Mini-cours d'optimisation

G. Vial

2 décembre 2008

Résumé

Le but de ces quelques pages est de rassembler, de manière synthétique, les principaux résultats généraux d'optimisation en dimension finie (avec ou sans contrainte). Il ne s'agit pas de donner des preuves complètes des résultats dans le cadre le plus général, mais plutôt des idées de démonstration dans des cas particuliers, ainsi que des interprétations géométriques des résultats. Ce texte s'adresse à un large public, de formation en mathématiques ou dans d'autres sciences. Pour une présentation plus détaillée et des preuves complètes, on renvoie à la bibliographie [1, 2, 3, 4].

1 Quelques rappels de calcul différentiel

On rappelle ici les formules de Taylor à l'ordre 1 et 2 pour une fonction $f : \mathbb{R}^d \to \mathbb{R}$.

◇ **Ordre 1 :** si f est de classe $\mathscr{C}^1$, alors

$$f(x+h) = f(x) + \langle \nabla f(x), h \rangle + o(\|h\|).$$

◇ **Ordre 2 :** si f est de classe $\mathscr{C}^2$, alors

$$f(x+h) = f(x) + \langle \nabla f(x), h \rangle + \frac{1}{2}\langle \nabla^2 f(x)h, h \rangle + o(\|h\|^2).$$

Remarque 1 Ces formules sont valides dès que le segment reliant x à $x+h$ est contenu dans l'ensemble de définition de la fonction f. Il est d'ailleurs intéressant de remarquer que la fonction d'une variable réelle $t \mapsto f(x+th)$ a pour dérivée en $t = 0$ la quantité $\langle \nabla f(x), h \rangle$ et sa dérivée seconde en $t = 0$ s'écrit $\langle \nabla^2 f(x)h, h \rangle$. Ce constat permet de ramener l'étude des problèmes multi-dimensionnels à des problèmes mono-dimensionnels via les dérivées directionnelles. Il faut cependant garder à l'esprit que la connaissance de f dans chaque direction n'est parfois pas suffisante pour comprendre pleinement son comportement (on se convaincra par exemple que la fonction $(x,y) \mapsto |xy|$ est convexe dans chaque direction autour du point $(0,0)$ mais qu'elle n'est convexe dans aucun voisinage de ce point).

Remarque 2 Le *vecteur gradient* $\nabla f(x) \in \mathbb{R}^d$ est défini comme le vecteur des dérivées partielles. La *matrice hessienne* est la matrice symétrique de taille $d \times d$ des dérivées secondes :

$$(1) \qquad \nabla f(x) = \left(\frac{\partial f}{\partial x_i}(x)\right)_{1 \le i \le d} \quad \text{et} \quad \nabla^2 f(x) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(x)\right)_{1 \le i,j \le d}.$$

En fait, la notion de gradient n'est pas intrinsèque, elle dépend du produit scalaire choisi : la définition générale de $\nabla f(x)$ résulte du théorème de représentation de Riesz appliqué à la différentielle de f en x. Toutefois, en dimension finie, on fixe le plus souvent le produit scalaire canonique et les formules (1) définissent le gradient et la hessienne tout aussi bien. Il en va de même pour la hessienne.

2 Remarques sur l'existence et l'unicité

On considère le problème de minimisation suivant :

$$(*) \qquad \text{Trouver } x^* \text{ dans } A \text{ tel que } f(x^*) = \min_{x \in A} f(x),$$

où $A \subset \mathbb{R}^d$ est un sous-ensemble de $\mathbb{R}^d$ et $f : A \to \mathbb{R}$ une fonction continue. La question de l'existence et de l'unicité d'un minimiseur x^* pour le problème $(*)$ est délicate. On peut retenir comme principe général que la compacité fournit des résultats d'existence, et la convexité un cadre favorable pour l'unicité.

Théorème 1 *Si f est continue et A est compact, alors le problème $(*)$ admet au moins une solution.*

Preuve. On considère (x_n) une suite minimisante, i.e. $x_n \in A$ et $f(x_n)$ converge vers la borne inférieure $\inf_A f$ de f sur A (éventuellement égale à $-\infty$). Par compacité, on peut supposer que (x_n) est convergente, quitte à extraire ; notons x^* la limite. Par continuité de f, il vient que $\inf_A f = f(x^*) = \min_A f$. ∎

Théorème 2 *Si f est continue et coercive, i.e. infinie à l'infini :*

$$\lim_{\|x\| \to \infty} f(x) = +\infty,$$

alors le problème $(*)$ admet au moins une solution.

