Module B5
Théorème de Karush–Kuhn–Tucker

Dans ce module, E désigne l'espace euclidien $\mathbb{R}^n$ avec $n \in \mathbb{N}$, muni du produit scalaire usuel et de norme associée la norme euclidienne, notée $\|\cdot\|_2$. Pour plus de détails, le lecteur pourra se référer au module **A2 : Différentiabilité sur les espaces euclidiens**.

L'objectif de ce module est d'établir des conditions d'optimalité du premier ordre pour des problèmes d'optimisation sous contraintes dont l'ensemble admissible est **fermé** et décrit à l'aide de contraintes d'égalité et d'inégalité.

1 Théorème de Karush–Kuhn–Tucker

1.1 Position du problème

Soit $U \subset E$ un ouvert. Soit $f : U \to \mathbb{R}$, $g_i : U \to \mathbb{R}$ avec $i \in [\![1\,;\,p]\!]$ et $h_j : U \to \mathbb{R}$ avec $j \in [\![1\,;\,q]\!]$ des fonctions $\mathcal{C}^1$. On considère le problème d'optimisation sous contraintes suivant

$$\text{Minimiser } f(x) \text{ sous les contraintes } \begin{cases} g_i(x) = 0 & \text{pour } i \in [\![1\,;\,p]\!] \\ h_j(x) \leq 0 & \text{pour } j \in [\![1\,;\,q]\!] \end{cases} \qquad (\mathcal{P}_c)$$

L'ensemble admissible $\mathcal{A}$ de ce problème est décrit à l'aide de contraintes d'égalité et d'inégalité sous forme normale. Il peut s'écrire comme l'intersection finie suivante :

$$\mathcal{A} = \left(\bigcap_{i=1}^{p} \left\{ x \in U \mid g_i(x) = 0 \right\} \right) \cap \left(\bigcap_{j=1}^{q} \left\{ x \in U \mid h_j(x) \leq 0 \right\} \right)$$

soit encore
$$\mathcal{A} = \left(\bigcap_{i=1}^{p} g_i^{-1}(\{0\}) \right) \cap \left(\bigcap_{j=1}^{q} h_j^{-1}(]-\infty\,;0\,]) \right)$$

L'ensemble admissible est donc fermé comme intersection finie d'images réciproques de fermés par des fonctions continues (car différentiables). Notons que, comme toute intersection d'ensembles (vides ou non), l'ensemble admissible n'est pas nécessairement non vide.

Remarque : Puisque le problème $(\mathcal{P}_c)$ est la minimisation d'une fonction sur un ensemble **fermé**, on ne peut pas appliquer les conditions d'optimalité du premier ordre et du second ordre énoncées dans le module **B4 : Optimisation sous contraintes**.

1.2 Énoncé général du théorème

On va dès maintenant énoncer la forme générale du théorème de Karush–Kuhn–Tucker. Les hypothèses d'application de ce théorème, ainsi que les preuves, seront précisées dans les deux sections suivantes.

Théorème 1 (KARUSH–KUHN–TUCKER)

Soit $U \subset E$ un ouvert. Soit $f : U \to \mathbb{R}$, $g_i : U \to \mathbb{R}$, avec $i \in [\![1\,;p]\!]$ et $h_j : U \to \mathbb{R}$, avec $j \in [\![1\,;q]\!]$ des fonctions de classe $\mathcal{C}^1$. On considère le problème d'optimisation sous contraintes suivant

$$\text{Minimiser } f(x) \text{ sous les contraintes } \begin{cases} g_i(x) = 0 & \text{pour } i \in [\![1\,;p]\!] \\ h_j(x) \leq 0 & \text{pour } j \in [\![1\,;q]\!] \end{cases} \qquad (\mathcal{P}_c)$$

Soit $x^* \in U$ une solution de $(\mathcal{P}_c)$. On suppose que **les contraintes sont qualifiées**. Alors

$$\forall\, i \in [\![1\,;p]\!], \quad g_i(x^*) = 0 \qquad \text{et} \qquad \forall\, j \in [\![1\,;q]\!], \quad h_j(x^*) \leq 0$$

et il existe $p+q$ réels $\lambda_1, \ldots, \lambda_p, \mu_1, \ldots, \mu_q$, appelés *multiplicateurs de LAGRANGE* associés à x^*, tels que

$$\nabla f(x^*) + \sum_{i=1}^{p} \lambda_i \nabla g_i(x^*) + \sum_{j=1}^{q} \mu_j \nabla h_j(x^*) = 0$$

avec $\qquad \forall\, j \in [\![1\,;q]\!], \quad \mu_j\, h_j(x^*) = 0 \qquad \text{et} \qquad \mu_j \geq 0$

> Le sens précis de la phrase
>
> **Les contraintes sont qualifiées**
>
> sera précisé ultérieurement, mais on peut le traduire par “le problème $(\mathcal{P}_c)$ vérifie des hypothèses sous lesquelles le théorème 1 est vrai”, ou encore “le problème $(\mathcal{P}_c)$ vérifie des hypothèses qui assurent l'existence de multiplicateurs de LAGRANGE pour toute solution du problème".

On appelle *conditions de KARUSH–KUHN–TUCKER*, ou *conditions KKT*, les trois (ensembles) de conditions qui apparaissent dans l'énoncé précédent. Plus précisément, les conditions

$$\forall\, i \in [\![1\,;p]\!], \quad g_i(x^*) = 0 \qquad \text{et} \qquad \forall\, j \in [\![1\,;q]\!], \quad h_j(x^*) \leq 0$$

sont appelées *conditions d'admissibilité*, puisqu'elles traduisent simplement le fait qu'une solution est admissible ; l'égalité

$$\nabla f(x^*) + \sum_{i=1}^{p} \lambda_i \nabla g_i(x^*) + \sum_{j=1}^{q} \mu_j \nabla h_j(x^*) = 0$$

est appelée dans ce cours *condition du premier ordre*, car elle ne fait intervenir que des gradients ; enfin, les égalités

$$\mu_j\, h_j(x^*) = 0$$

sont appelées *conditions de complémentarité*. En effet, elle assure que soit le multiplicateur de LAGRANGE μ_j est nul, soit la contrainte associée est *saturée* (c'est-à-dire que la contrainte d'inégalité s'annule).

REMARQUE : Le théorème 1 est essentiellement un résultat d'existence des multiplicateurs de LAGRANGE, mais pas des solutions du problème : l'existence de celles-ci fait partie des hypothèses du théorème.

1.3 Cas des contraintes non saturées

Commençons par faire quelques observations d'ordre général sur le théorème 1. Si x^* est une solution du problème $(\mathcal{P}_c)$ et si le théorème 1 est vrai, alors on remarque que le gradient de la fonction objectif f en x^* peut ne s'exprimer qu'en fonction des gradients des fonctions g_i et h_j qui s'annulent en x^*, puisque les μ_j sont nuls si $h_j(x^*) < 0$. L'idée derrière cette observation est que, localement autour de x^*, les contraintes non saturées "ne se voient pas" (cf. Figure 1). Si l'on considère un problème plus simple

$$\text{Minimiser } f \text{ sous les contraintes } h(x) \leq 0$$

où les fonctions f et h sont différentiables et définies sur l'ouvert U, alors deux situations se présentent : soit la solution x^* sature la contrainte (donc $h(x^*) = 0$), et la condition du premier ordre dans le théorème 1 assure que

$$\nabla f(x^*) = -\mu \nabla h(x^*)$$

avec $\mu \geq 0$; soit x^* ne sature pas la contrainte (donc $h(x^*) < 0$), et la condition de complémentarité dans le théorème 1 assure que le multiplicateur de Lagrange associé est nul, de sorte que la condition du premier ordre devient

$$\nabla f(x^*) = -\mu \nabla h(x^*) = 0$$

Autrement dit, x^* est un point critique de f. Ce n'est pas surprenant : si on considère l'ensemble $U_{x^*} = \{x \in U \mid h(x) < 0\}$, on peut montrer qu'il est ouvert (comme image réciproque de l'ouvert $]-\infty\,;0\,[$ par la fonction continue h), contenu dans U et non vide (puisqu'il contient x^* par hypothèse). Aussi, on peut considérer le problème

$$\text{Minimiser } \tilde{f} \text{ sous les contraintes } \tilde{h}(x) \leq 0$$

où $\tilde{f}$ et $\tilde{h}$ sont respectivement les restrictions à U_{x^*} des fonctions f et h. Puisque x^* est admissible pour ce problème, et que l'ensemble admissible de ce problème est contenu dans l'ensemble admissible du premier problème, on en déduit que x^* est solution de ce nouveau problème. Enfin, l'ensemble admissible de ce problème étant un ouvert, la proposition 3 du module **B4 : Optimisation sous contraintes** assure que $\nabla f(x^*) = 0$.

