# OPTIMISATION ET CONTRÔLE

Grégoire Allaire
Département de mathématiques appliquées
Ecole Polytechnique

1. Conditions d'optimalité avec contraintes d'inégalité
2. Lagrangien, point selle, théorème de Kuhn et Tucker, dualité
3. Algorithmes d'optimisation sans contraintes (début)

# I - Conditions d'optimalité avec contraintes d'inégalité

Soit $F(v) = (F_1(v), ..., F_M(v))$: $V \rightarrow \mathbb{R}^M$. On considère

$$\inf_{v\in K} J(v) \quad \text{avec} \quad K = \{v \in V,\ F(v) \leq 0\}$$

où $F(v) \leq 0$ signifie que $F_i(v) \leq 0$ pour $1 \leq i \leq M$.

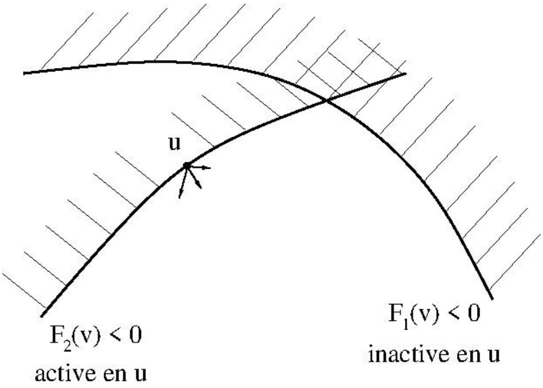
**Définitions.** Soit $u$ tel que $F(u) \leq 0$.

1. On appelle ensemble des contraintes actives en $u$
   $I(u) = \{i \in \{1, \ldots, M\}, F_i(u) = 0\}$.
2. On dit que les contraintes d'inégalité sont qualifiées en $u \in K$ si la famille $(F_i'(u))_{i\in I(u)}$ est libre.

**Remarque.** Il existe d'autres conditions de qualification, plus générales mais plus compliquées (voir plus loin).

# Contraintes et variations admissibles

- Les contraintes inactives $F_i(u) < 0$ ne jouent aucun rôle.
- Pour les contraintes actives $F_i(u) = 0$ on peut faire des variations dans certaines directions (mais pas dans d'autres).



# Conditions d'optimalité

**Théorème.** Soit $u \in V$ tel que $F(u) \leq 0$. On suppose que $J$ et les contraintes $(F_i)_{1 \leq i \leq M}$ sont dérivables en $u$. On suppose de plus que les contraintes sont qualifiées en $u$. Alors, si $u$ est un minimum local de $J$ sur $K$, il existe $\lambda_1, \ldots, \lambda_M \in \mathbb{R}^+$, appelés **multiplicateurs de Lagrange**, tels que

$$J'(u) + \sum_{i=1}^{M} \lambda_i F_i'(u) = 0, \quad \lambda_i \geq 0, \quad F_i(u) \leq 0,$$
$$\lambda_i = 0 \ \text{ si } \ F_i(u) < 0.$$

**Remarque.** Les conditions sur les $\lambda_i$ peuvent se réécrire

$$\lambda \cdot F(u) = 0, \quad \lambda \geq 0 \ \text{ et } \ F(u) \leq 0.$$

et s'appellent conditions de complémentarité.

**Remarque.** Pour calculer les multiplicateurs de Lagrange $\lambda_i$ des contraintes actives $i \in I(u)$ on peut utiliser les égalités $F_i(u) = 0$.

# Preuve

Soit $K(u)$ le cône des directions admissibles au point $u$

$$K(u) = \left\{ \begin{array}{l} w \in V \text{ tel que } \exists\, (v^n)_{n\geq 0} \in K\,,\ \exists\, (\varepsilon^n)_{n\geq 0} \in \mathbb{R}_+^* \text{ vérifiant} \\ \displaystyle\lim_{n\to+\infty} v^n = u\,,\ \lim_{n\to+\infty} \varepsilon^n = 0\,,\ \lim_{n\to+\infty} \frac{v^n - u}{\varepsilon^n} = w \end{array} \right\}$$

**Rappel (Proposition 2.5.5):** si $u$ est un minimum local, alors

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

La preuve est constituée de deux étapes:

1. On caractérise $K(u)$ en montrant que $K(u) = \tilde{K}(u)$ avec

$$\tilde{K}(u) = \left\{ w \in V \text{ tel que } \langle F_i'(u), w\rangle \leq 0 \quad \forall\, i \in I(u) \right\}\,.$$

2. On conclut grâce au lemme de Farkas.

# Preuve (suite)

Montrons que $K(u) = \tilde{K}(u) = \{w \in V, \langle F_i'(u), w\rangle \leq 0, i \in I(u)\}$.

