Conditions d'optimalité II

(P. Carpentier)

6 mars 2020

Plan du cours

Problème

Notations :

- $\mathbb{U}$: espace de Hilbert de dimension finie ($\mathbb{R}^n$),
- U^{ad} : sous-ensemble convexe fermé non vide de $\mathbb{U}$,
- $\mathbb{V}$: Hilbert dans lequel vivent les contraintes ($\mathbb{R}^m$),
- C : cône convexe fermé non vide de $\mathbb{V}$,

- $J \ : \ \mathbb{U} \longrightarrow \mathbb{R}$: continue et différentiable (critère),
- $\Theta \ : \ \mathbb{U} \longrightarrow \mathbb{V}$: continue et différentiable (contrainte).

Problème $(\mathcal{P})$:

$$\min_{u \in U^{\mathrm{ad}}} J(u),$$

$$\text{sous} \quad : \quad \Theta(u) \in -C.$$

Ensemble admissible regroupant toutes les contraintes :

$$U^{\mathrm{ad}}_{\Theta} = \{u \in U^{\mathrm{ad}} \mid \Theta(u) \in -C\}.$$

Quelques rappels du cours précédent

Soit $u_0 \in U^{\mathrm{ad}}_{\Theta} \subset \mathbb{U}$

- Cône tangent à U^{ad}_{Θ} en u_0

$$T_{u_0} U^{\mathrm{ad}}_{\Theta} = \big\{ d \in \mathbb{U} \mid \exists d_k \to d,\ \exists \epsilon_k \downarrow 0 \ \text{ t.q. } \ u_0 + \epsilon_k d_k \in U^{\mathrm{ad}}_{\Theta} \big\}.$$

- Cône dual positif

$$\big(T_{u_0} U^{\mathrm{ad}}_{\Theta} \big)^+ = \big\{ d \in \mathbb{U} \mid \langle d \,, d' \rangle \geq 0 \ \forall d' \in T_{u_0}\big(U^{\mathrm{ad}}_{\Theta}\big) \big\}.$$

Cas U^{ad}_{Θ} **convexe** : $\big(T_{u_0} U^{\mathrm{ad}}_{\Theta} \big)^+ = \big\{ d \in \mathbb{U} \mid \langle d \,, u - u_0 \rangle \geq 0 \ \ \forall u \in U^{\mathrm{ad}}_{\Theta} \big\}$.

- Condition nécessaire d'optimalité

$$u^\sharp \in \arg\min_{u \in U^{\mathrm{ad}}_{\Theta}} J(u) \implies \nabla J(u^\sharp) \in \Big(T_{u^\sharp} U^{\mathrm{ad}}_{\Theta} \Big)^+ .$$

- Lemme de Farkas (généralisé)

$$\big\{ d \in \mathbb{U} \mid A.d \in K \big\}^+ = \operatorname{adh}\big(A^{\star}.q \mid q \in K^+ \big).$$

Condition d'optimalité avec hypothèse de qualification

Difficulté : exprimer le dual positif du cône tangent...

- Définition du cône linéarisant à U_Θ^{ad} en u_0 :

$$T_{u_0}^{\ell} U_\Theta^{\mathrm{ad}} = \left\{ d \in \mathbb{U} \mid d \in T_{u0} U^{\mathrm{ad}} \ \textbf{et} \ \Theta'(u_0).d \in T_{\Theta(u_0)}(-C) \right\}.$$

- L'inclusion suivante est toujours vraie :

$$T_{u_0} U_\Theta^{\mathrm{ad}} \subset T_{u_0}^{\ell} U_\Theta^{\mathrm{ad}}.$$

- **Hypothèse de qualification des contraintes** en u_0 **(QC)** :

$$T_{u_0} U_\Theta^{\mathrm{ad}} = T_{u_0}^{\ell} U_\Theta^{\mathrm{ad}}.$$

- Sous **(QC)**, la condition nécessaire d'optimalité devient :

$$u^\sharp \in \operatorname*{arg\,min}_{u \in U_\Theta^{\mathrm{ad}}} J(u) \implies \nabla J(u^\sharp) \in \left(T_{u^\sharp}^{\ell} U_\Theta^{\mathrm{ad}} \right)^+.$$

$$T^{\ell}_{u^\sharp} U^{\mathrm{ad}}_{\Theta} = \Big\{ d \in \mathbb{U} \mid \underbrace{\big(d, \Theta'(u^\sharp).d\big)}_{A.d} \in \underbrace{T_{u^\sharp} U^{\mathrm{ad}} \times T_{\Theta(u^\sharp)}(-C)}_{K} \Big\}.$$