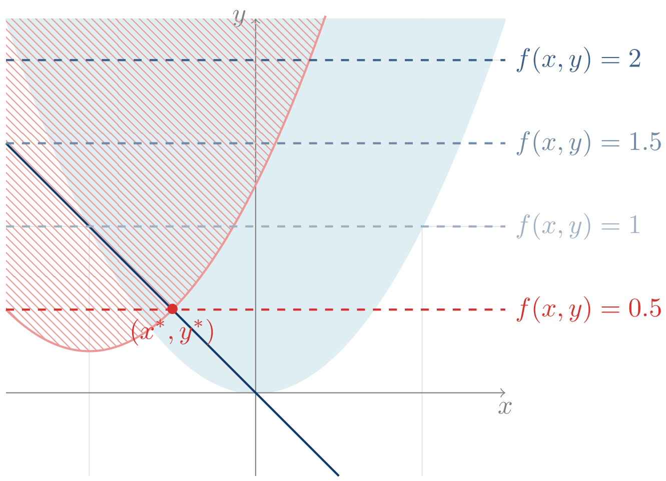

Figure 1 – Cas d'une contrainte non saturée à l'optimum. Ici, on a choisi $f(x,y) = y$, une seule contrainte d'égalité $g(x,y) = x + y$ (en bleu foncé), deux contraintes d'égalité définies par $h_1(x,y) = x^2 - y$ (en bleu clair) et $h_2(x,y) = (x+1)^2 + 1/4 - y$ (en rouge hachuré). La solution vaut $(x^*, y^*) = (-1/2, 1/2)$, qui sature g et h_2.

Appliquons ce raisonnement au cas général du problème $(\mathcal{P}_c)$. On note $\mathcal{J}_{x^*}$ l'ensemble des indices $j \in [\![1\,;\,q]\!]$ correspondant aux contraintes saturées par x^*, c'est-à-dire pour lesquels $h_j(x^*) = 0$. Posons

$$U_{x^*} = \left\{ x \in U \mid \forall\, j \notin \mathcal{J}_{x^*}, h_j(x) < 0 \right\}$$

Notons que cet ensemble est ouvert (comme intersection d'ouverts) et qu'il contient x^*. Considérons à présent le problème d'optimisation sous contraintes

$$\text{Minimiser } \tilde{f} \text{ sous les contraintes } \begin{cases} \tilde{g}_i(x) = 0 & \text{pour } i \in [\![1\,;\,p]\!] \\ \tilde{h}_j(x) \leq 0 & \text{pour } j \in \mathcal{J}_{x^*} \end{cases}$$

où à nouveau $\tilde{f}$, $\tilde{g}_i$ et $\tilde{h}_j$ sont les restrictions à l'ouvert U_{x^*} des fonctions f, g_i et h_j. Si $x \in U_{x^*}$ est admissible pour ce problème, alors il est admissible pour le problème $(\mathcal{P}_c)$. Il s'ensuit que $f(x) \geq f(x^*)$ pour tout tel x, de sorte que x^* est également solution de ce problème. Ainsi, on peut virtuellement ignorer les contraintes h_j pour $j \notin \mathcal{J}_{x^*}$ (en les intégrant de manière implicite dans la définition des fonctions apparaissant dans le problème d'optimisation). Autrement dit, pour la démonstration du théorème 1, on peut supposer que x^* sature toutes les contraintes d'inégalité.

Dans les deux sections qui suivent, on va s'attacher à préciser les conditions d'application du théorème 1, en définissant précisément la notion de *qualification des contraintes*, et fournir les preuves associées.

2 Optimisation sur une sous-variété de $\mathbb{R}^n$

Dans cette section, on va commencer par considérer le cas où l'ensemble admissible $\mathcal{A}$ de $(\mathcal{P}_c)$ est une sous-variété, décrite à l'aide d'une submersion. On va ensuite étendre le résultat obtenu à des ensembles plus généraux.

2.1 Cas de contraintes d'égalité

Dans ce paragraphe, on suppose que $\mathcal{A}$ n'est décrit qu'à l'aide de contraintes d'égalité, c'est-à-dire qu'on considère le problème

$$\text{Minimiser } f(x) \text{ sous les contraintes } g_i(x) = 0 \text{ pour } i \in [\![1\,;\,p]\!] \qquad (\mathcal{P}_{ce})$$

On pose

$$G : \left\{ \begin{array}{ccc} U & \to & \mathbb{R}^p \\ x & \mapsto & \big(g_i(x)\big)_{1 \leq i \leq p} \end{array} \right.$$

Ainsi, par définition, toute solution x^* du problème $(\mathcal{P}_c)$ est un point appartenant à l'ensemble admissible $G^{-1}(\{0\})$ vérifiant

$$\forall\, x \in G^{-1}(\{0\}), \qquad f(x^*) \leq f(x)$$

On suppose que G est une submersion de classe $\mathcal{C}^1$ sur U, c'est-à-dire que $dg(x)$ surjective pour tout $x \in U$, ou encore que les $\nabla g_i(x)$ forment une famille libre de E. Il s'ensuit que l'ensemble admissible $G^{-1}(\{0\})$ est une sous-variété de $\mathbb{R}^n$. Dans ce cas, le théorème 1 n'est rien d'autre qu'une réécriture du théorème des extrema liés (module **A5 : Sous-variétés de $\mathbb{R}^n$. Théorème des extrema liés.**). Celui-ci assure en effet que, si x^* est une solution de $(\mathcal{P}_c)$, alors il existe $\tilde{\lambda}_1, \ldots, \tilde{\lambda}_p \in \mathbb{R}$ tels que

$$\nabla f(x^*) = \sum_{i=1}^{p} \tilde{\lambda}_i \nabla g_i(x^*)$$

En définissant $\lambda_i = -\tilde{\lambda}_i$, on retrouve la condition du premier ordre du théorème 1. Les conditions d'admissibilité sont acquises par définition d'une solution et les conditions de complémentarité n'ont pas lieu d'être car il n'y a pas de contrainte d'inégalité. On est donc amené à proposer une première définition pour la phrase *les contraintes sont qualifiées en* x^*, valable lorsqu'il n'y a pas de contraintes d'inégalité :

Définition 1 (Qualification des contraintes d'égalité)

Soit $U \subset \mathbb{R}^n$ un ouvert. Soit $f : U \to \mathbb{R}$ et $g_i : U \to \mathbb{R}$, avec $i \in [\![1\,;\,p]\!]$ des fonctions de classe $\mathcal{C}^1$. On considère le problème d'optimisation sous contraintes **d'égalité** suivant

$$\text{Minimiser } f(x) \text{ sous les contraintes } g_i(x) = 0 \text{ pour } i \in [\![1\,;\,p]\!] \qquad (\mathcal{P}_{ce})$$

On dit que les contraintes sont *qualifiées* si, pour **tout** x **admissible**, les $\nabla g_i(x)$ forment une famille libre pour $i \in [\![1\,;\,p]\!]$.