Soit $w \in K(u)$. Il existe $v^n \in K$ et $\varepsilon^n > 0$ tel que $F(v^n) \leq 0$,

$$\lim_{n\to+\infty} v^n = u, \quad \lim_{n\to+\infty} \varepsilon^n = 0, \quad \lim_{n\to+\infty} (v^n - u)/\varepsilon^n = w.$$

Un développement de Taylor conduit à

$$(1) \qquad 0 \geq F_i(v^n) = F_i(u) + \varepsilon^n \langle F_i'(u), w\rangle + o(\varepsilon^n).$$

Si $F_i(u) < 0$, alors (1) ne donne aucune information sur $w$. Par contre, si $F_i(u) = 0$ (contrainte active), alors (1) implique

$$0 \geq \langle F_i'(u), w\rangle \quad \text{c'est-à-dire que } w \in \tilde{K}(u).$$

Réciproquement, $w \in Int\tilde{K}(u) = \{w \in V, \langle F_i'(u), w\rangle < 0, i \in I(u)\}$

$$F_i(u + \varepsilon^n w) = F_i(u) + \varepsilon^n \langle F_i'(u), w\rangle + o(\varepsilon^n) \leq 0$$

Donc $Int\tilde{K}(u) \subset K(u) \subset \tilde{K}(u)$ et on conclut car $K(u)$ est fermé.

# Preuve (fin)

On a donc montré que

$$\langle J'(u), w\rangle \geq 0 \quad \forall\, w \in K(u) = \left\{ w \in V, \langle F_i'(u), w\rangle \leq 0, i \in I(u)\right\}$$

Grâce au lemme de Farkas, on en déduit que

$$J'(u) = -\sum_{i \in I(u)} \lambda_i F_i'(u)$$

avec $\lambda_i \geq 0$ et la somme est réduite aux contraintes actives.

# Lemme de Farkas

**Lemme.** (version simplifiée) Soient $a_1, \ldots, a_M$ une famille libre de $V$. On considère les ensembles

$$\mathcal{K} = \left\{ w \in V\,, \quad \langle a_i, w\rangle \leq 0 \;\text{ pour }\; 1 \leq i \leq M \right\},$$

et

$$\hat{\mathcal{K}} = \left\{ q \in V\,, \quad \exists\, \lambda_1, \ldots, \lambda_M \geq 0\,,\; q = -\sum_{i=1}^{M} \lambda_i a_i \right\}.$$

Alors pour tout $p \in V$, on a l'implication

$$\langle p, w\rangle \geq 0 \;\forall\, w \in \mathcal{K} \implies p \in \hat{\mathcal{K}}\,.$$

**Remarque.** La réciproque étant évidente, il s'agit en fait d'une équivalence.

**Application.** Le théorème précédent se démontre avec le choix $p = J'(u)$ et $a_i = F_i'(u)$.

# Preuve du lemme de Farkas

On caractérise $\mathcal{K} = \{w \in V \,, \langle a_i, w \rangle \leq 0 \text{ pour } 1 \leq i \leq M\}$.
Soit $A$ le s.e.v. engendré par $a_1, \ldots, a_M$ et $A^\perp$ son orthogonal:

$\forall w \in V$ il existe un unique $w_1 \in A, w_2 \in A^\perp$ tel que $w = w_1 + w_2$.

Donc $w \in \mathcal{K}$ si et seulement si $w_1 \in \mathcal{K}$ (aucune condition sur $w_2$).
Soit la base duale $b_1, \ldots, b_M$ de $A$, définie par $\langle a_i, b_j \rangle = \delta_{ij}$

$$b_j = \sum_{k=1}^{M} c_{jk} a_k, \quad (c_{jk})_{1 \leq j,k \leq M} = D^{-1}, \quad D = (\langle a_j, a_k \rangle)_{1 \leq j,k \leq M}$$

et $D$ est inversible car $a_1, \ldots, a_M$ est libre. Comme $w_1 \in A$, on a

$$w_1 = -\sum_{i=1}^{M} \mu_i b_i \quad \text{et } w_1 \in \mathcal{K} \;\Leftrightarrow\; \mu_i \geq 0 \;\text{ pour tout } 1 \leq i \leq M.$$