Preuve. Soit $x_0 \in A$ et $m = f(x_0)$. Par coercivité de f, il existe une boule B centrée en x_0 en dehors de laquelle f est strictement supérieure à m. Ainsi les bornes inférieures de f sur A et sur $A \cap B$ sont les mêmes. Le résultat précédent appliqué à $A \cap B$, compact, permet de conclure. ∎

Théorème 3 *Si A est convexe et f strictement convexe, alors le problème $(*)$ admet au plus une solution.*

Preuve. Supposons que x_1^* et x_2^* soient deux solutions distinctes, alors $x^* = (x_1^* + x_2^*)/2 \in A$ et, par stricte convexité de f,

$$f(x^*) < \frac{f(x_1^*) + f(x_2^*)}{2} = \min_A f,$$

qui fournit une contradiction. ∎

3 Optimisation sans contrainte

Dans ce paragraphe, on suppose que l'ensemble des contraintes A est $\mathbb{R}^d$ tout entier. On dispose alors de conditions nécessaires d'optimalité locale :

Théorème 4 (Équation d'Euler) *Soit $f \in \mathscr{C}^1(\mathbb{R}^d, \mathbb{R})$ et $x^* \in \mathbb{R}^d$. Si f admet en x^* un minimum local sur $\mathbb{R}^d$, alors*

$$\nabla f(x^*) = 0. \tag{2}$$

L'équation précédente est appelée équation d'Euler *et toute solution de* (2) *est appelée* point critique de f.

Preuve. Fixons $u \in \mathbb{R}^d$ et posons $h = tu$ pour $t > 0$. La formule de Taylor s'écrit à l'ordre 1 autour de x^* :

$$f(x^* + h) = f(x^*) + \langle \nabla f(x^*), h \rangle + o(\|h\|),$$

soit encore

$$\frac{f(x^* + tu) - f(x^*)}{t} = \langle \nabla f(x^*), u \rangle + o(1).$$

Comme x^* est un minimiseur local de f, la quantité précédente est positive pour t petit. En faisant tendre t vers 0, il vient

$$\forall u \in \mathbb{R}^d, \quad \langle \nabla f(x^*), u \rangle \geq 0. \tag{3}$$

Le vecteur u étant quelconque, on en déduit $\nabla f(x^*) = 0$. ■

Remarque 3 L'inégalité (3) est valable pour tout u car il est permis d'effectuer des variations autour de x^* *dans toutes les directions*. On verra que cette inégalité est encore valable en présence de contraintes, mais seulement pour les *directions admissibles*.

Notons que l'équation d'Euler est un résultat local, qui ne distingue pas les minima locaux du minimum global. Par ailleurs, minima et maxima ne sont pas discriminés (ni des points cols/selles ou des points d'inflexion). Toutefois, on a la condition d'ordre 2 suivante :

Théorème 5 (Condition nécessaire d'ordre 2) *Soit $f \in \mathscr{C}^2(\mathbb{R}^d, \mathbb{R})$ et $x^* \in \mathbb{R}^d$. Si f admet en x^* un minimum local sur $\mathbb{R}^d$, alors*

$$\nabla^2 f(x^*) \textit{ est semi-définie positive.} \tag{4}$$

Preuve. On reprend la preuve du théorème précédent. On a déjà $\nabla f(x^*) = 0$ et la formule de Taylor à l'ordre 2 permet d'écrire pour toute direction $u \in \mathbb{R}^d$,

$$0 \leq \langle \nabla^2 f(x^*) u, u \rangle + o(1),$$

d'où le résultat par passage à la limite $t \to 0$. ■

Les conditions nécessaires précédentes ne fournissent malheureusement pas de condition suffisante, le cas où la hessienne est semi-définie positive sans être définie positive restant indéterminé :

Théorème 6 (Condition suffisante d'optimalité) *Soit $f \in \mathscr{C}^2(\mathbb{R}^d, \mathbb{R})$ et $x^* \in \mathbb{R}^d$. On suppose*

$$\nabla f(x^*) = 0 \quad \textit{et} \quad \nabla^2 f(x^*) \textit{ est définie positive,}$$

alors f admet en x^* un minimum local.