Notons que, selon cette définition, si les contraintes sont qualifiées, alors l'ensemble admissible est une sous-variété. La réciproque est fausse ; en effet, si une contrainte est redondante (c'est-à-dire si elle apparaît deux fois dans les fonctions g_i), l'ensemble admissible n'est pas modifié, mais les $\nabla g_i(x)$ ne peuvent plus définir une famille libre.

Exemple

Minimisation sur le cercle. On considère le problème d'optimisation sous contrainte d'égalité suivant

$$\text{Minimiser } f(x, y) = x \text{ sous la contrainte } g(x, y) = (x-1)^2 + y^2 - 1 = 0$$

L'ensemble admissible est le cercle de centre $(1, 0)$ et de rayon 1. En particulier, on voit que la fonction f est positive sur l'ensemble admissible (car $x \in [\,0\,;2\,]$). Puisque (x, y) est admissible si et seulement si $x = 1 \pm \sqrt{1-y^2}$, et que

$$\forall\, y \in \mathbb{R}, \qquad f(1 \pm \sqrt{1-y^2}, y) \geq 0 = f(0, 0)$$

avec $(0, 0)$ un point admissible, on en déduit que $(0, 0)$ est une solution du problème considéré. On a par ailleurs

$$\nabla f(0, 0) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \text{et} \qquad \nabla g(0, 0) = \begin{pmatrix} -2 \\ 0 \end{pmatrix}$$

On a bien $\nabla f(0, 0) + \nabla g(0, 0)/2 = 0$. On en déduit que le multiplicateur de Lagrange associé à la solution $(0, 0)$ vaut $1/2$.

La liberté de la famille des $\nabla g_i(x)$ est une condition essentielle, comme le montre le contre-exemple suivant :

Contre-exemple

Contraintes non qualifiées. On considère le problème d'optimisation sous contrainte d'égalité suivant

$$\text{Minimiser } f(x, y) = x \text{ sous la contrainte } g(x, y) = x^3 - \sin y - 1 = 0$$

L'ensemble admissible est l'ensemble des points $(\sqrt[3]{\sin y + 1}, y)$, pour $y \in \mathbb{R}$. En particulier, on voit que la fonction f est positive sur l'ensemble admissible (puisque $\sin y \geq -1$). Puisque

$$\forall\, y \in \mathbb{R}, \qquad f(\sqrt[3]{\sin y + 1}, y) \geq 0 = f(0, \pi/2)$$

on en déduit que $(0, -\pi/2)$ est une solution du problème considéré. On a par ailleurs

$$\nabla f(0, -\pi/2) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \text{et} \qquad \nabla g(0, -\pi/2) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Il apparaît clairement que ces deux vecteurs ne sont pas colinéaires. Cela est dû au fait que g n'est pas une submersion en $(0, -\pi/2)$.

La qualification des contraintes d'égalité est une condition suffisante pour l'application du théorème 1, mais pas nécessaire.

Contre-exemple

Contraintes affines non qualifiées au sens de l'indépendance linéaire. On considère le problème d'optimisation sous contrainte d'égalité suivant

$$\text{Minimiser } f(x, y) = x \text{ sous les contraintes } \begin{cases} g_1(x, y) = x = 0 \\ g_2(x, y) = 2\,x = 0 \end{cases}$$

Il est clair que ce problème admet une infinité de solutions, car la fonction f est constante sur l'ensemble admissible, qui est l'axe des ordonnées $x = 0$. Or, on a pour tout $y \in \mathbb{R}$

$$\nabla f(0, y) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad \nabla g_1(0, y) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \text{et} \qquad \nabla g_2(0, y) = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$$

En particulier, on a

$$\forall\, y \in \mathbb{R}, \qquad \nabla f(0, y) - 2\,\nabla g_1(0, y) - \frac{1}{2}\,\nabla g_2(0, y) = 0$$

les multiplicateurs de Lagrange associés à la solution $(0, y)$ étant $(-2, -1/2)$.

> En réalité, on verra dans la section suivante que les contraintes de tout problème d'optimisation convexe dont l'ensemble admissible décrit à l'aide de contraintes d'égalité affines sont qualifiées (au sens de Slater) dès lors que le problème admet au moins un point admissible. De manière plus générale, il n'existe pas de condition de qualification absolue ; il s'agit toujours de conditions suffisantes, mais non nécessaires.

La qualification des contraintes est une propriété qui n'est pas intrinsèque à l'ensemble admissible du problème ; au contraire, elle dépend de la **description** de cet ensemble.

Contre-exemple

Contraintes non qualifiées. On considère le problème d'optimisation sous contrainte d'égalité suivant

$$\text{Minimiser } f(x, y) = x \text{ sous les contraintes } \begin{cases} g_1(x, y) = x^2 + (y-1)^2 - 1 = 0 \\ g_2(x, y) = x^2 + (y+1)^2 - 1 = 0 \end{cases}$$

L'ensemble admissible est le singleton $\{(0, 0)\}$. Ainsi, le point $(0, 0)$ est donc l'unique solution du problème considéré. On a par ailleurs

$$\nabla f(0, 0) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad \nabla g_1(0, 0) = \begin{pmatrix} 0 \\ -1 \end{pmatrix} \qquad \text{et} \qquad \nabla g_2(0, 0) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Ainsi, il n'est pas possible d'écrire le vecteur $\nabla f(0,0)$ comme combinaison linéaire des vecteurs $\nabla g_1(0,0)$ et $\nabla g_2(0,0)$. Les contraintes ne sont en effet pas qualifiées en $(0,0)$ (puisque les deux vecteurs $\nabla g_i(0,0)$ sont colinéaires). Dans le problème considéré, l'ensemble admissible est en réalité décrit comme l'intersection entre deux cercles tangents ; il est possible de le décrire autrement. Par exemple, comme l'intersection entre l'axe des abscisses et l'axe des ordonnées. Le problème d'optimisation associé est donc donné par

$$\text{Minimiser } f(x,y) = x \text{ sous les contraintes } \begin{cases} \tilde g_1(x,y) = x = 0 \\ \tilde g_2(x,y) = y = 0 \end{cases}$$

Puisque la fonction objectif est la même, et que l'ensemble admissible reste inchangé (malgré le changement d'écriture), ce problème d'optimisation est le même que celui considéré précédemment (en particulier, il admet le même ensemble de solutions). On a cette fois

$$\nabla f(0,0) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad \nabla \tilde g_1(0,0) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \text{et} \qquad \nabla \tilde g_2(0,0) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

On voit donc que les vecteurs $\nabla \tilde g_1(0,0)$ et $\nabla \tilde g_2(0,0)$ sont linéairement indépendants. On a par ailleurs

$$\nabla f(0,0) - 1 \times \nabla \tilde g_1(0,0) + 0 \times \nabla \tilde g_2(0,0) = 0$$

2.2 Cas de contraintes d'inégalité

Ajoutons des contraintes d'inégalité. On a vu dans le module **B4 : Optimisation sous contraintes** qu'il est toujours possible, moyennant l'ajout d'une variable, de transformer une contrainte d'inégalité en contrainte d'égalité. Plus précisément, en posant

$$\forall\,(x,\varepsilon_1,\ldots,\varepsilon_q) \in U \times \mathbb{R}^q, \qquad \tilde f(x,\varepsilon_1,\ldots,\varepsilon_q) = f(x)$$

$$\text{et} \quad \forall\, i \in [\![\, 1\,;\, p+q \,]\!], \qquad \tilde g_i(x,\varepsilon_1,\ldots,\varepsilon_q) = \begin{cases} g_i(x) & \text{si } i \in [\![\, 1\,;\, p \,]\!] \\ h_{i-p}(x) + \varepsilon_{i-p}^2 & \text{si } i \in [\![\, p+1\,;\, p+q \,]\!] \end{cases}$$

on a montré que le problème $(\mathcal{P}_c)$ est équivalent au problème

$$\text{Minimiser } \tilde f(x,\varepsilon) \text{ sous la contrainte } \tilde g(x,\varepsilon) = 0 \qquad (\mathcal{P}_{ce})$$

dans le sens où :

- si (x^*,ε^*) est solution de $(\mathcal{P}_{ce})$, alors x^* est solution de $(\mathcal{P}_c)$;
- si x^* est solution de $(\mathcal{P}_c)$, alors $\left(x^*, \sqrt{-h_1(x^*)}, \ldots, \sqrt{-h_q(x^*)}\right)$ est solution de $(\mathcal{P}_{ce})$.