Autrement dit,

$$\mathcal{K} = \left\{ w \in V \text{ tel que } w_1 = -\sum_{i=1}^{M} \mu_i b_i \text{ et } \mu \in (\mathbb{R}_+)^M \right\}.$$

# Preuve du lemme de Farkas (fin)

$$\mathcal{K} = \left\{ w \in V \text{ tel que } w_1 = -\sum_{i=1}^{M} \mu_i b_i \text{ et } \mu \in (\mathbb{R}_+)^M \right\} .$$

Si $\langle p, w \rangle \geq 0$ pour tout $w \in \mathcal{K}$, on prend $w_1 = 0$ et successivement $w_2$ et $-w_2 \in A^\perp$, ce qui prouve que $p \in A$. Donc, il existe des réels $\lambda_1, \ldots, \lambda_M$ tels que

$$p = -\sum_{i=1}^{M} \lambda_i a_i .$$

Enfin, pour $w_2 = 0$ et $w_1 = -\sum_{i=1}^{M} \mu_i b_i$ avec $\mu \geq 0$

$$\langle p, w \rangle = \sum_{i,j=1}^{M} \lambda_i \mu_j \langle a_i, b_j \rangle = \mu \cdot \lambda \geq 0 \text{ pour tout } \mu \in (\mathbb{R}_+)^M,$$

ce qui n'est possible que si $\lambda \geq 0$, c'est-à-dire que $p \in \hat{\mathcal{K}}$.

# Autre condition de qualification des contraintes

Dans le polycopié on propose une condition de qualification, plus générale mais plus compliquée:
les contraintes sont qualifiées s'il existe $\overline{w} \in V$ tel que $\forall i \in I(u)$

$$(C) \qquad \begin{array}{ll} \text{ou bien} & \langle F_i'(u), \overline{w} \rangle < 0 \ , \\ \text{ou bien} & \langle F_i'(u), \overline{w} \rangle = 0 \quad \text{et} \quad F_i \quad \text{est affine} \ . \end{array}$$

**Exercice.** Si la famille $(F_i'(u))_{i \in I(u)}$ est libre alors $(C)$ est vérifiée.

La condition d'optimalité reste vraie avec $(C)$ mais la démonstration du lemme de Farkas est plus compliquée (voir polycopié).

**Intérêt.** Si toutes les contraintes $F_i$ sont affines, il n'y a rien à vérifier ! Prendre $\overline{w} = 0$.

**Définition.** On appelle Lagrangien du problème la fonction

$$\mathcal{L}(v,\mu) = J(v) + \sum_{i=1}^{M} \mu_i F_i(v) = J(v) + \mu \cdot F(v) \quad \forall (v,\mu) \in V \times (\mathbb{R}^+)^M .$$

La nouvelle variable **positive** $\mu \in (\mathbb{R}^+)^M$ est appelée multiplicateur de Lagrange pour la contrainte $F(v) \leq 0$.

**Lemme.** Le problème de minimisation est équivalent à

$$\inf_{v \in V,\ F(v) \leq 0} J(v) = \inf_{v \in V} \sup_{\mu \in (\mathbb{R}^+)^M} \mathcal{L}(v,\mu).$$

**Preuve.** On vérifie facilement que

$$\sup_{\mu \in (\mathbb{R}^+)^M} \mathcal{L}(v,\mu) = \begin{cases} J(v) & \text{si } \ F(v) \leq 0 \\ +\infty & \text{s'il existe } i \text{ tel que } \ F_i(v) > 0 \end{cases}$$

# Stationnarité du Lagrangien

Rappel de la définition du Lagrangien:

$$\mathcal{L}(v,\mu) = J(v) + \mu \cdot F(v) \qquad \forall (v,\mu) \in V \times (\mathbb{R}^+)^M .$$

On peut réécrire le précédent théorème sous la forme suivante.

**Théorème.** On suppose que les contraintes sont qualifiées en $u$ tel que $F(u) \leq 0$. Alors, si $u$ est un minimum local, il existe $\lambda_1, \ldots, \lambda_M \geq 0$, appelés multiplicateurs de Lagrange, tels que

$$\frac{\partial \mathcal{L}}{\partial v}(u,\lambda) = 0 \quad \text{et} \quad \frac{\partial \mathcal{L}}{\partial \mu}(u,\lambda) \cdot (\mu - \lambda) \leq 0 \quad \forall \mu \in (\mathbb{R}^+)^M$$

**Remarque.** La deuxième condition est l'inéquation d'Euler pour la **maximisation** du Lagrangien par rapport à $\mu$ dans le convexe fermé $(\mathbb{R}^+)^M$.

