.

Approche de la méthode du Lagrangien

Mise en place du contexte

On souhaite optimiser une fonction φ définie sur un domaine $\mathcal{D} \subset \mathbb{R}^n$ sous les contraintes *égalitaires* définies par des fonctions $g_1, g_2, \cdots, g_p$.

Si l'on note $\mathcal{C}$ le domaine contraint, alors nous cherchons x_0 tel que :

$$\varphi(x_0) = \max_{x \in \mathcal{C} \cap \mathcal{D}} \varphi(x) \qquad \text{ou} \qquad \varphi(x_0) = \min_{x \in \mathcal{C} \cap \mathcal{D}} \varphi(x)$$

Nous souhaitons expliquer en quoi l'étude du Lagrangien défini en cours sous la forme $L(x;\lambda) = \varphi(x) - \langle \lambda \ ; \ (g_i(x))_{i \leq p} \rangle$ fournit une réponse au problème, et comment on en est amené à poser sa définition.

Etude des contraintes

On remarque immédiatement qu'écrire $h(x_1 \ldots x_n) = t(x_1 \ldots x_n)$ revient à demander $(h - t)(x_1 \ldots x_n) = 0$ et qu'ainsi, toute contrainte égalitaire s'écrit sous la forme $g(x) = 0$ avec g définie sur $\mathbb{R}^n$ et $x = (x_1 \ldots x_n)$.
Par ailleurs, si une contrainte $\psi(x) = 0$ s'écrit avec $\psi : \mathbb{R}^n \longrightarrow \mathbb{R}^k$ pour $k > 1$ alors on peut décomposer ψ de la façon suivante :

$$\psi(x_1 \ldots x_n) = (g_1(x_1 \ldots x_n) \ ; \ \ldots \ ; \ g_k(x_1 \ldots x_n))$$

où $g_i : \mathbb{R}^n \longrightarrow \mathbb{R}$ et ainsi, toute famille de contraintes égalitaires se ramène à l'écriture d'égalités du type $g_i(x) = 0$ avec g_i définies sur $\mathbb{R}^n$ et à valeurs dans $\mathbb{R}$.

Dans toute la suite, on pourra donc considérer que l'ensemble *contraint* des contraintes égalitaires, noté $\mathcal{C}$, se définira toujours par :

$$\mathcal{C} : \psi(x) = O \ \Leftrightarrow \ \begin{cases} g_1(x) = 0 \\ g_2(x) = 0 \\ \vdots \\ g_p(x) = 0 \end{cases}$$

avec p le nombre de contraintes et $g_i : \mathbb{R}^n \longrightarrow \mathbb{R}$ et $x = (x_1 \ldots x_n)$. Du point de vue ensembliste, on a donc :

$$\mathcal{C} = \{x \in \mathbb{R}^n \mid g_1(x) = 0 \ ; \ \ldots \ ; \ g_2(x) = 0 \ ; \ \ldots \ ; \ g_p(x) = 0\}$$

et si l'on écrit $\psi = (g_1 \ldots g_p)$ ceci revient à :

$$\mathcal{C} = \{x \in \mathbb{R}^n \mid \psi(x) = O\} = \psi^{-1}(\{O\})$$

Dans toute la suite, on considérera donc que $\psi : \mathbb{R}^n \longrightarrow \mathbb{R}^p$ avec $\psi = (g_i)_{i \leq p}$ sont les contraintes égalitaires associées *en ces termes* au problème d'optimisation sous contrainte $\mathcal{C}$.

Lien contraintes et Lagrangien

Nous définissons donc le Lagrangien d'un problème d'optimisation par :

$$\boxed{L(x;\lambda) := \varphi(x) - \langle \lambda \mid \psi(x) \rangle} = \varphi(x_1 \ldots x_n) - \sum_{k=1}^{p} \lambda_k \cdot g_k(x_1 \ldots x_p)$$

Certains auteurs y préfèrent la somme $\varphi(x) + \langle \lambda \mid \psi(x) \rangle$, les coefficients λ_i trouvés seront les mêmes au signe près. Comme la solution au problème ne s'intéresse pas aux valeurs de ces coefficients, les solutions finales seront inchangées et il n'y pas de raison de s'obliger à faire un choix plutôt qu'un autre.

