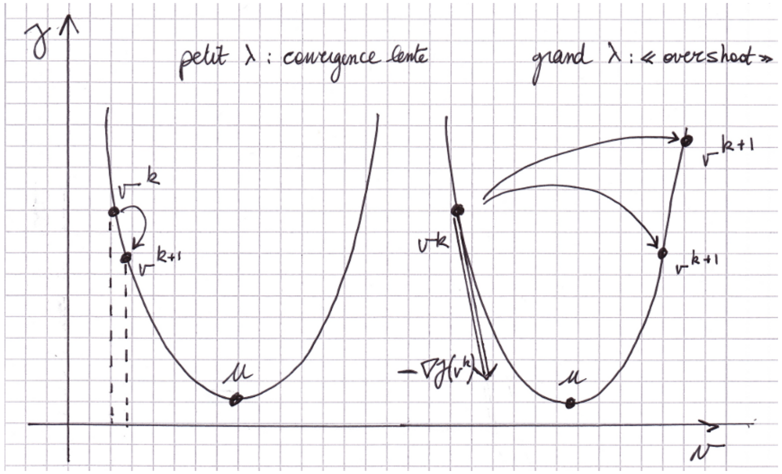
Méthode de gradient (2)

- **Initialisation**
  - choisir $v^0 \in V$ et poser $k := 0$
  - choisir le pas $\lambda > 0$
  - fixer un seuil de convergence $\varepsilon > 0$
- **Itérations** (boucle sur k)
  - calculer $\nabla J(v^k)$
  - choisir comme direction de descente $d^k = -\nabla J(v^k)$
  - déterminer v^{k+1} selon la formule

$$v^{k+1} = v^k + \lambda d^k$$

  - test de convergence : $\dfrac{\|v^{k+1} - v^k\|_V}{\|v^0\|_V} \leqslant \varepsilon$ ou $\dfrac{|J(v^{k+1}) - J(v^k)|}{J(v^0)} \leqslant \varepsilon$
- Méthode de **gradient à pas fixe** : choix de λ ? convergence ?

Méthode de gradient (3)



- La convergence est très lente si λ est trop petit
- Si λ est trop grand, on peut ne pas converger !

En résumé et à venir...

- **Aujourd'hui** : optimisation sans contraintes
  - exemples de problèmes d'optimisation, vocabulaire
  - différentiabilité, convexité
  - existence/unicité de minimiseurs (optimisation sans contraintes)
  - conditions nécessaires (voire suffisantes) d'optimalité
  - algorithme de gradient
- **Jeudi 5 mars** : optimisation avec contraintes
  - existence/unicité de minimiseurs
  - conditions nécessaires : inéquations d'Euler–Lagrange
- **Jeudi 12 mars** : méthodes numériques
  - convergence de l'algorithme du gradient
  - gradient projeté (contraintes)
- **TP informatique** : jeudi 12 mars
- **TDs** : 5 et 26 mars