# Preuve

Les conditions d'optimalité sont

$$J'(u)+\sum_{i=1}^{M} \lambda_i F_i'(u) = 0 \ , \quad \lambda_i \geq 0 \ , \quad F_i(u) \leq 0 \ , \quad \lambda_i = 0 \text{ si } F_i(u) < 0 \ .$$

Cette condition est bien la stationnarité du Lagrangien puisque

$$\frac{\partial \mathcal{L}}{\partial v}(u,\lambda) = J'(u) + \lambda \cdot F'(u) = 0,$$

et que l'inéquation d'Euler

$$\frac{\partial \mathcal{L}}{\partial \mu}(u,\lambda) \cdot (\mu - \lambda) = F(u) \cdot (\mu - \lambda) \leq 0 \quad \forall \mu \in (\mathbb{R}^+)^M$$

implique que $\lambda_i = 0$ si $F_i(u) < 0$.

# Exemple: régularisation d'un signal

Soit $A$ une matrice carrée d'ordre $n$, symétrique définie positive. Soit $b \in \mathbb{R}^n$ et $\epsilon > 0$. On considère

$$\min_{x\in\mathbb{R}^n \text{ tel que } \|x-b\|^2\leq\epsilon^2} Ax \cdot x.$$

Il existe une unique solution $x^* \in \mathbb{R}^n$ qui vérifie

(1) $$\text{soit } x^* = 0 \quad \text{ et } \|b\| < \epsilon,$$

(2) soit $x^* = (\mathrm{Id} + \lambda^{-1}A)^{-1}b$ avec $\lambda > 0$ tel que $\|x^* - b\| = \epsilon$.

**Preuve.** Soit la contrainte n'est pas active et on trouve (1). Soit la contrainte est active et on trouve (2) avec la condition d'optimalité (la contrainte est qualifiée car $x^* - b \neq 0$ quand $\|x^* - b\| = \epsilon$).

# Contraintes d'égalité et d'inégalité

Soit $F(v)$: $V \to \mathbb{R}^M$ et $G(v)$: $V \to \mathbb{R}^N$. On considère

$$\inf_{v \in K} J(v) \quad \text{avec} \quad K = \{v \in V,\ G(v) = 0,\ F(v) \leq 0\}.$$

**Théorème.** Soit $u \in K$. On suppose que $J$ et $F$ sont dérivables en $u$, que $G$ est dérivable dans un voisinage de $u$, et les contraintes sont qualifiées en $u$:

$$\left(G_i'(u)\right)_{1 \leq i \leq N} \bigcup \left(F_i'(u)\right)_{i \in I(u)} \text{ est une famille libre}$$

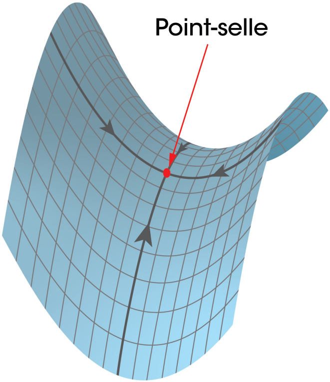
où $I(u)$ est l'ensemble des contraintes actives $F_i(u) = 0$. Alors, si $u$ est un minimum local de $J$ sur $K$, il existe des multiplicateurs de Lagrange $\mu_1, \ldots, \mu_N$, et $\lambda_1, \ldots, \lambda_M \geq 0$, tels que

$$J'(u) + \sum_{i=1}^{N} \mu_i G_i'(u) + \sum_{i=1}^{M} \lambda_i F_i'(u) = 0 \ , \ \lambda \geq 0 \ , \ F(u) \leq 0 \ , \ \lambda \cdot F(u) = 0 \ .$$

# Point selle

**Définition.** Soit un Lagrangien $\mathcal{L}(v,q)$. On dit que $(u,p) \in V \times P$ est un point-selle (ou col, ou min-max) de $\mathcal{L}$ sur $V \times P$ si

$$\forall\, q \in P \qquad \mathcal{L}(u,q) \leq \mathcal{L}(u,p) \leq \mathcal{L}(v,p) \quad \forall\, v \in V\ .$$




On minimise en $v$ et on maximise en $q$.

