RÉSUMÉ DE COURS : RECHERCHE D'EXTREMA

Dans ce chapitre, on va s'intéresser à la recherche des extrema d'une fonction de plusieurs variables. Plus précisément, on va déterminer des conditions pour qu'une fonction de plusieurs variables admette un maximum ou un minimum (local ou global) en un point. Dans le cas d'une fonction f d'une variable, on sait déjà comment trouver les extrema. En effet, dans ce cas, les points éventuels où f peut admettre un extremum sont les réels a tels que $f'(a) = 0$. De plus, pour déterminer si un tel point a correspond effectivement à un maximum ou un minimum local, on s'intéresse au signe de $f''(a)$. Plus précisément, la fonction f admet un maximum (resp. minimum) local en a si $f''(a) < 0$ (resp. $f''(a) > 0$), ce que l'on peut voir à l'aide du développement limité de f à l'ordre 2 en a calculé par exemple via la formule de Taylor-Young. Dans ce qui suit, nous allons voir comment généraliser ces résultats au cas des fonctions de plusieurs variables, à l'aide notamment des notions de gradient et de hessienne, d'abord pour les extrema locaux et globaux en général, puis pour les extrema locaux et globaux de fonctions soumises à des contraintes.

1 Existence d'extrema sur un fermé borné

Définition 1.1 *Soit f une application d'une partie U de $\mathbb{R}^n$ dans $\mathbb{R}$, et soit $a \in U$. On dit que :*

1. *f admet un maximum local en a si : $\exists r > 0,\ \forall x \in U \cap B(a,r),\ f(x) \leq f(a)$.*
2. *f admet un minimum local en a si : $\exists r > 0,\ \forall x \in U \cap B(a,r),\ f(x) \geq f(a)$.*
3. *f admet un extremum local en a si elle admet un maximum ou un minimum local en a.*
4. *f admet un maximum global en a si : $\forall x \in U,\ f(x) \leq f(a)$.*
5. *f admet un minimum global en a si : $\forall x \in U,\ f(x) \geq f(a)$.*
6. *f admet un extremum global en a si elle admet un maximum ou un minimum global en a.*

On dispose alors du théorème suivant qui garantit l'existence d'extrema sous certaines conditions :

Théorème 1.2 *Toute application f continue d'un fermé borné U de $\mathbb{R}^n$ dans $\mathbb{R}$ est bornée et atteint ses bornes. En particulier, f admet un maximum global et un minimum global sur U.*

A noter que ceci n'est ni plus ni moins que la généralisation aux fonctions de $\mathbb{R}^n$ dans $\mathbb{R}$ du résultat bien connu pour les fonctions de $\mathbb{R}$ dans $\mathbb{R}$, à savoir "toute fonction continue sur un segment (de $\mathbb{R}$) est bornée et atteint ses bornes". En d'autres termes, les fermés bornés de $\mathbb{R}^n$ jouent à peu de choses près le même rôle que les segments de $\mathbb{R}$. A noter aussi que ce résultat nous assure de l'existence d'extrema globaux, mais ne nous dit pas comment les trouver. Aux paragraphes suivants, on verra comment les déterminer.

2 Recherche d'extrema sans contrainte

Définition 2.1 *Soit U un ouvert de $\mathbb{R}^n$ et soit $f : U \longrightarrow \mathbb{R}$ une application de classe $\mathcal{C}^1$. On dit qu'un point $a \in U$ est un point critique de f si $\nabla(f)(a) = 0$.*

A noter qu'en un point critique a de f, toutes les dérivées directionnelles de f en a à l'ordre 1 sont nulles (vu qu $f'_u(a) = \langle \nabla(f)(a), u \rangle$ pour tout $u \in \mathbb{R}^n$). A présent, nous allons rappeler une condition nécessaire (du premier ordre) pour l'existence d'un extremum local.

Théorème 2.2 *Soit U un ouvert de $\mathbb{R}^n$, et soit $f : U \longrightarrow \mathbb{R}$ une application de classe $\mathcal{C}^1$. Si f admet un extremum local en un point $a \in U$, alors a est un point critique de f.*

En d'autres termes et au vu de l'expression du gradient, on retiendra que :

$$\boxed{\text{si } f \text{ admet un extremum local en } a, \text{ alors : } \forall i \in \{1, ..., n\},\ \ \partial_i(f)(a) = 0.}$$

A noter que ce résultat n'est ni plus ni moins que la généralisation de la condition "$f'(a) = 0$" pour la recherche d'extrema des fonctions d'une variable. A noter aussi que ce résultat n'est pas une équivalence, mais seulement une condition nécessaire d'existence d'un extremum local. En effet, il existe des fonctions admettant des points critiques qui ne sont pas des extrema locaux. A présent, nous allons voir une condition suffisante (du second ordre) d'existence d'un extremum local.