Nous observons que le vecteur extrait du gradient de L sur $\mathbb{R}^n \times \mathbb{R}^p$, défini par les coordonnées associées à λ est :

$$\left(\frac{\partial}{\partial\lambda_1}L(x;\lambda)\ ;\ \dots\ ;\ \frac{\partial}{\partial\lambda_p}L(x;\lambda)\right) = (-g_1(x)\ ;\ \dots\ ;\ -g_p(x))$$

Et ainsi, si $(x_0;\lambda_0)$ est un point critique de L alors il vient en particulier : $\psi(x_0) = O$ c'est-à-dire que $x_0 \in \mathcal{C}$.

Propriété : *Si le Lagrangien du problème d'optimisation sous contraintes égalitaire admet un gradient, alors tout point critique* $(x_0;\lambda_0)$ *vérifie* $x_0 \in \mathcal{C}$ *et donc satisfait aux contraintes du problème.*

Le problème de la réciproque se pose évidemment : et si l'on a $x_0 \in \mathcal{C}$, peut-on alors affirmer que x_0 est la partie constituée des n premières coordonnées d'un point critique de L ? Eh bien la réponse ne saurait être positive si L n'admet déjà pas de point critique.
Il est clair que :

$$\begin{aligned}\nabla L(x;\lambda) &= \left(\frac{\partial\varphi}{\partial x_1}(x)\ ;\ \dots\ ;\ \frac{\partial\varphi}{\partial x_n}(x)\ ;\ 0\ ;\ \dots\ ;\ 0\right) - \left(\sum_{k=1}^{p}\lambda_k\frac{\partial}{\partial x_1}g_k(x)\ ;\ \dots\ ;\ \sum_{k=1}^{p}\lambda_k\frac{\partial}{\partial x_n}g_k(x)\ ;\ g_1(x)\ ;\dots\ ;\ g_p(x)\right)\\ &= \nabla(\varphi)(x) - \left(\sum_{k=1}^{p}\lambda_k\left(\frac{\partial}{\partial x_1}g_k(x)\ ;\ \dots\ ;\ \frac{\partial}{\partial x_n}g_k(x)\right)\ ;\ g_1(x)\ ;\dots\ ;\ g_p(x)\right)\end{aligned}$$

à condition que L admette bien un gradient, et en complétant par des "0" chaque fois que la dimension de l'objet écrit est inférieure à $n+p$.
Si un point critique existe, il vérifie bien, en particulier, que $(g_1(x)\ ;\dots\ ;\ g_p(x))$ s'annule. En choisissant $x \in \mathcal{C}$, on obtient immédiatement que cette partie du gradient s'annule : on ne saurait compléter, en général, chaque point qui satisfait aux contraintes en un point critique de L, sinon tout point satisfaisant aux contraintes serait critique pour φ... alors que les contraintes sont définies indépendamment de φ !

La réciproque n'est donc pas vérifiée. C'est pourquoi on parle souvent de *condition nécessaire de premier ordre*.

Formes différentielles, différentielles totales

Définition : Pour $f\ :\ \mathbb{R}^n \longrightarrow \mathbb{R}$ admettant des dérivées partielles en tout point $x \in \mathcal{D} \subset \mathbb{R}^n$ ouvert, on appelle *forme différentielle* de f en $a \in \mathcal{D}$ l'application :

$$\begin{aligned} D_f(a) : \mathbb{R}^n &\longrightarrow \mathbb{R}\\ h &\mapsto \langle\nabla(f)(a)\mid h\rangle = \sum_{k=1}^{n}\frac{\partial f}{\partial x_i}(a)\cdot h\end{aligned}$$

Les formes différentielles sont alors clairement des applications linéaires de $\mathbb{R}^n$ dans $\mathbb{R}$ <u>donc</u> des formes linéaires. Lorsqu'elles sont non nulles, elles admettent un noyau $ker(D_f(a))$ qui est un hyperplan de $\mathbb{R}^n$ donc de dimension $n-1$ et ainsi engendré par une famille libre de $n-1$ vecteurs libres.
En faisant varier $a \in \mathbb{R}^n$, on décrit une famille de formes différentielles associées à f que l'on nomme alors *différentielle totale* de f notée $D(f)$ ou aussi df (comme en calcul intégral) soit encore :

$$D(f) = df : (x \mapsto df(x) = D_f(x))$$

On peut retenir que la linéarité permettrait d'aboutir à $D(f)(x) = df(x) = \sum_{k=1}^{n} D_f(x_i)$ lorsque x s'écrit $x = (x_1 \dots x_n)$ dans la base canonique (ce serait fastidieux -et hors programme- de le prouver soigneusement).