# Point selle (2)

**Proposition.** Soit des fonctions $J, F_1, \ldots, F_M$ dérivables sur $V$.
Ou bien $K = \{F(v) = 0\}$ et $P = \mathbb{R}^M$.
Ou bien $K = \{F(v) \leq 0\}$ et $P = (\mathbb{R}_+)^M$.
Pour $(v, q) \in V \times P$, on considère le Lagrangien

$$\mathcal{L}(v, q) = J(v) + q \cdot F(v).$$

Si $(u, p)$ est un point-selle de $\mathcal{L}$ sur $V \times P$, alors $u \in K$ et $u$ est un minimum global de $J$ sur $K$. De plus, on a

$$J'(u) + \sum_{i=1}^{M} p_i F_i'(u) = 0 \, .$$

# Preuve

Dans le cas de contraintes d'égalité, la condition de point-selle est

$$\forall\, q \in \mathbb{R}^M \quad J(u)+q\cdot F(u) \leq J(u)+p\cdot F(u) \leq J(v)+p\cdot F(v) \quad \forall\, v \in V.$$

la première inégalité montre que $F(u) = 0$, i.e. $u \in K$. Il reste alors

(1) $$J(u) \leq J(v) + p \cdot F(v) \quad \forall\, v \in V,$$

qui montre, pour $v \in K$, que $u$ est un minimum global de $J$ sur $K$.

Si $J$ et $F$ sont dérivables en $u$, (1) montre que $u$ est un point de minimum sans contrainte de $J + p \cdot F$ dans $V$, ce qui implique que la dérivée s'annule en $u$, $J'(u) + p \cdot F'(u) = 0$.

Même principe de preuve pour le cas de contraintes d'inégalité.

**Théorème de Kuhn et Tucker.** Soit des fonctions $J, F_1, \dots, F_M$ convexes et dérivables sur $V$. Soit $K = \{F(v) \leq 0\}$, $P = (\mathbb{R}_+)^M$ et le Lagrangien

$$\mathcal{L}(v, q) = J(v) + q \cdot F(v) \quad \forall (v, q) \in V \times P \; .$$

Soit $u \in K$ où les contraintes sont qualifiées. Alors les 3 affirmations suivantes sont équivalentes:

1. $u$ est un minimum global de $J$ sur $K$,
2. $\exists\, p \in (\mathbb{R}_+)^M$ tel que $(u, p)$ est un point-selle du Lagrangien $\mathcal{L}$ sur $V \times P$,
3. $\exists\, p \in (\mathbb{R}_+)^M$ tel que

$$F(u) \leq 0 \,, \quad p \geq 0 \,, \quad p \cdot F(u) = 0 \,, \quad J'(u) + p \cdot F'(u) = 0 \, .$$

**Remarques.**
Ce théorème ne s'applique que pour des contraintes d'inégalité.
Les conditions d'optimalité sont donc nécessaires et suffisantes dans ce cas.

# Preuve

$(\mathbf{1}) \Rightarrow (\mathbf{3})$ Si $u$ est un minimum de $J$ sur $K$, on peut écrire la condition d'optimalité: $\exists\, p \in (\mathbb{R}_+)^M$ tel que

$$F(u) \leq 0\,, \quad p \geq 0\,, \quad p \cdot F(u) = 0\,, \quad J'(u) + p \cdot F'(u) = 0\,.$$

$(\mathbf{3}) \Rightarrow (\mathbf{2})$ On en déduit $J(u) + q \cdot F(u) \leq J(u) + p \cdot F(u) = J(u)$ pour tout $q \in (\mathbb{R}_+)^M$. De plus, comme $J(v) + p \cdot F(v)$ est convexe, la condition d'optimalité est suffisante et on en déduit

$$J(u) + p \cdot F(u) \leq J(v) + p \cdot F(v) \quad \forall v \in V,$$

c'est-à-dire que $(u, p)$ est point-selle de $\mathcal{L}(v, q)$ sur $V \times (\mathbb{R}_+)^M$.

$(\mathbf{2}) \Rightarrow (\mathbf{1})$ Si $(u, p)$ est point-selle, on a déjà montré dans la Proposition précédente que $u$ est un minimum global de $J$ sur $K$.