2.3 Qualification au sens de l'indépendance linéaire

Appliquons les résultats obtenus dans le premier paragraphe de cette section au problème $(\mathcal{P}_{ce})$. Si pour tout $(x,\varepsilon_1,\ldots,\varepsilon_q)$ admissible, les $\nabla \tilde g_i(x,\varepsilon_1,\ldots,\varepsilon_q)$ forment une famille libre, alors toute solution $(x^*,\varepsilon_1^*,\ldots,\varepsilon_q^*)$ de $(\mathcal{P}_{ce})$ est telle qu'il existe $p+q$ réels $\tilde\lambda_1,\ldots,\tilde\lambda_{p+q}$ tels que

$$\nabla \tilde f(x^*,\varepsilon_1^*,\ldots,\varepsilon_q^*) + \sum_{i=1}^{p+q} \tilde\lambda_i\, \nabla \tilde g_i(x^*,\varepsilon_1^*,\ldots,\varepsilon_q^*) = 0$$

Dans ce qui suit, on va

1. traduire les hypothèses de ce résultat en hypothèses sur les fonctions f, g_i et h_j ;

2. exprimer la conclusion en fonction des fonctions f, g_i et h_j.

Étape 1. Commençons par considérer les hypothèses du théorème 1. D'après le paragraphe précédent, le problème $(\mathcal{P}_c)$ admet une solution x^* si et seulement si le problème $(\mathcal{P}_{ce})$ admet une solution (x^*, ε^*). Intéressons-nous à présent à la condition de qualification donnée dans la définition 1. Puisque $\tilde{g}_i(x, \varepsilon_1, \ldots, \varepsilon_q) = g_i(x)$ pour tout $x \in U$ lorsque $i \in [\![1\,;\, p]\!]$, on a d'une part

$$\forall\, i \in [\![1\,;\, p]\!]\,, \qquad \nabla \tilde{g}_i(x, \varepsilon_1, \ldots, \varepsilon_q) = \begin{pmatrix} \nabla g_i(x) \\ 0 \end{pmatrix} \in \mathbb{R}^{p+q}$$

D'autre part, comme $\tilde{g}_i(x, \varepsilon_1, \ldots, \varepsilon_q) = h_{i-p}(x) + \varepsilon_{i-p}^2$ lorsque $i \in [\![p+1\,;\, p+q]\!]$, on a

$$\forall\, i \in [\![p+1\,;\, p+q]\!]\,, \qquad \nabla \tilde{g}_i(x, \varepsilon_1, \ldots, \varepsilon_q) = \begin{pmatrix} \nabla h_{i-p}(x) \\ 2\, \varepsilon_{i-p}\, e_{i-p} \end{pmatrix}$$

où $e_{i-p} \in \mathbb{R}^q$ est le $(i-p)$-ème vecteur de la base canonique de $\mathbb{R}^q$. Ainsi, les $\nabla \tilde{g}_i(x, \varepsilon_1, \ldots, \varepsilon_q)$ forment une famille libre si et seulement si pour tout $(t_1, \ldots, t_p, u_1, \ldots, u_q) \in \mathbb{R}^{p+q}$

$$\begin{pmatrix} \displaystyle\sum_{i=1}^{p} t_i\, \nabla g_i(x) + \sum_{j=1}^{q} u_j\, \nabla h_j(x) \\ 2\, u_1\, \varepsilon_1 \\ \vdots \\ 2\, u_q\, \varepsilon_q \end{pmatrix} = 0 \qquad \Longrightarrow \qquad t_1 = \cdots = t_p = u_1 = \cdots = u_q = 0$$

Rappelons que $\varepsilon_j = \sqrt{-h_j(x)}$ pour tout $(x, \varepsilon_1, \ldots, \varepsilon_q)$ admissible. En particulier, on a

$$2\, u_j\, \varepsilon_j = 0 \qquad \Longleftrightarrow \qquad \big(u_j = 0 \quad \text{ou} \quad h_j(x) = 0\big)$$

Autrement dit, si $h_j(x) \neq 0$, on a nécessairement $u_j = 0$. Ainsi, on a

$$\sum_{i=1}^{p} t_i\, \nabla g_i(x) + \sum_{j=1}^{q} u_j\, \nabla h_j(x) = \sum_{i=1}^{p} t_i\, \nabla g_i(x) + \sum_{\substack{j=1 \\ h_j(x)=0}}^{q} u_j\, \nabla h_j(x)$$

Il s'ensuit que les $\nabla g_i(x)$ pour $i \in [\![1\,;\, p]\!]$ et les $\nabla h_j(x)$ tels que $h_j(x) = 0$ forment une famille libre si et seulement si les $\nabla \tilde{g}_i(x, \varepsilon_1, \ldots, \varepsilon_q)$ forment une famille libre.

On va anticiper le succès de ce travail de transposition et introduire dès à présent la définition suivante :

Définition 2 (Condition de l'indépendance linéaire (IL))

Soit $U \subset E$ un ouvert. Soient $f : U \to \mathbb{R}$, $g_i : U \to \mathbb{R}$, avec $i \in [\![1\,;\, p]\!]$ et $h_j : U \to \mathbb{R}$, avec $j \in [\![1\,;\, q]\!]$ des fonctions de classe $\mathcal{C}^1$. On considère le problème d'optimisation sous contraintes suivant

$$\text{Minimiser } f(x) \text{ sous les contraintes } \begin{cases} g_i(x) = 0 & \text{pour } i \in [\![1\,;\, p]\!] \\ h_j(x) \leq 0 & \text{pour } j \in [\![1\,;\, q]\!] \end{cases} \qquad (\mathcal{P}_c)$$

On dit que les contraintes sont *qualifiées* au sens de l'indépendance linéaire (IL) si, pour tout x admissible, les $\nabla g_i(x)$ pour $i \in [\![1\,;\, p]\!]$ et les $\nabla h_j(x)$ pour $j \in \mathcal{J}_x$ (c'est-à-dire tels que $h_j(x) = 0$) forment une famille libre.

On vient donc de démontrer que si les contraintes du problème $(\mathcal{P}_c)$ sont qualifiées au sens de l'indépendance linéaire (définition 2), alors les contraintes du problème $(\mathcal{P}_{ce})$ sont qualifiées au sens de la définition 1.

Étape 2. Considérons maintenant la condition du premier ordre. Puisque

$$\nabla \tilde{f}(x, \varepsilon_1, \ldots, \varepsilon_q) = \begin{pmatrix} \nabla f(x) \\ 0 \end{pmatrix}$$

il s'ensuit que le théorème 1 assure qu'il existe $\tilde{\lambda}_1, \ldots, \tilde{\lambda}_{p+q}$ tels que

$$\begin{pmatrix} \nabla f(x^*) + \sum_{i=1}^{p} \tilde{\lambda}_i \, \nabla g_i(x^*) + \sum_{j=1}^{q} \tilde{\lambda}_{p+j} \, \nabla h_j(x^*) \\ 2\, \tilde{\lambda}_{p+1}\, \varepsilon_1^* \\ \vdots \\ 2\, \tilde{\lambda}_{p+q}\, \varepsilon_q^* \end{pmatrix} = 0$$

En particulier, on a d'une part

$$\nabla f(x^*) + \sum_{i=1}^{p} \tilde{\lambda}_i \, \nabla g_i(x^*) + \sum_{j=1}^{q} \tilde{\lambda}_{p+j} \, \nabla h_j(x^*) = 0$$

ce qui, en posant $\lambda_i = \tilde{\lambda}_i$ pour tout $i \in [\![1 \,;\, p]\!]$ et $\mu_j = \tilde{\lambda}_{p+j}$, correspond à la **condition du premier ordre** du théorème 1. D'autre part,

$$\forall\, j \in [\![1 \,;\, q]\!], \qquad \tilde{\lambda}_{p+j}\, \varepsilon_j^* = \mu_j \sqrt{-h_j(x^*)} = 0$$

Autrement dit, on a soit $\mu_j = 0$, soit $h_j(x^*) = 0$, c'est-à-dire que

$$\forall\, j \in [\![1 \,;\, q]\!], \qquad \mu_j \, h_j(x^*) = 0$$

ce qui est la première partie de la **condition de complémentarité** qui apparaît dans le théorème 1.