Lemme de Farkas : $\big\{ d \in \mathbb{U} \mid A.d \in K \big\}^+ = \mathrm{adh}\big(A^\star.q \mid q \in K^+ \big)$

- $A^\star.(w, \lambda) = w + (\Theta'(u^\sharp))^\star.\lambda$,
- $K^+ = (T_{u^\sharp} U^{\mathrm{ad}})^+ \times \big(T_{\Theta(u^\sharp)}(-C) \big)^+$.

La condition d'optimalité s'écrit donc :

$$\nabla J(u^\sharp) \in \Big\{ w + (\Theta'(u^\sharp))^\star.\lambda \mid w \in (T_{u^\sharp} U^{\mathrm{ad}})^+ ,\ \lambda \in \big(T_{\Theta(u^\sharp)}(-C) \big)^+ \Big\}$$

- $\exists \lambda^\sharp,\ \forall u \in U^{\mathrm{ad}},\ \Big\langle \nabla J(u^\sharp) - (\Theta'(u^\sharp))^\star.\lambda^\sharp \,,\, u - u^\sharp \Big\rangle \geq 0$,

  car $(T_{u^\sharp} U^{\mathrm{ad}})^+ = \big\{ w \in \mathbb{U} \mid \langle w \,,\, u - u^\sharp \rangle \geq 0 \ \ \forall u \in U^{\mathrm{ad}} \big\}$.

- $\lambda^\sharp \in \big(T_{\Theta(u^\sharp)}(-C) \big)^+ \iff \langle \lambda^\sharp \,,\, \Theta(u^\sharp) \rangle = 0$ et $-\lambda^\sharp \in C^+$,

  car $\big(T_{\Theta(u^\sharp)}(-C) \big)^+ = \big\{ \lambda \in \mathbb{V} \mid \langle \lambda \,,\, v + \Theta(u^\sharp) \rangle \leq 0 \ \ \forall v \in C \big\}$.

Conditions nécessaires d'optimalité de **KKT**

Soit $u^\sharp$ un minimum local du problème $(\mathcal{P})$. On suppose, en plus des hypothèses initiales sur le problème, que les contraintes sont qualifiées en $u^\sharp$. Alors, il existe $\lambda^\sharp \in \mathbb{V}$ tel que l'on ait :

1. $\left\langle \nabla J(u^\sharp) + \left(\Theta'(u^\sharp)\right)^\star . \lambda^\sharp \, , u - u^\sharp \right\rangle \geq 0 \quad \forall u \in U^{\mathrm{ad}}$,
2. $\Theta(u^\sharp) \in -C$,
3. $\lambda^\sharp \in C^+$,
4. $\left\langle \lambda^\sharp \, , \Theta(u^\sharp) \right\rangle = 0$.

La condition 4 est appelée condition des écarts complémentaires, et $\lambda^\sharp$ est appelé multiplicateur optimal associé aux contraintes Θ.

Remarque : les conditions de **KKT** sont des conditions nécessaires et suffisantes d'optimalité lorsque $(\mathcal{P})$ est un problème **convexe**, pourvu que l'on vérifie l'hypothèse de qualification des contraintes.

On s'intéresse d'abord à la première condition de KKT :

$$\left\langle \nabla J(u^\sharp) + \left(\Theta'(u^\sharp)\right)^\star . \lambda^\sharp \, , u - u^\sharp \right\rangle \geq 0 \quad \forall u \in U^{\mathrm{ad}}.$$

Cas sans aucune contrainte : $U_\Theta^{\mathrm{ad}} = \mathbb{U}$ (Θ non présent).