# Exemple essentiel

Soit $A$ une matrice carrée d'ordre $n$, symétrique définie positive.
Soit $B$ une matrice de taille $m \times n$ avec $m \leq n$ et $rg(B) = m$.
Soit $b \in \mathbb{R}^n$ et $c \in \mathbb{R}^m$. On considère

$$\min_{x \in \mathbb{R}^n \text{ tel que } Bx \leq c} \left\{ J(x) = \frac{1}{2} Ax \cdot x - b \cdot x \right\}.$$

Il existe une unique solution $x^* \in \mathbb{R}^n$ qui vérifie, pour $p \in (\mathbb{R}^+)^m$,

$$Ax^* - b + B^* p = 0\,, \quad p \cdot (Bx^* - c) = 0\,, \quad p \geq 0\,, \quad Bx^* \leq c.$$

Comment calculer $p$ et $x^*$ ?

- La condition d'optimalité semble non-linéaire...
- **En fait c'est un problème combinatoire !**
- Problème très important en pratique car à la base de l'algorithme SQP.

**Stratégie de résolution:**

On prédit l'ensemble $I$ des contraintes actives, $\{1, ..., m\} = I \cup I^c$.

On résout

$$\min_{x \in \mathbb{R}^n \text{ tel que } (Bx)_i = c_i,\ \forall i \in I,} J(x)$$

qui admet une unique solution $x_I$ et un unique multiplicateur de Lagrange $p_I$.

On teste:

1. si $p_I \geq 0$ et si $(Bx)_i \leq c_i$ pour $i \in I^c$, on a trouvé la bonne solution grâce au théorème de Kuhn-Tucker,
2. sinon, on recommence avec un nouveau choix de $I$.

Il y a un nombre fini (mais exponentiel) de choix possibles de $I$ !
Il existe un algorithme qui fait mieux qu'une simple énumération.

# 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

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

**Exemple.** Lagrangien $\mathcal{L}(v,q) = J(v) + q \cdot F(v)$ avec $P = \mathbb{R}^M$ (ou $\mathbb{R}^M_+$). Dans ce cas $\mathcal{J}(v) = J(v)$ si $F(v) = 0$ (ou $F(v) \leq 0$) et $\mathcal{J}(v) = +\infty$ sinon, tandis qu'il n'y a pas de contraintes pour le problème dual (autres que $q \in P$).

# Dualité (2)

**Lemme (dualité faible).** On a toujours

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

**Preuve:** $\mathcal{L}(v,q) \geq \inf_v \mathcal{L}(v,q) \Rightarrow \sup_q \mathcal{L}(v,q) \geq \sup_q \inf_v \mathcal{L}(v,q)$
Donc

$$\inf_v \sup_q \mathcal{L}(v,q) \geq \sup_q \inf_v \mathcal{L}(v,q).$$

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

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

Notion très importante aussi bien du point de vue théorique que numérique !

- Le problème dual peut avoir une interprétation "physique".
- Le problème dual n'a pas de contraintes.
- Le problème dual peut être plus simple à résoudre que le primal.
- S'il existe un point selle, une fois connue la solution $p$ du dual, on obtient la solution $u$ du primal grâce à une minimisation sans contrainte de $v \to \mathcal{L}(v, p)$.
- Des algorithmes numériques utilisent cette dualité.

# Preuve (dualité forte)

Soit $(u, p)$ un point-selle de $\mathcal{L}$

$$\forall\, q \in P \quad \mathcal{L}(u, q) \le \mathcal{L}(u, p) = \mathcal{L}^* \le \mathcal{L}(v, p) \quad \forall\, v \in V.$$

On rappelle que $\mathcal{J}(v) = \sup_{q\in P} \mathcal{L}(v, q)$.
Donc l'inégalité de gauche donne $\mathcal{J}(u) = \mathcal{L}^*$
et l'inégalité de droite $\mathcal{L}^* \le \mathcal{L}(v, p) \le \mathcal{J}(v)$
Par conséquent $\mathcal{J}(u) = \min_{v\in U} \mathcal{J}(v) = \mathcal{L}^*$.
On montre de la même façon que $\mathcal{G}(p) = \sup_{q\in P} \mathcal{G}(q) = \mathcal{L}^*$.
**Réciproquement**, supposons que

$$\mathcal{J}(u) = \min_{v\in V} \mathcal{J}(v) = \mathcal{L}^* = \max_{q\in P} \mathcal{G}(q) = \mathcal{G}(p)\ .$$

Par définition de $\mathcal{J}$ on a

$$\mathcal{L}(u, q) \le \mathcal{J}(u) = \mathcal{L}^* \quad \forall\, q \in P\ .$$

Par définition de $\mathcal{G}$ on a

$$\mathcal{L}(v, p) \ge \mathcal{G}(p) = \mathcal{L}^* \quad \forall\, v \in V\ ,$$

ce qui montre que $(u, p)$ est point-selle et $\mathcal{L}(u, p) = \mathcal{L}^*$.