Il reste à montrer que les μ_j sont **positifs**. Commençons par remarquer que, d'après ce qui précède, si $h_j(x^*) \neq 0$, alors $\mu_j = 0$. Ainsi, on suppose que $\mathcal{J}_{x^*}$ est non vide ; il suffit d'étudier le signe des μ_j pour $j \in \mathcal{J}_{x^*}$. Puisque, si $j \notin \mathcal{J}_{x^*}$, alors $\mu_j = 0$, on a

$$\nabla f(x^*) + \sum_{i=1}^{p} \lambda_i \, \nabla g_i(x^*) + \sum_{j \in \mathcal{J}_{x^*}} \mu_j \, \nabla h_j(x^*) = 0$$

On définit $\varphi_0 = (g_1, \ldots, g_p, (h_j)_{j \in \mathcal{J}_{x^*}})$ et $\varphi = (g_1, \ldots, g_p, (h_j)_{j \in \mathcal{J}_{x^*} \setminus \{j_0\}})$ pour $j_0 \in \mathcal{J}_{x^*}$. Les applications φ_0 et φ sont des submersions en x^* (par liberté des gradients de ses composantes). Ainsi, $\varphi_0^{-1}(\{0\})$ et $\varphi^{-1}(\{0\})$ définissent des sous-variétés. La dimension d'une sous-variété étant égale à celle de l'ensemble de ses vecteurs tangents en un de leur point, on a $\dim \ker d_{x^*}\varphi = \dim \ker d_{x^*}\varphi_0 + 1$. Par définition de ces deux sous-espaces vectoriels, on a par ailleurs

$$\ker d_{x^*}\varphi_0 \subset \ker d_{x^*}\varphi$$

Il s'ensuit qu'il existe un vecteur $v \in \mathbb{R}^n$ non nul tel que $v \in \ker d_{x^*}\varphi \setminus \ker d_{x^*}\varphi_0$; autrement dit,

$$\forall\, i \in [\![1 \,;\, p]\!], \quad \langle \nabla g_i(x^*), v \rangle = 0, \qquad \forall\, j \in \mathcal{J}_{x^*} \setminus \{j_0\}, \quad \langle \nabla h_j(x^*), v \rangle = 0$$

mais $\langle \nabla h_{j_0}(x^*), v \rangle \neq 0$. Il s'ensuit donc que

$$\langle \nabla f(x^*), v \rangle = -\mu_{j_0} \, \langle \nabla h_{j_0}(x^*), v \rangle$$

> En particulier, on voit que, dans le cas d'un problème sous contraintes d'égalité qualifiées au sens de l'indépendance linéaire, lorsqu'il existe une solution x^*, alors on a
>
> $$\langle \nabla f(x^*), v \rangle = 0$$
>
> pour tout v vecteur tangent à l'ensemble admissible en x^*.

Par définition d'un vecteur tangent, il existe une courbe $\gamma : \left]-\delta\,;\delta\right[\to U$ telle que

$$\forall\, t \in \left]-\delta\,;\delta\right[, \qquad \begin{cases} \forall\, i \in \mathcal{I}_{x^*}, & g_i(\gamma(t)) = 0, \\ \forall\, j \in \mathcal{J}_{x^*} \setminus \{j_0\}, & h_j(\gamma(t)) = 0 \end{cases}$$

et telle que $(\gamma(0), \gamma'(0)) = (x^*, v)$. Soit $t \in \left]-\delta\,;\delta\right[$. Puisque les fonctions en jeu sont différentiables sur l'ouvert U, on a pour t voisin de zéro

$$f(\gamma(t)) = f(x^*) + t\,\langle \nabla f(x^*), v\rangle + o(|t|) = f(x^*) - t\,\mu_{j_0}\,\langle \nabla h_{j_0}(x^*), v\rangle + o(|t|)$$

et
$$h_{j_0}(\gamma(t)) = t\,\langle \nabla h_{j_0}(x^*), v\rangle + o(|t|)$$

De cette dernière égalité, on établit que, puisque $h_{j_0}(\gamma(0)) = h_{j_0}(x^*) = 0$ par définition de $\mathcal{J}_{x^*}$ et que $\langle \nabla h_{j_0}(x^*), v\rangle \neq 0$, la quantité $h_{j_0}(\gamma(t))$ change de signe en $t = 0$. Plus précisément, quitte à remplacer v par $-v$, on peut supposer que $\langle \nabla h_{j_0}(x^*), v\rangle < 0$. Dans ce cas, $h_{j_0}(\gamma(t)) < 0$ pour $t > 0$ voisin de 0. Il s'ensuit que, pour de tels t,

$$\begin{cases} \forall\, i \in \mathcal{I}_{x^*}, & g_i(\gamma(t)) = 0, \\ \forall\, j \in \mathcal{J}_{x^*} \setminus \{j_0\}, & h_j(\gamma(t)) = 0 \\ h_{j_0}(\gamma(t)) < 0 \end{cases}$$

Autrement dit, les points $\gamma(t)$ sont admissibles pour le problème considéré. On en déduit par optimalité que $f(\gamma(t)) \geq f(x^*)$ pour de tels t. Il faut donc nécessairement que

$$t\,\mu_{j_0}\,\langle \nabla h_{j_0}(x^*), v\rangle \leq 0$$

Puisque $t > 0$ et $\langle \nabla h_{j_0}(x^*), v\rangle < 0$, il s'ensuit que $\mu_{j_0} \geq 0$.

Ainsi, on a finalement démontré que, si les contraintes du problème sont qualifiées au sens de l'indépendance linéaire, le théorème 1 est vrai.

> Dans la démonstration ci-dessus, on n'utilise en réalité que le fait que les vecteurs $\nabla g_i(x^*)$ pour $i \in [\![1\,;p]\!]$ et $\nabla h(x^*)$ pour $j \in [\![1\,;q]\!]$ tel que $h_j(x^*) = 0$ forment une famille libre. Théoriquement, il n'est donc pas nécessaire de vérifier cette propriété pour tout x admissible. Cependant, en pratique, puisque x^* n'est pas connu, il est général plus simple de faire cette étude sur tout l'ensemble admissible.

3 Optimisation convexe

3.1 Condition de qualification de Slater

Les deux conditions de qualification des contraintes introduites dans la section précédente (qui n'en forment en réalité qu'une seule, puisque la première est un cas particulier de la seconde) reposent sur une condition de liberté d'une famille de vecteurs. Or, au-delà d'un certain cardinal, une famille de vecteurs ne peut être libre. Puisque la famille de vecteurs considérée est *a priori* d'autant plus grande qu'il y a de contraintes dans le problème, il s'ensuit mécaniquement qu'un problème très contraint (c'est-à-dire présentant un nombre important de contraintes distinctes) a peu de chances d'avoir des contraintes qualifiées (au sens IL). Dans ce paragraphe, on va donc introduire une autre condition de qualification, qui va nous permettre d'éviter l'écueil que l'on vient de décrire.