Il viendrait :

$$D_\psi(x) = D_{\psi_1}(x) + \dots + D_{\psi_n}(x) = \sum_{k=1}^{n} D_{g_k}(x)$$

par les définitions introduites.

Contexte du problème d'optimisation :

Dans notre contexte, on obtiendrait alors que demander x_0 critique pour φ sous la contrainte $x_0 \in \mathcal{C}$ nous amène donc à demander :

$$ker\,(D_\psi(x_0)) \;\perp\; \nabla(\varphi)(x_0)$$

Commençons par réécrire plus simplement :

$$\begin{aligned}
\nabla L(x;\lambda) &= \nabla(\varphi)(x) - \left(\sum_{k=1}^{p} \lambda_k \left(\frac{\partial}{\partial x_1} g_k(x)\;;\;\dots\;;\;\frac{\partial}{\partial x_n} g_k(x)\right)\;;\; g_1(x)\;;\dots\;;\; g_p(x)\right)\\
&= \nabla(\varphi)(x) - \left(\sum_{k=1}^{p} \lambda_k D_{g_k}(x)\;;\; g_1(x)\;;\dots\;;\; g_p(x)\right)\\
&= \nabla(\varphi)(x) - (D_\psi(x)(\lambda)\;;\;\psi(x))
\end{aligned}$$

où l'on rappelle que $\nabla(\varphi)(x)$ est une contraction abusive d'un vecteur de $\mathbb{R}^{n+k}$ dont les k dernières composantes sont nulles.

Ainsi, si $(x_0;\lambda_0)$ est critique pour L on obtient :

$$\nabla(L)(x_0;\lambda_0) = 0 \;\Leftrightarrow\; \psi(x_0) = 0 \;\text{ et }\; \nabla(\varphi)(x_0) = D_\psi(x_0)(\lambda_0)$$

Mais x_0 *critique pour* φ signifie $\nabla(\varphi)(x_0) = 0$ et donc $D_\psi(x_0)(\lambda_0) = 0$ et ainsi :

$$D_\psi(x_0)(\lambda_0) = \sum_{k=1}^{p} D_{g_k}(x_0)h_k = 0 = \langle \nabla(\psi)(x_0) \mid \lambda_0 \rangle = 0$$

ce qui nous conduit bien à l'orthogonalité mentionnée. De plus, l'équation reliant x_0 et λ_0, la connaissance de λ_0 (donc des coefficients $\lambda_1 \dots \lambda_p$) nous renseigne sur les valeurs de $x_1 \dots x_n$.

A noter que nous ne nous intéresserons pas à la réciproque : en pratique, on se contente d'une condition de premier ordre *suffisante*

Critique mais surtout extremum !

De la même façon que l'on a obtenu pour un point critique x_0 pour φ sous contrainte $\mathcal{C}$ un point critique $(x_0;\lambda_0)$ pour le Lagrangien L, on transmet x_0 point extremal de φ sous contrainte $\mathcal{C}$ en $(x_0;\lambda_0)$ *point extremal de* L en exploitant les conditions de second ordre. Mais il faudra garder l'écritures des contraintes $\psi(x) = 0$.

Ceci nous amène donc à définir une Hessienne de L complétée par les contraintes sous forme matricielle que l'on nomme *Hessienne Bordée* et qui s'écrira :

$$\partial H(x;\lambda) = \begin{pmatrix}
0 & \dots & 0 & \frac{\partial}{\partial x_1} g_1(x) & \dots & \frac{\partial}{\partial x_n} g_1(x)\\
\vdots & \ddots & \vdots & \vdots & & \vdots\\
0 & \dots & 0 & \frac{\partial}{\partial x_1} g_p(x) & \dots & \frac{\partial}{\partial x_n} g_p(x)\\
\frac{\partial}{\partial x_1} g_1(x) & \dots & \frac{\partial}{\partial x_1} g_p(x) & & & \\
\vdots & & \vdots & & \nabla^2(L)(x;\lambda) & \\
\frac{\partial}{\partial x_n} g_1(x) & \dots & \frac{\partial}{\partial x_n} g_p(x) & & &
\end{pmatrix}$$

Réécrite plus simplement avec la notation du jacobien :

$$\partial H(x;\lambda) = \begin{pmatrix} O & Jac_\psi(x) \\ {}^t Jac_\psi(x) & \nabla^2(L)(x;\lambda) \end{pmatrix}$$

et dont on pourra vraiment admettre les propriétés données en cours caractérisant la reconnaissance d'extrema sous contraintes.