# Exemple

$A$ sym. déf. $\geq 0$, $B$ $m \times n$ avec $rg(B) = m \leq n$, $b \in \mathbb{R}^n$ et $c \in \mathbb{R}^m$.

$$\min_{x \in \mathbb{R}^n \text{ tel que } Bx = c} \left\{ J(x) = \frac{1}{2} Ax \cdot x - b \cdot x \right\}.$$

Fonction duale, pour $p \in \mathbb{R}^m$,

$$\mathcal{G}(p) = \min_{x \in \mathbb{R}^n} \left\{ \mathcal{L}(x,p) = \frac{1}{2} Ax \cdot x - b \cdot x + p \cdot (Bx - c) \right\}.$$

On calcule $\quad \mathcal{G}(p) = -\frac{1}{2} A^{-1}(b - B^* p) \cdot (b - B^* p) - p \cdot c.$

Problème dual (sans contraintes !)

$$\max_{p \in \mathbb{R}^m} \mathcal{G}(p)$$

Une fois calculé le point de maximum $p^*$, il reste une minimisation (sans contraintes !) pour trouver le point de minimum $x^*$

$$\min_{x \in \mathbb{R}^n} \left\{ \mathcal{L}(x,p^*) = \frac{1}{2} Ax \cdot x - b \cdot x + p^* \cdot (Bx - c) \right\}.$$

# III - Algorithmes d'optimisation sans contraintes.

- Tous les algorithmes d'optimisation sont itératifs.
- On se limite aux algorithmes déterministes utilisant la connaissance des dérivées.
- Un algorithme est dit d'ordre 1 s'il utilise le gradient.
- Un algorithme est dit d'ordre 2 s'il utilise le gradient et la dérivée seconde.

1. Gradient à pas optimal (ordre 1).
2. Gradient à pas fixe (ordre 1).

Dans tout ce qui suit, $V = \mathbb{R}^N$ est un espace vectoriel de dimension finie, muni d'un produit scalaire $\langle u, v \rangle$ et d'une norme associée.

# Principe des algorithmes de gradient

On veut résoudre

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

Initialisation: choisir $u^0$ dans $V$.
Itérations: pour $n \geq 0$

$$u^{n+1} = u^n - \mu w^n \quad \text{avec } \mu > 0$$

avec $w^n$ = direction de descente et $\mu$ = pas de descente. Si $\mu$ est petit, on a

$$J(u^{n+1}) = J(u^n) - \mu \langle J'(u^n), w^n \rangle + \mathcal{O}(\mu^2),$$

donc pour descendre il vaut mieux prendre $w^n$ colinéaire à $J'(u^n)$.
**Conclusion:** algorithme de gradient

$$u^{n+1} = u^n - \mu^n J'(u^n) \quad \text{avec } \mu^n > 0.$$

# Gradient à pas optimal

On veut résoudre

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

Initialisation: choisir $u^0$ dans $V$.
Itérations: pour $n \geq 0$

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

où $\mu^n \in \mathbb{R}$ est choisi à chaque étape tel que

$$J(u^{n+1}) = \inf_{\mu\in\mathbb{R}^+} J\big(u^n - \mu J'(u^n)\big) .$$

$\mu^n$ est appelé pas de descente optimal.

# Convergence de l'algorithme

**Théorème** Soit $J$ une fonction différentiable. On suppose que $J$ est $\alpha$-convexe, 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 tout borné de $V$: $\forall\, M > 0$ il existe $C_M > 0$ tel que

$$\|v\| + \|w\| \leq M \;\Rightarrow\; \big\|J'(v) - J'(w)\big\| \leq C_M \|v - w\| \,.$$

Alors l'algorithme de gradient à pas optimal converge: $\forall\, u^0$, la suite $(u^n)$ converge vers la solution $u$.

**Remarque.** Si $J$ n'est pas $\alpha$-convexe:

- l'algorithme peut converger vers un minimum local,
- le résultat de l'algorithme peut varier selon l'initialisation,
- l'algorithme peut ne pas converger (il oscille entre plusieurs limites bien que la fonction décroisse).