On considère toujours le problème d'optimisation sous contraintes $(\mathcal{P}_c)$ et on suppose que les g_i sont **affines** et que les h_j sont **convexes**, de sorte que l'ensemble admissible soit convexe. On introduit la définition suivante :

Définition 3 (Condition de qualification de SLATER)

Soit $U \subset E$ un ouvert. Soit $f : U \to \mathbb{R}$ une fonction, $h_j : U \to \mathbb{R}$, $j \in [\![1\,;\,q]\!]$ des fonctions différentiables et **convexes** et $g_i : U \to \mathbb{R}$, $i \in [\![1\,;\,p]\!]$ des fonctions **affines**. On considère le problème d'optimisation sous contraintes suivant

$$\text{Minimiser } f(x) \text{ sous les contraintes } \begin{cases} g_i(x) = 0 & \text{pour } i \in [\![1\,;\,p]\!] \\ h_j(x) \leq 0 & \text{pour } j \in [\![1\,;\,q]\!] \end{cases} \qquad (\mathcal{P}_c)$$

Posons $\mathcal{J}$ l'ensemble des indices $1 \leq j \leq q$ correspondant à une contrainte h_j affine. On dit que les contraintes sont *qualifiées* au sens de SLATER s'il existe un vecteur **admissible** $x_0 \in U$ tel que

$$\forall\, 1 \leq j \leq q, \qquad \begin{cases} h_j(x_0) \leq 0 & \text{si } j \in \mathcal{J} \\ h_j(x_0) < 0 & \text{si } j \notin \mathcal{J} \end{cases}$$

> Notons que, dans la définition 3, il y a une hypothèse sur la convexité de la fonction objectif, alors que la formulation "les contraintes sont qualifiées" sous-entend que seules des hypothèses sur les contraintes sont énoncées. En réalité, il s'agit d'un choix fait ici pour simplifier l'énoncé du théorème KKT.

Dans le cas de la condition de SLATER, le caractère qualifié des contraintes dépend à nouveau de la description de l'ensemble admissible.

CONTRE-EXEMPLE

Contraintes non qualifiées. Considérons le problème suivant

$$\text{Minimiser } f(x,y) = x \text{ sous les contraintes } \begin{cases} y \leq 0 \\ \min(x,0)^2 - y \leq 0 \end{cases}$$

L'ensemble des points admissibles de ce problème sont les points

$$\mathcal{A} = \left\{ (x,0) \in \mathbb{R}^2 \mid x \geq 0 \right\}$$

Or, pour tous ces points, la seconde contrainte s'annule ; les contraintes de ce problème ne sont donc pas qualifiées. Considérons maintenant le problème

$$\text{Minimiser } f(x,y) = x \text{ sous les contraintes } \begin{cases} -x \leq 0 \\ \quad y = 0 \end{cases}$$

Il est alors aisé de vérifier qu'il s'agit du même problème. Cependant, sous cette écriture, les contraintes sont qualifiées (car elles sont affines, et que le problème est réalisable).

3.2 Démonstration du théorème 1

La démonstration du théorème 1 dans le cas de la condition de SLATER *utilise des résultats non vus dans ce cours. Aussi, on se contente ici d'en expliquer le principe.*

Rappelons que, dans le paragraphe §1.3, on a montré que, quitte à restreindre la fonction objectif et les fonctions définissant les contraintes à U_{x^*}, on peut supposer que x^* sature toutes les contraintes. Notons par ailleurs que, sous les hypothèses de cette section, l'ensemble U_{x^*} est convexe. Montrons que, si les contraintes du problème initial $(\mathcal{P}_c)$ sont qualifiées au sens de SLATER, alors les contraintes du problème restreint le sont aussi. Autrement dit, qu'il existe $\tilde{x}_0 \in U$ admissible pour le problème $(\mathcal{P}_c)$ tel que $h_j(\tilde{x}_0) < 0$ pour tout $j \notin \mathcal{J}$ (tel que h_j n'est pas affine) et pour tout $j \notin \mathcal{J}_{x^*}$ (tel que x^* ne sature pas h_j). Si x_0 satisfait ces conditions, il suffit de choisir $\tilde{x}_0 = x_0$. On suppose donc que ce n'est pas le cas. La convexité de l'ensemble admissible $\mathcal{A}$ assure qu'il contient le segment $[\, x_0\,;x^*\,]$ (non vide puisque x^* sature toutes les contraintes, au contraire de x_0. On va pour cela utiliser le lemme suivant :

Lemme 1

Soient $\mathcal{C} \subset \mathcal{X}$ un ensemble convexe et $\phi : \mathcal{C} \to \mathbb{R}$ une fonction convexe. Soient $x_1, x_2 \in \mathcal{C}$. On suppose que $\phi(x_1) > \phi(x_2)$. Alors,

$$\forall\, x \in \,]\, x_1\,; x_2\,]\,, \qquad \phi(x_1) > \phi(x)$$

DÉMONSTRATION : Laissé en exercice.

Soit $j \notin \mathcal{J}$. Puisque $h_j(x^*) = 0$ et que $h_j(x_0) < 0$, on en déduit qu'aucun point de $]\, x^*\,; x_0\,]$ ne sature la contrainte h_j. On a ainsi démontré que les contraintes du problème restreint à l'ouvert U_{x^*} satisfait la condition de SLATER.

On peut donc se contenter de démontrer le théorème 1 sous la condition de SLATER en supposant que x^* **sature toutes les contraintes**. L'idée principale est d'exploiter la convexité des fonctions pour remplacer les contraintes non affines par des contraintes affines, en linéarisant les fonctions associées au voisinage de x^*. Plus précisément, pour tout $j \in [\![\, 1\,;\, q\,]\!]$, on considère la fonction affine

$$\tilde{h}_j : x \mapsto \langle \nabla h_j(x^*), x - x^* \rangle$$

tandis qu'on dédouble les contraintes d'égalité $g_i(x) = 0$ (avec $i \in [\![\, 1\,;\, p\,]\!]$) en introduisant les deux fonctions

$$\tilde{h}_{q+2\,i-1} : x \mapsto \langle \nabla g_i(x^*), x - x^* \rangle \qquad \text{et} \qquad \tilde{h}_{q+2\,i} : x \mapsto -\langle \nabla g_i(x^*), x - x^* \rangle$$

On notera que pour tout $j \in \mathcal{J}$, on a l'identité $h_j = \tilde{h}_j$, tandis que $\tilde{h}_{q+2\,i-1} = g_i$ et $\tilde{h}_{q+2\,i} = -g_i$ pour tout $i \in [\![\, 1\,;\, p\,]\!]$. Pour tout $x \in \mathcal{A}$, on a par convexité

$$0 \geq h_j(x) \geq \langle \nabla h_j(x^*), x - x^* \rangle = \tilde{h}_j(x)$$

On peut maintenant s'intéresser au problème auxiliaire

$$\text{Minimiser } f(x) \text{ sous les contraintes } \tilde{h}_j(x) \leq 0 \text{ pour } j \in [\![\, 1\,;\, q + 2\,p\,]\!] \qquad (\mathcal{P}_{\text{aux}})$$

d'ensemble admissible noté $\mathcal{C}$. D'après ce qui précède, on a $\mathcal{A} \subset \mathcal{C}$.

L'ensemble convexe

$$\mathcal{C} = \left\{ x \in \mathbb{R}^n \mid \forall\, j \in [\![\, 1\,;\, q + 2\,p\,]\!]\,, \quad \langle \nabla \tilde{h}_j(x^*), x - x^* \rangle \leq 0 \right\}$$

est appelé *cône des directions admissibles* en x^*.