Preuve. La formule de Taylor à l'ordre 2 permet d'écrire

$$f(x^* + h) - f(x^*) = \langle \nabla^2 f(x^*)h, h\rangle + o(\|h\|^2).$$

Or $\langle \nabla^2 f(x^*)h, h\rangle \geq \lambda_{\min}\|h\|^2$ où $\lambda_{\min} > 0$ désigne la plus petite valeur propre de la matrice hessienne. On en déduit que pour $\|h\|$ assez petit, $f(x^* + h) - f(x^*) \geq 0$. ∎

Exemple 1 On considère le problème de la régression polynomiale au sens des moindres carrés. Étant donné un nuage de points $(x_i, y_i)_{1\leq i\leq N}$ et un entier n, on cherche un polynôme p_n de degré inférieur ou égal à n qui l'approche au mieux au sens suivant :

$$\sum_{i=1}^{N} |y_i - p_n(x_i)|^2 \text{ est minimale.}$$

Si l'on recherche le polynôme p_n sous la forme $p_n = a_0 + a_1 X + \cdots + a_n X^n$, alors le problème de minimisation se récrit

$$\min_{a\in\mathbb{R}^{n+1}} \Phi(a) \quad \text{avec} \quad \Phi(a) = \|Va - y\|_2^2, \tag{5}$$

où la matrice de Vandermonde V et les vecteurs a, y sont définis par

$$V = \begin{pmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^n \\ 1 & x_2 & x_2^2 & \cdots & x_2^n \\ \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_N & x_N^2 & \cdots & x_N^n \end{pmatrix} \qquad a = \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ \vdots \\ a_n \end{pmatrix} \qquad y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ \vdots \\ y_N \end{pmatrix}.$$

Le problème (5) est un problème d'optimisation sans contrainte, et relève des résultats des paragraphes précédents.

◇ La fonction Φ est continue et coercive :

$$\Phi(a) = \|Va\|^2 - 2\langle Va, y\rangle + \|y\|^2 \geq \alpha\|a\|^2 - 2\beta\|a\| + \gamma,$$

où α est la plus petite valeur propre de la matrice définie positive[1] $V^\mathsf{T} V$, $\beta = \|V\|\,\|y\|$ et $\gamma = \|y\|^2$. On est donc assuré de l'existence pour le problème (5).

◇ La fonction Φ est de classe $\mathscr{C}^2$ et

$$\nabla\Phi(a) = 2(V^\mathsf{T} Va - V^\mathsf{T} y), \quad \nabla^2\Phi(a) = 2V^\mathsf{T} V.$$

◇ La hessienne de Φ est définie positive pour tout $a \in \mathbb{R}^{n+1}$, donc Φ est strictement convexe sur $\mathbb{R}^{n+1}$, ce qui assure l'unicité pour (5).

◇ L'équation d'Euler s'écrit

$$V^\mathsf{T} Va = V^\mathsf{T} y,$$

qu'on appelle *système des équations normales*. Signalons qu'on peut aussi résoudre le problème (5) à l'aide de la décomposition QR.

[1] La matrice $V^\mathsf{T} V$ est définie positive dès que V est injective, i.e. dès qu'au moins $n+1$ valeurs des x_i sont distinctes. Ce n'est pas une forte restriction quand N est très grand devant n, qui est la situation courante.

4 Optimisation sous contraintes

Afin de pouvoir écrire des conditions d'optimalité, on écrit l'ensemble des contraintes A sous la forme générale d'*égalités-inégalités* :

$$A = \{x \in \mathbb{R}^d \ ; \ g_1(x) = 0, \ldots, g_p(x) = 0 \text{ et } h_1(x) \leq 0, \ldots, h_q(x) \leq 0\},$$

où les fonctions g_i et h_j sont définie de $\mathbb{R}^d$ dans $\mathbb{R}$.