# Preuve

A chaque itération $n$ on minimise $f(\mu) = J\big(u^n - \mu\, J'(u^n)\big)$ et $\mu^n$ est caractérisé par $f'(\mu^n) = 0$, ce qui s'écrit aussi

$$\langle J'(u^{n+1}),\, J'(u^n)\rangle = 0\,. \tag{1}$$

Autrement dit $\langle J'(u^{n+1}), u^{n+1} - u^n\rangle = 0$ et on déduit de l'$\alpha$-convexité de $J$ que

$$J(u^n) - J(u^{n+1}) \geq \frac{\alpha}{2}\|u^n - u^{n+1}\|^2\,,$$

ce qui prouve que la suite $J(u^n)$ est décroissante, minorée par $J(u)$, donc elle converge et $u^{n+1} - u^n$ tend vers 0.
Comme $J$ est $\alpha$-convexe et $J(u^n)$ est bornée, la suite $(u^n)$ est aussi bornée par une constante $M$. Alors

$$\|J'(u^n) - J'(u^{n+1})\|^2 \leq C_M^2\|u^{n+1} - u^n\|^2$$

et, à cause de (1),

$$\|J'(u^n) - J'(u^{n+1})\|^2 = \|J'(u^n)\|^2 + \|J'(u^{n+1})\|^2 \geq \|J'(u^n)\|^2$$

# Preuve (fin)

On en déduit

$$\|J'(u^n)\| \leq C_M \|u^{n+1} - u^n\|,$$

donc $J'(u^n)$ tend vers 0. L'$\alpha$-convexité de $J$ donne alors

$$\alpha\|u^n - u\|^2 \leq \langle J'(u^n) - J'(u), u^n - u\rangle = \langle J'(u^n), u^n - u\rangle$$
$$\leq \left\|J'(u^n)\right\| \ \|u^n - u\|$$

ce qui implique

$$\alpha\|u^n - u\| \leq \|J'(u^n)\|,$$

d'où l'on déduit la convergence de l'algorithme.

# Variante: gradient à pas fixe

On veut résoudre

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

Initialisation: choisir $u^0$ dans $V$. Itérations: pour $n \geq 0$

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

**Théorème.** On suppose que $J$ est $\alpha$-convexe, différentiable et que $J'$ est Lipschitzien sur $V$: il existe $L > 0$ tel que

$$\big\| J'(v) - J'(w) \big\| \leq L \|v - w\| \quad \forall v, w \in V.$$

Alors, si $\mu > 0$ est suffisamment petit, l'algorithme de gradient à pas fixe converge : $\forall \, u^0$, la suite $(u^n)$ converge vers la solution $u$.

**Remarque.** Si $J$ est $C^2$, alors l'$\alpha$-convexité est équivalente à $J''(u)(w,w) \geq \alpha \|w\|^2$ pour tout $u, w \in V$, tandis que l'hypothèse sur $J'$ Lipschitz est vérifiée si $J''(u)(w,w) \leq L \|w\|^2 \;\; \forall u, w \in V$.

# Preuve

Posons $v^n = u^n - u$. Comme $J'(u) = 0$, on a

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

donc

$$\begin{aligned} \|v^{n+1}\|^2 &= \|v^n\|^2 - 2\mu\langle J'(u^n) - J'(u), u^n - u\rangle + \mu^2\big\|J'(u^n) - J'(u)\big\|^2 \\ &\leq \left(1 - 2\alpha\mu + L^2\mu^2\right) \|v^n\|^2 , \end{aligned}$$

d'après l'$\alpha$-convexité et $J'$ $L$-Lipschitz. Si $0 < \mu < 2\alpha/L^2$, on a $0 < 1 - 2\alpha\mu + L^2\mu^2 < 1$, d'où la convergence.

**Remarque.** (vitesse de convergence géométrique) On en déduit

$$\|u^n - u\| \leq \gamma^n\|u^0 - u\| \quad \text{avec } \gamma = \sqrt{1 - \frac{\alpha^2}{L^2}} \text{ pour } \mu = \frac{\alpha}{L^2}.$$

D'autres algorithmes d'ordre 1 améliorent la valeur de $\gamma$ mais pas la puissance...

On vient d'étudier les premiers algorithmes "de base".

La semaine prochaine on étudiera:

- la méthode de Newton (algorithme d'ordre 2),
- des variantes adaptées à des problèmes spécifiques (algorithmes d'ordre 1),
- les vitesses de convergence (optimales ou pas),