On va à présent suivre le raisonnement suivant :

1. on va montrer que si les contraintes du problème $(\mathcal{P}_c)$ sont qualifiées au sens de SLATER, alors les solutions de $(\mathcal{P}_c)$ sont solution du problème $(\mathcal{P}_{\text{aux}})$:
2. à l'aide du lemme de FARKAS–MINKOWSKI (lemme 3, admis), on va démontrer le théorème 1 pour le problème $(\mathcal{P}_{\text{aux}})$;

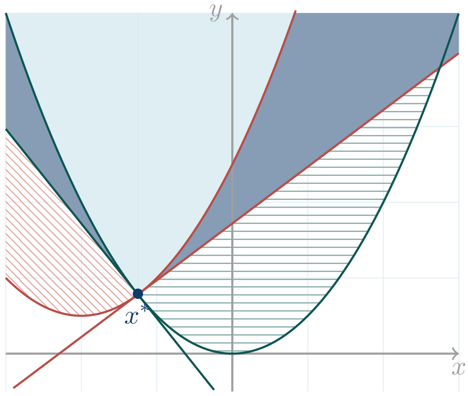

FIGURE 2 – Exemple de cône des directions admissibles en x^*. La contrainte $x^2 \leq y$ est représentée en rouge (graphe de h_1 et de $\tilde{h}_1$). La contrainte $(x+1)^2 + 1/4 \leq y$ est représentée en vert (graphe de h_2 et de $\tilde{h}_2$). L'ensemble admissible $\mathcal{A}$ est la région en bleu clair et le cône $\mathcal{C}$ correspond à l'union de $\mathcal{A}$ et des régions en bleu foncé.

3. enfin, on va traduire ce résultat sur le problème initial $(\mathcal{P}_c)$.

Étape 1. Démontrons que si x^* est solution de $(\mathcal{P}_c)$, alors x^* est solution de $(\mathcal{P}_{\text{aux}})$. Commençons par remarquer que, puisque l'ensemble admissible $\mathcal{C}$ du problème auxiliaire contient $\mathcal{A}$ et a même fonction objectif que $(\mathcal{P}_c)$, toute solution $\hat{x}$ du problème $(\mathcal{P}_{\text{aux}})$ vérifie $f(\hat{x}) \leq f(x^*)$. Il reste donc à démontrer que $f(\hat{x}) = f(x^*)$, d'où l'on déduira que x^* est solution du problème auxiliaire (car admissible). Pour cela, on va raisonner par l'absurde en supposant que $f(\hat{x}) < f(x^*)$ et suivre les étapes suivantes :

1.a. on suppose que $\hat{x}$ ne sature aucune contrainte du problème $(\mathcal{P}_{\text{aux}})$ et on montre que l'intersection entre $]\,x^*\,;\hat{x}\,]$ et $\mathcal{A}$ est non vide ;

1.b. puisque $f(x^*) > f(\hat{x})$, le lemme 1 assure que $f(x) < f(x^*)$ pour tout $x \in \,]\,x^*\,;\hat{x}\,]$;

1.c. on en déduit l'existence d'un point $x \in \mathcal{A} \cap \,]\,x^*\,;\hat{x}\,]$ tel que $f(x) < f(x^*)$, ce qui constitue une contradiction ;

1.d. si $\hat{x}$ sature au moins une contrainte, on utilise la continuité de f sur U pour démontrer l'existence d'un point x_ε dans $\mathcal{C}$ et voisin de $\hat{x}$ tel que x_ε ne sature aucune contrainte $(\mathcal{P}_{\text{aux}})$ et $f(x_\varepsilon) < f(x^*)$;

1.e. on applique les points 1.a. – 1.c. au point x_ε.

L'étape 1.a. est une conséquence du résultat (admis) suivant :

Lemme 2

Soit $\mathcal{U} \subset \mathcal{X}$ un ouvert convexe. Soit $\phi : \mathcal{U} \to \mathbb{R}$ une fonction convexe différentiable en $x^0 \in \mathcal{U}$. Si $p \in \mathcal{X}$ est tel que

$$\forall\, x \in \mathcal{U}, \qquad \phi(x) \geq \phi(x^0) + \langle p, x - x^0 \rangle$$

alors $p = \nabla \phi(x^0)$.

DÉMONSTRATION : Admis. La preuve se trouve dans le module **A2 : Sous-différentiabilité** (5MM71) (Proposition 3).

Une première conséquence de ce lemme est que la fonction affine

$$\tilde{\phi} : x \mapsto \phi(x^0) + \langle p, x - x^0 \rangle$$

(appelée *minorante affine de ϕ exacte en x^0*) définit un hyperplan, frontière de deux demi-espaces, dont l'un contient l'épigraphe de ϕ, c'est-à-dire la région située au-dessus du graphe de ϕ (cf. Figure 3). Une seconde conséquence est que toute autre fonction affine $G \neq \tilde{\phi}$ telle que $G(x^*) = \phi(x^*)$ ne peut minorer ϕ, c'est-à-dire qu'il existe $x \in U$ tel que $G(x) > \phi(x)$. Il s'ensuit que (cf. Figure 3) la fonction G définit deux demi-espaces ouverts, dont les intersections avec l'épigraphe de f sont nécessairement non vides. En particulier, l'intersection entre l'hyperplan $G(x) = 0$ et l'épigraphe de f est non réduit à x^* et contient un segment $[\,x^* \,; x_G\,]$ avec $x_G \neq x^*$.

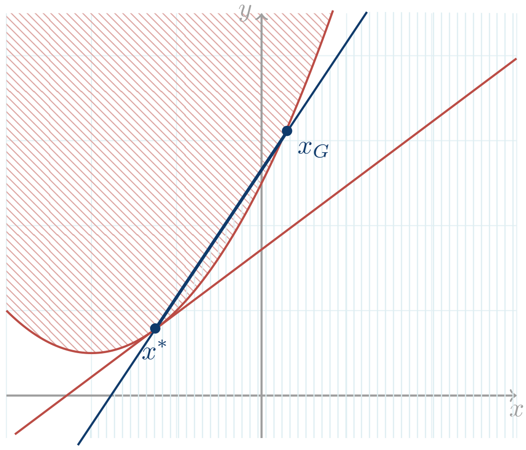

FIGURE 3 – Exemple de minorante affine exacte (en rouge), qui contient $(x^*, \phi(x^*))$. Tout autre hyperplan (en bleu), différent de la minorante affine, passant par $(x^*, \phi(x^*))$, définit deux demi-plans ouverts dont les intersections avec l'épigraphe de ϕ (hachuré rouge) est non vide. Le segment $[\,x^* \,; x_G\,]$ est contenu dans l'épigraphe.

Ainsi, si $\hat{x}$ ne sature aucune contrainte de $(\mathcal{P}_{\text{aux}})$, alors on peut associer à chaque contrainte non affine un point x_j tel que l'intersection entre l'ensemble $h_j(x) \leq 0$ et le segment $[\,x^* \,; \hat{x}\,]$ soit le segment $[\,x^* \,; x_j\,]$ avec $x^* \neq x_j$. Puisqu'il y a un nombre fini de contraintes, l'intersection de ces segments est un segment $[\,x^* \,; \tilde{x}\,]$ d'intérieur non vide. Celui-ci est donc contenu dans l'ensemble admissible $\mathcal{A}$. Les étapes 1.b. et 1.c. suivent immédiatement.

On suppose maintenant que $\hat{x}$ sature au moins une contrainte de $(\mathcal{P}_{\text{aux}})$. Si les contraintes saturées par $\hat{x}$ sont toutes affines, alors on se reporte au cas précédent en ne définissant que des points x_j associé aux contraintes non affines non saturées par $\hat{x}$. Le segment obtenu $[\,x^* \,; \tilde{x}\,]$ est contenu dans $\mathcal{C}$, donc satisfait toutes les contraintes affines de $(\mathcal{P}_c)$, tout en satisfaisant (par construction) les contraintes non affines. La conclusion reste donc valable.

On considère enfin le cas où $\hat{x}$ sature au moins une contrainte non affine. Puisque f est continue sur U, pour tout $0 < \varepsilon < f(x^*) - f(\hat{x})$, il existe $\delta > 0$ tel que

$$\forall\, x \in U, \qquad \|x - \hat{x}\| \leq \delta \qquad \Longrightarrow \qquad |f(x) - f(\hat{x})| \leq \varepsilon$$

En particulier, on en déduit qu'il existe $x_\varepsilon \in [\,x_0 \,; \hat{x}\,[$ tel que $f(x_\varepsilon) < f(x^*)$. Puisque la condition de SLATER est vérifiée, on a $\tilde{h}_j(x_0) < 0$ pour tout $j \notin \mathcal{J}$. Or, $\tilde{h}_j$ est convexe et $\tilde{h}_j(\hat{x}) = 0$, donc le lemme 1 assure qu'aucun point de $[\,x_0 \,; \hat{x}\,[$ ne sature la contrainte $\tilde{h}_j \leq 0$. Autrement dit, $\tilde{h}_j(x_\varepsilon)$ pour tout $j \notin \mathcal{J}$. On peut donc appliquer le raisonnement précédent en remplaçant $\hat{x}$ par x_ε.