4.1 Contraintes de type égalité

On se place dans ce paragraphe dans le cas $q = 0$ et donc

$$A = \{x \in \mathbb{R}^d \ ; \ g_1(x) = 0, \ldots, g_p(x) = 0\}.$$

Théorème 7 (extréma liés) *Soit $f, g_i : \mathbb{R}^d \to \mathbb{R}$ des fonctions de classe $\mathscr{C}^1$. On suppose que f admet en $x^* \in A$ un minimum local sur A et que*

$$\text{(6)} \qquad \textit{la famille } \nabla g_1(x^*), \ldots, \nabla g_p(x^*) \textit{ est libre.}$$

Alors il existe des multiplicateurs de Lagrange $\lambda_1, \ldots, \lambda_p \in \mathbb{R}$ *tels que*

$$\text{(7)} \qquad \nabla f(x^*) + \sum_{i=1}^{p} \lambda_i \nabla g_i(x^*) = 0.$$

L'équation (7) *est appelée* équation d'Euler-Lagrange*, la condition* (6) condition de qualification des contraintes*.*

Preuve. On présente ici une preuve dans le cas $p = 1$ et $d = 2$, pour en limiter la technicité, mais aucune difficulté conceptuelle supplémentaire n'apparaît pour le cas général. On note donc

$$A = \{x \in \mathbb{R}^d \ ; \ g(x) = 0\}.$$

L'idée de la preuve est de paramétrer l'ensemble A au voisinage de x^* et d'écrire l'équation d'Euler en le paramètre (non contraint !). La condition (6) indique ici que le vecteur $\nabla g(x^*)$ est non-nul. Sans perte de généralité, on supposera que sa seconde composante est non-nulle :

$$\frac{\partial g}{\partial x_2}(x^*) \neq 0.$$

Le théorème des fonctions implicites permet d'écrire A comme un graphe dans une petite boule centrée en x^* :

$$A \cap B(x^*, \varepsilon) = \{x = (x_1, x_2) \in B(x^*, \varepsilon), \ x_2 = \alpha(x_1)\}.$$

La fonction $\varphi(x_1) = f(x_1, \alpha(x_1))$ admet ainsi un minimum relatif autour de x_1^*. N'ayant plus de contraintes sur x_1, on peut écrire l'équation d'Euler (2) :

$$\frac{\partial f}{\partial x_1}(x_1^*, \alpha(x_1^*)) + \frac{\partial f}{\partial x_2}(x_1^*, \alpha(x_1^*))\, \alpha'(x_1^*) = 0$$

Par ailleurs, on a $g(x_1, \alpha(x_1)) \equiv 0$ d'où, par dérivation,

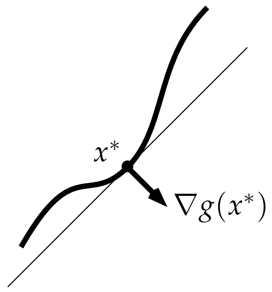
$$\frac{\partial g}{\partial x_1}(x_1^*, \alpha(x_1^*)) + \frac{\partial g}{\partial x_2}(x_1^*, \alpha(x_1^*))\, \alpha'(x_1^*) = 0$$

En remarquant que $\alpha(x_1^*) = x_2^*$ et en posant

$$\lambda = -\left[\frac{\partial g}{\partial x_2}(x^*)\right]^{-1} \frac{\partial f}{\partial x_2}(x^*),$$

on obtient immédiatement $\nabla f(x^*) + \lambda \nabla g(x^*) = 0$. ■

Remarque 4 On peut interpréter géométriquement le résultat précédent :



Le vecteur $\nabla g(x^*)$ – quand il est non-nul – fournit la direction normale à la courbe A (sous-variété de dimension $d-p$ dans le cas général) en x^*.
Par ailleurs, la dérivée par rapport à x_1 de la fonction φ introduite dans la démonstration précédente s'interprète comme la *dérivée tangentielle* de f en x^*. Ainsi, le fait que la dérivée tangentielle soit nulle implique que le gradient est parallèle à la normale, ce qu'exprime exactement l'équation d'Euler-Lagrange.