L'équation d'optimalité se réduit à l'expression :

$$\nabla J(u^\sharp) = 0.$$

On retrouve la condition « classique » de stationnarité...

Cas sans ensemble admissible : $U^{\mathrm{ad}} = \mathbb{U}$.

L'inéquation variationnelle ci-dessus devient une **équation** :

$$\nabla J(u^\sharp) + \left(\Theta'(u^\sharp)\right)^\star . \lambda^\sharp = 0.$$

On s'intéresse ensuite aux trois dernières conditions de KKT :

$$\Theta(u^\sharp) \in -C \quad , \quad \lambda^\sharp \in C^+ \quad , \quad \left\langle \lambda^\sharp , \Theta(u^\sharp) \right\rangle = 0.$$

- Contraintes de type égalité : $C = \{0\} \quad \rightsquigarrow \quad C^+ = \mathbb{R}^m$.
  Les trois dernières conditions de **KKT** se résument à :

$$\Theta(u^\sharp) = 0.$$

- Contraintes de type inégalité : $C = \mathbb{R}^m_+ \quad \rightsquigarrow \quad C^+ = \mathbb{R}^m_+$.
  Les trois dernières conditions de **KKT** prennent la forme :

$$\Theta_i(u^\sharp) \leq 0 \text{ et } \lambda_i^\sharp \geq 0 \ \forall i = 1 \ldots m \quad , \quad \sum_{i=1}^m \lambda_i^\sharp \Theta_i(u^\sharp) = 0.$$

$\rightsquigarrow$ Contraintes disjonctives : $\lambda_i^\sharp = 0 \quad$ **ou** $\quad \Theta_i(u^\sharp) = 0 \ \ \forall i$

Pour les **résoudre**, on considère les 2^m alternatives possibles...

Interprétation lagrangienne des conditions de **KKT**

On définit le Lagrangien du problème d'optimisation

$$\begin{array}{rcl} L \; : \; U^{\mathrm{ad}} \times C^{+} & \longrightarrow & \mathbb{R} \\ (u, \lambda) & \longrightarrow & J(u) + \langle \lambda \,, \Theta(u) \rangle . \end{array}$$

Dans le cas $U^{\mathrm{ad}} = \mathbb{U}$ et $C = \{0\}$ (contraintes égalité uniquement), les conditions de **KKT** s'expriment à l'aide des gradients de L :

$$\left. \begin{array}{l} \nabla J(u^\sharp) + \big(\Theta'(u^\sharp)\big)^{\star} . \lambda^\sharp = 0 \\ \Theta(u^\sharp) = 0 \end{array} \right\} \iff \left\{ \begin{array}{l} \nabla_u L(u^\sharp, \lambda^\sharp) = 0 \\ \nabla_\lambda L(u^\sharp, \lambda^\sharp) = 0 \end{array} \right. .$$

La recherche d'un point vérifiant les conditions de **KKT** se ramène alors à l'étude de la stationnarité du lagrangien ! La solution de ce système n'est cependant pas toujours solution du problème $(\mathcal{P})$.

Exemple : utilisation de KKT dans le cas linéaire quadratique

$$\min_{B.u - b = 0} \quad \frac{1}{2} u^\top . A . u \quad \rightsquigarrow \quad \begin{pmatrix} A & B^\top \\ B & 0 \end{pmatrix} . \begin{pmatrix} u^\sharp \\ \lambda^\sharp \end{pmatrix} = \begin{pmatrix} 0 \\ b \end{pmatrix} .$$

Exemple : projection sur un demi-espace

Soit $u_0 \in \mathbb{R}^n$ et soit l'hyperplan de $\mathbb{R}^n$ d'équation : $a^\top.u - b = 0$.

$$\min_{u \in \mathbb{R}^n} \frac{1}{2}\|u - u_0\|^2 \quad \text{sous} \quad a^\top.u - b \leq 0.$$

On écrit les **conditions de KKT** :

$$u^\sharp - u_0 + \lambda^\sharp a = 0,$$
$$a^\top u^\sharp - b \leq 0 \quad , \quad \lambda^\sharp \geq 0 \quad , \quad \lambda^\sharp\left(a^\top u^\sharp - b\right) = 0,$$

et on joue aux **devinettes** :