On a donc finalement démontré que x^* est solution du problème $(\mathcal{P}_{\text{aux}})$.

Étape 2. On va à présent démontrer le théorème 1 pour le problème ($\mathcal{P}_{\text{aux}}$). On rappelle que x est admissible pour ce problème si et seulement si

$$\forall\, j \in [\![1\,;\, q+2\,p]\!]\,, \qquad \langle \nabla \tilde{h}_j(x^*), x - x^* \rangle \leq 0 \qquad \text{soit} \qquad \langle (-\nabla \tilde{h}_j(x^*)), x - x^* \rangle \geq 0$$

Par ailleurs, puisque x^* est solution de ($\mathcal{P}_{\text{aux}}$) et que f est convexe, on en déduit en dérivant la fonction $t \mapsto f(t\,x + (1-t)\,x^*)$, définie et croissante sur $[\,0\,;1\,]$, que

$$\forall\, x \in \mathcal{C}, \qquad \langle \nabla f(x^*), x - x^* \rangle \geq 0$$

On peut alors appliquer le lemme (admis) suivant :

Lemme 3 (Farkas–Minkowski)

Soit $\mathcal{X}$ un espace de Hilbert. Soient $n \in \mathbb{N}$, $(a_j)_{1\leq j \leq n} \in \mathcal{X}^n$ et $b \in \mathcal{X}$ tels que pour tout $v \in \mathcal{X}$, on a l'implication suivante :

$$\big(\forall\, j \in [\![1\,;\, n]\!]\,, \quad \langle a_j, v \rangle \geq 0\big) \qquad \Longrightarrow \qquad \langle b, v \rangle \geq 0$$

Alors il existe un vecteur $(\tilde{\lambda}_j)_{1\leq j\leq n} \in (\mathbb{R}^+)^n$ tel que

$$b = \sum_{j=1}^{n} \tilde{\lambda}_j\, a_j$$

Démonstration : Admis. La réciproque est vraie.

En appliquant ce lemme, on en déduit l'existence de réels positifs $\tilde{\lambda}$ tels que

$$\nabla f(x^*) = -\sum_{j=1}^{q+2p} \tilde{\lambda}_j \nabla \tilde{h}_j(x^*)$$

Il suffit de remplacer $\tilde{h}_j$ par leur définition pour retrouver le théorème 1.

3.3 Condition suffisante d'optimalité

Lorsque le problème est convexe, les conditions KKT sont des conditions suffisantes.

Théorème 2 (CNS du premier ordre dans le cas convexe)

Soit $U \subset E$ un ouvert. Soit $f : U \to \mathbb{R}$ et $h_j : U \to \mathbb{R}$, $j \in [\![1\,;\, q]\!]$ des fonctions **convexes** différentiables et $g_i : U \to \mathbb{R}$, $i \in [\![1\,;\, p]\!]$ des fonctions **affines**. On considère le problème d'optimisation sous contraintes suivant

$$\text{Minimiser } f(x) \text{ sous les contraintes } \begin{cases} g_i(x) = 0 & \text{pour } i \in [\![1\,;\, p]\!] \\ h_j(x) \leq 0 & \text{pour } j \in [\![1\,;\, q]\!] \end{cases} \qquad (\mathcal{P}_c)$$

Soit $x^* \in U$ tel que

$$\forall\, i \in [\![1\,;\, p]\!]\,, \quad g_i(x^*) = 0 \qquad \text{et} \qquad \forall\, j \in [\![1\,;\, q]\!]\,, \quad h_j(x^*) \leq 0$$

et tel qu'il existe $p+q$ réels $\lambda_1, \ldots, \lambda_p, \mu_1, \ldots, \mu_q$ tels que

$$\nabla f(x^*) + \sum_{i=1}^{p} \lambda_i \nabla g_i(x^*) + \sum_{j=1}^{q} \mu_j \nabla h_j(x^*) = 0$$

avec $\qquad \forall\, j \in [\![1\,;\, q]\!]\,, \quad \mu_j\, h_j(x^*) = 0 \quad \text{et} \quad \mu_j \geq 0$

Alors x^* est solution de ($\mathcal{P}_c$).

Autrement dit, si le problème est convexe et qu'il existe un point admissible x^* pour lesquels on peut associer des multiplicateurs de LAGRANGE vérifiant les conditions KKT, alors ce point est solution du problème considéré. Il suffit donc de résoudre le système de conditions KKT pour trouver les solutions du problème. Ce théorème peut donc être vu comme la réciproque du théorème 1.

DÉMONSTRATION : Soit x un point admissible. Par convexité de f, on a

$$\begin{aligned} f(x) &\geq f(x^*) + \langle \nabla f(x^*), x - x^* \rangle \\ &\geq f(x^*) - \sum_{i=1}^{p} \lambda_i \langle \nabla g_i(x^*), x - x^* \rangle - \sum_{j=1}^{q} \mu_j \langle \nabla h_j(x^*), x - x^* \rangle \end{aligned}$$

Puisque x est admissible et que les h_j sont convexes, on a pour tout $j \in [\![1\,;\, q]\!]$

$$0 \geq \mu_j\, h_j(x) \geq \mu_j\, h_j(x^*) + \langle \mu_j \nabla h_j(x^*), x - x^* \rangle = \langle \mu_j \nabla h_j(x^*), x - x^* \rangle$$

Par ailleurs, puisque les g_i sont affines, on a

$$0 = g_i(x) = g_i(x^*) + \langle \nabla g_i(x^*), x - x^* \rangle = \langle \nabla g_i(x^*), x - x^* \rangle$$

On en déduit que

$$\sum_{i=1}^{p} \lambda_i \langle \nabla g_i(x^*), x - x^* \rangle + \sum_{j=1}^{q} \mu_j \langle \nabla h_j(x^*), x - x^* \rangle \leq 0$$

Il s'ensuit que, pour tout point admissible $x \in U$, on a

$$f(x) \geq f(x^*)$$

Autrement dit, x^* est une solution du problème $(\mathcal{P}_c)$. ■

> On voit que la condition de qualification des contraintes n'apparaît pas dans le théorème précédent. C'est logique, puisque la qualification est une condition introduite pour assurer l'existence des multiplicateurs de LAGRANGE, et que, dans les hypothèses du théorème considéré, ceux-ci sont déjà supposés existants.

En pratique, si l'on souhaite résoudre un problème d'optimisation sous contraintes **convexe**, on suit la procédure suivante :

1. on vérifie que les conditions sont qualifiées (au sens de SLATER ou de l'indépendance linéaire par exemple) ;
2. si c'est le cas, l'existence des multiplicateurs de LAGRANGE est assurée ; on cherche donc un ou les points vérifiant les conditions KKT. Ces points sont solutions du problème considéré.

Si le problème n'est pas convexe, on peut suivre la même procédure (mais en se restreignant à la condition de qualification IL), mais il est possible qu'elle fournisse des points non solutions. Cependant, comme toute solution vérifie nécessairement les conditions KKT, il suffit de comparer leur valeur par la fonction objectif ; le ou les points de valeur minimale sont solutions (si le problème admet une solution !).

Enfin, si les conditions ne sont pas qualifiées au sens SLATER ou de l'indépendance linéaire, il est toujours possible de chercher si des points vérifient les conditions KKT. Les conditions de qualifications suffisent à assurer leur existence, mais ne sont pas nécessaires ; par ailleurs, il est possible que les contraintes soient qualifiées suivant d'autres conditions de qualification, auquel cas les multiplicateurs existent bien.