Exemple 2 Soit $A \in \mathbb{R}^{d\times d}$ une matrice symétrique. On considère le problème de minimisation sous contrainte :

$$\min_{\|x\|_2=1} \langle Ax, x\rangle. \tag{8}$$

Il entre dans le cadre précédent avec

$$f(x) = \langle Ax, x\rangle \quad \text{et} \quad g(x) = \|x\|^2 - 1.$$

Bien sûr, les fonctions f et g sont de classe $\mathscr{C}^1$ et

$$\nabla f(x) = (A + A^\mathsf{T})x = 2Ax \quad \text{et} \quad \nabla g(x) = 2x.$$

Le problème (8) admet au moins une solution par compacité, notée x^*. Par ailleurs, la condition de qualification des contraintes (6) est trivialement vérifiée. On peut donc écrire l'équation d'Euler-Lagrange en x^* : il existe $\lambda \in \mathbb{R}$ tel que

$$2Ax^* + 2\lambda x^* = 0,$$

soit encore $Ax^* = -\lambda x^*$. On a ainsi montré l'existence d'un couple propre pour toute matrice symétrique[2].

4.2 Contraintes de type inégalité

On ne considère ici que des contraintes de type inégalité : $p = 0$ et

$$A = \{x \in \mathbb{R}^d \ ; \ h_1(x) \leq 0, \ldots, h_q(x) \leq 0\}.$$

[2]Cela constitue le point de départ de la preuve du théorème spectral : il suffit de procéder par récurrence sur la dimension d en travaillant dans l'orthogonal de x^*, stable par A car A est symétrique !

Théorème 8 (Karush-Kuhn-Tucker) *Soit $f, h_j : \mathbb{R}^d \to \mathbb{R}$ des fonctions de classe $\mathscr{C}^1$. On suppose que f admet en $x^* \in A$ un minimum local sur A et qu'il existe un vecteur $z \in \mathbb{R}^d$ tel que*

$$\forall j \quad \Big(h_j(x^*) = 0 \Longrightarrow \langle \nabla h_j(x^*), z\rangle < 0\Big). \tag{9}$$

Alors il existe des multiplicateurs de Kuhn et Tucker $\mu_1, \dots, \mu_q \in \mathbb{R}$ *tels que*

$$\nabla f(x^*) + \sum_{j=1}^{q} \mu_j \nabla h_j(x^*) = 0, \tag{10}$$

avec $\mu_j \geq 0$ (positivité) et $\mu_j h_j(x^*) = 0$ (relations d'exclusion).

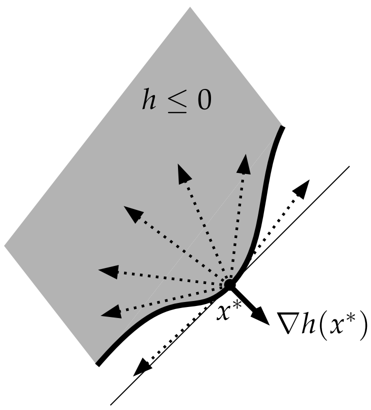
L'équation (10) *est appelée* condition KKT, *la condition* (9) condition de qualification des contraintes*.*

Preuve. Ici encore, on se place dans le cas particulier d'une seule contrainte en dimension 2 :

$$A = \{x \in \mathbb{R}^d \ ; \ h(x) \leq 0\}.$$

Si $h(x^*) < 0$, alors le point minimiseur x^* est situé à l'intérieur de l'ensemble A, si bien qu'on peut effectuer des petites variations autour de x^* dans toutes les directions, et l'équation d'Euler (2) est encore valable. On vient d'exprimer la relation d'exclusion $\mu h(x^*) = 0$.

On est donc ramené au cas $h(x^*) = 0$. Signalons que l'utilisation du théorème des extréma liés est possible car x^* minimise *a fortiori* f sur l'ensemble frontière ∂A défini par $h(x) = 0$. On obtient bien la condition (10), mais pas la positivité des multiplicateurs.