- **Si** $a^\top u^\sharp - b < 0$, on a $\lambda^\sharp = 0$ (écarts complémentaires).
  $\rightsquigarrow$ $u^\sharp = u_0$ pourvu que les données vérifient : $a^\top u_0 - b < 0$.
- **Sinon**, on résout le système :

$$\left\{\begin{array}{l} a^\top u^\sharp - b = 0 \\ u^\sharp - u_0 + \lambda^\sharp a = 0 \end{array}\right. \quad \rightsquigarrow \quad \left\{\begin{array}{l} \lambda^\sharp = \frac{1}{\|a\|^2}\left(a^\top u_0 - b\right) \\ u^\sharp = u_0 - \frac{1}{\|a\|^2}\left(a^\top u_0 - b\right)a \end{array}\right.$$

pourvu que l'on ait $\lambda^\sharp \geq 0$, ce qui implique : $a^\top u_0 - b \geq 0$.

Rappel des conditions nécessaires d'optimalité de KKT

Soit $u^\sharp$ un minimum local du problème $(\mathcal{P})$:

$$\min_{u \in U^{\mathrm{ad}}} J(u) \quad \text{sous} \quad \Theta(u) \in -C.$$

Sous l'hypothèse de **qualification des contraintes (QC)** en $u^\sharp$:

$$T_{u^\sharp} U^{\mathrm{ad}}_{\Theta} = T^{\ell}_{u^\sharp} U^{\mathrm{ad}}_{\Theta},$$

il existe $\lambda^\sharp \in \mathbb{V}$ tel que l'on ait :

- $\left\langle \nabla J(u^\sharp) + \left(\Theta'(u^\sharp)\right)^{\star}.\lambda^\sharp \,, u - u^\sharp \right\rangle \geq 0 \quad \forall u \in U^{\mathrm{ad}}$,
- $\Theta(u^\sharp) \in -C$,
- $\lambda^\sharp \in C^+$,
- $\left\langle \lambda^\sharp \,, \Theta(u^\sharp) \right\rangle = 0$.

Comment vérifier la condition QC : $T_{u^\sharp} U^{\mathrm{ad}}_{\Theta} = T^{\ell}_{u^\sharp} U^{\mathrm{ad}}_{\Theta}$?
(sans elle, les conditions de **KKT** *ne sont pas valides).*

Idée : donner des conditions **suffisantes** qui assurent l'égalité.

Cas général : condition de Robinson

$$0 \in \mathrm{int}\Big(\Theta(u^\sharp) + \Theta'(u^\sharp).\big(U^{\mathrm{ad}} - u^\sharp\big) + C\Big).$$

Cas convexe : condition de Slater

$$0 \in \mathrm{int}\Big(\Theta\big(U^{\mathrm{ad}}\big) + C\Big).$$

- $C = \{0\}$: contrainte égalité (convexité $\Rightarrow$ Θ affine)

$$0 \in \mathrm{int}\Big(\Theta\big(U^{\mathrm{ad}}\big)\Big).$$

- $\mathrm{int}(C) \neq \emptyset$: contrainte de cône (convexité $\Rightarrow$ Θ $C-$convexe)

$$\exists \bar{u} \in U^{\mathrm{ad}},\ \Theta(\bar{u}) \in \mathrm{int}\big(-C\big).$$

Interprétation marginaliste des multiplicateurs

Problème perturbé et fonction valeur :

$$G(v) = \left\{ \min_{u \in U^{\mathrm{ad}}} J(u) \quad \text{sous} \quad \Theta(u) - v \in -C \right\}.$$

G fournit le coût optimal en fonction du niveau de perturbation v.

Soit $(u^\sharp, \lambda^\sharp)$ une solution primale-duale de **KKT**.

Il est clair que l'on a toujours :

$$J(u^\sharp) = G(0),$$

Dans le cas convexe, G est convexe et sous-différentiable, et l'on dispose alors de l'interprétation marginaliste du multiplicateur $\lambda^\sharp$:

$$-\lambda^\sharp \in \partial G(0).$$

Un des résultats les plus utilisés de la théorie de l'optimisation !