En pointillés sont représentées les directions *admissibles*, i.e. les vecteurs u pour lesquels $x^* + tu$ est dans l'ensemble A quand $t > 0$ est proche de 0. Dans le cas où $\nabla h(x^*) \neq 0$, ce dernier fournit la direction normale extérieure en x^* et on peut caractériser les directions admissibles comme les vecteurs $u \in \mathbb{R}^d$ tels que

$$\langle \nabla h(x^*), u\rangle \leq 0.$$

La même méthode que dans le cas non contraint permet d'écrire que pour de tels u, on a l'inégalité (3) :

$$\langle \nabla f(x^*), u\rangle \geq 0.$$

On obtient donc le résultat suivant :

$$\forall u \in \mathbb{R}^d \quad \langle \nabla h(x^*), u\rangle \leq 0 \Longrightarrow \langle \nabla f(x^*), u\rangle \geq 0, \tag{11}$$

d'où on déduit facilement qu'il existe $\mu \geq 0$ tel que $\nabla f(x^*) = -\mu \nabla h(x^*)$. ■

Remarque 5 Contrairement au cas des contraintes de type égalité, la généralisation de cette démonstration dans le cas $p > 1$ et $d > 2$ est plus délicate. Le demi-plan tangent est remplacé par un cône tangent et, de la même manière, la condition de qualification des contraintes traduit le fait qu'il soit non vide. Par ailleurs, la relation (11) devient :

$$\forall u \in \mathbb{R}^d \quad \Big(\forall j \quad \langle \nabla h_j(x^*), u\rangle \leq 0\Big) \Longrightarrow \Big(\langle \nabla f(x^*), u\rangle \geq 0\Big).$$

On peut encore en déduire qu'il existe des réels positifs μ_j tels que

$$\nabla f(x^*) = \sum_{j=1}^{q} \mu_j \nabla h_j(x^*).$$

C'est le lemme de Farkas-Minkowski et la preuve en est non-triviale, voir [3, 1].

Remarque 6 On introduit souvent le *Lagrangien* du problème ; il s'agit ici de la fonction

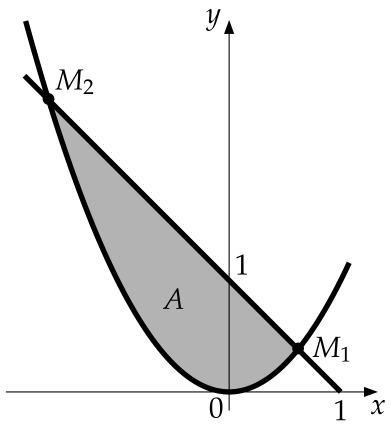
$$\mathscr{L}(x, \mu_1, \dots, \mu_q) = f(x) + \sum_{j=1}^{q} \mu_j h_j(x).$$

La condition (10) traduit le fait qu'il existe des multiplicateurs μ_j tels que x^* soit un point critique du Lagrangien :

$$\nabla_x \mathscr{L}(x, \mu_1, \dots, \mu_q) = 0.$$

Exemple 3 Soit à maximiser x sur le sous-ensemble A de $\mathbb{R}^2$:

$$A = \left\{ (x,y) \in \mathbb{R}^2 \quad ; \quad y \geq x^2 \quad \text{et} \quad x + y \leq 1 \right\}.$$



Le Lagrangien du problème s'écrit

$$\mathscr{L}(x, \mu_1, \mu_2) = f(x,y) + \mu_1 h_1(x,y) + \mu_2 h_2(x,y),$$

avec

$$f(x,y) = -x, \quad h_1(x,y) = x^2 - y \quad \text{et} \quad h_2(x,y) = x + y - 1.$$

(Noter la définition $f(x,y) = -x$ car il s'agit d'un problème de *maximisation*).

Étude de la condition de qualification des contraintes : les gradients des fonctions de contraintes s'écrivent

$$\nabla h_1(x,y) = (2x, -1)^{\mathsf{T}} \quad \text{et} \quad \nabla h_2(x,y) = (1,1)^{\mathsf{T}}.$$

Ces vecteurs ne s'annulent jamais, il reste donc seulement à vérifier qu'ils ne sont pas opposés en les sommets M_1 et M_2, ce qui est effectivement le cas. En conséquence, la condition de qualification des contraintes est partout vérifiée[3].

[3] Ici, l'ensemble A est convexe et la condition de qualification est équivalente à la non vacuité de A, cf. [4].

Conditions de Kuhn et Tucker : les contraintes étant qualifiées, tout point de minimum est point critique du Lagrangien et les conditions KKT s'écrivent

$$\begin{cases} -1+2x\mu_1+\mu_2 = 0 \\ -\mu_1+\mu_2 = 0 \end{cases} \quad \text{avec} \quad \begin{cases} \mu_1(x^2-y) = 0, \\ \mu_2(x+y-1) = 0. \end{cases}$$

S'ensuit une étude de cas selon que $\mu_1 = 0$ ou $\mu_1 \neq 0$. On en déduit que la seule solution est $(x^*, y^*) = M_1$ avec $\mu_1 = \mu_2 = 1/(1+2x^*)$.
L'existence d'une solution au problème de maximisation, par compacité, permet d'assurer que le point critique M_1 en est bien l'unique solution (ce qui est évident "à la main").

5 Contraintes et méthodes numériques

Si la présence de contraintes complique l'étude théorique des problèmes d'optimisation, elle pose aussi des difficultés au niveau numérique. Par exemple, les méthodes de gradient ne peuvent être utilisées telles quelles car une itération

$$x_{n+1} = x_n - \rho \nabla f(x_n)$$

n'est pas assurée de rester dans l'ensemble des contraintes. Une solution consiste à projeter x_{n+1} sur l'ensemble des contraintes ; on parle de méthode de *gradient projeté*. Toutefois, cet algorithme est difficilement utilisable en pratique car – en général – on ne sait pas calculer la projection.

Dans le cas de contraintes "égalités", il est possible de déterminer les points critiques à l'aide des équations d'Euler-Lagrange. En effet, si l'ensemble des contraintes est

$$A = \{x \in \mathbb{R}^d \; ; \; g(x) = 0\},$$

alors les conditions d'optimalité s'écrivent

$$\begin{cases} \nabla f(x) + \lambda \nabla g(x) & = & 0, \\ g(x) & = & 0. \end{cases}$$

Il s'agit d'un système de $d+1$ équations (en général non-linéaires) dont l'inconnue est le couple $(x, \lambda) \in \mathbb{R}^{d+1}$. On peut, par exemple, utiliser une méthode de Newton pour déterminer (x^*, λ^*) numériquement.

Citons enfin une méthode très utilisée dans les applications, qui ramène un problème sous contraintes à un problème sans contrainte. Il s'agit de la méthode de *pénalisation*. Toujours dans l'exemple d'une contrainte de type "égalité", on pose pour $\varepsilon > 0$,

$$f_\varepsilon(x) = f(x) + \varepsilon^{-1} \left[g(x)\right]^2.$$

La méthode s'appuie sur le constat suivant : si x_ε^* minimise f_ε sur $\mathbb{R}^d$ tout-entier, alors $g(x_\varepsilon^*)$ doit être proche de 0. En effet, un point $x \in \mathbb{R}^d$ tel que $g(x)$ est "grand" n'a que peu de chance de fournir un minimiseur de f_ε car le terme $\varepsilon^{-1} \left[g(x)\right]^2$ – très grand – le *pénalise*. Plus précisément, sous certaines hypothèses sur f et g, on peut montrer que x_ε^* tend vers x^* lorsque ε tend vers 0.

Quelques mathématiciens de l'optimisation

Leonhard Euler [1707-1783] mathématicien et physicien suisse.

Joseph-Louis Lagrange [1736-1813] mathématicien français.

William Karush [1917-1997] mathématicien américain.

Harold W. Kuhn [1925-] mathématicien et économiste américain.

Albert W. Tucker [1905-1995] mathématicien américano-canadien.

Gyula Farkas [1847-1930] mathématicien et physicien hongrois.

Hermann Minkowski [1864-1909] mathématicien et physicien allemand.

Références

[1] G. ALLAIRE. *Analyse numérique et optimisation*. Éditions de l'École Polytechnique. Ellipses, Paris 2005.

[2] V. BECK, J. MALICK, G. PEYRÉ. *Objectif Agrégation*. H&K 2004.

[3] P. G. CIARLET. *Introduction à l'analyse numérique matricielle et à l'optimisation*. Collection Mathématiques Appliquées pour la Maîtrise. [Collection of Applied Mathematics for the Master's Degree]. Masson, Paris 1982.

[4] J.-B. HIRIART-URRUTY. *L'optimisation*. Que sais-je ? PUF, Paris 1996.