Théorème 2.3 *Soit U un ouvert de $\mathbb{R}^n$, soit $f : U \longrightarrow \mathbb{R}$ une application de classe $\mathcal{C}^2$, soit $a \in U$ un point critique de f et soit $\nabla^2(f)(a)$ la matrice hessienne de f en a.*

1. *Si toutes les valeurs propres de $\nabla^2(f)(a)$ sont > 0, alors f admet un minimum local en a.*
2. *Si toutes les valeurs propres de $\nabla^2(f)(a)$ sont < 0, alors f admet un maximum local en a.*
3. *Si $\nabla^2(f)(a)$ admet deux valeurs propres $\neq 0$ de signes contraires, alors f n'a pas d'extremum en a.*

De façon générale, pour étudier la nature d'un point critique a, on commencera par calculer la matrice hessienne de f en a. Ensuite, on déterminera les valeurs propres de $\nabla^2(f)(a)$, en calculant par exemple le rang de $\nabla^2(f)(a) - \lambda \mathrm{I}_n$ en fonction de λ, et enfin on conclura à l'aide du théorème 2.3. A noter qu'en termes de formes quadratiques, ce résultat se retraduit de la façon suivante :

Théorème 2.4 *Soit U un ouvert de $\mathbb{R}^n$, soit $f : U \longrightarrow \mathbb{R}$ une application de classe $\mathcal{C}^2$, soit $a \in U$ un point critique de f et soit q_a la forme quadratique associée à la matrice hessienne de f en a.*

1. *Si, pour tout $h \in \mathbb{R}^n \setminus \{0\}$, on a $q_a(h) > 0$, alors f admet un minimum local en a.*
2. *Si, pour tout $h \in \mathbb{R}^n \setminus \{0\}$, on a $q_a(h) < 0$, alors f admet un maximum local en a.*
3. *Si q_a n'est pas de signe constant sur $\mathbb{R}^n$, alors f n'admet pas d'extremum en a.*

On rappelle que la forme quadratique q_a n'est pas de signe constant si elle peut prendre des valeurs > 0 et < 0. Dans ce cas, on dit que a est un *point-selle (ou col)*. A noter que, si $q_a(h) \geq 0$ pour tout $h \in \mathbb{R}^n$ et qu'il existe un vecteur $h_0 \neq 0$ tel que $q_a(h_0) = 0$, alors on ne peut pas conclure quant à la nature du point critique a de f. A noter aussi que le résultat ci-dessus n'est ni plus ni moins que la généralisation des conditions "$f'(a) = 0$" et "$f''(a) > 0$" (resp. $f''(a) < 0$") pour la recherche de minima (resp. maxima) des fonctions d'une variable. A noter enfin qu'une fonction peut très bien présenter un ou plusieurs extrema locaux sans pour autant avoir un maximum ou un minimum global. De façon générale, pour déterminer les extrema (éventuels) d'une fonction f de plusieurs variables sur un ouvert U de $\mathbb{R}^n$, on commencera par chercher les points critiques de f sur U. Puis, on déterminera la nature de ces points critiques à l'aide des théorèmes donnés plus haut. Enfin, si f présente un extremum local en a, on cherchera à montrer en étudiant le signe de $f(x) - f(a)$ que f admet effectivement un extremum global en a (si c'est le cas).

Définition 2.5 *On dit qu'une partie U de $\mathbb{R}^n$ est convexe si, pour tous points $A, B \in U$, le segment $[A, B]$ est inclus dans U.*

En d'autres termes, une partie U de $\mathbb{R}^n$ est convexe si et seulement si, pour tous points $A, B \in U$ et pour tout $t \in [0, 1]$, le point $tA + (1-t)B$ appartient à U. En général, c'est cette condition que l'on vérifiera pour montrer que U est convexe. A titre d'exemple, les ensembles $\mathbb{R}^n$ et $\mathbb{R}_+^*$ sont convexes, alors que $\mathbb{R}^*$ ne l'est pas. Pour le voir, il suffit de remarquer que -1 et 1 appartiennent à $\mathbb{R}^*$ mais que le segment $[-1, 1]$ n'est pas inclus dans $\mathbb{R}^*$. De plus, tout intervalle de $\mathbb{R}$ est convexe et tout produit cartésien d'ensembles convexes est convexe. En particulier, tout produit cartésien d'intervalles est un ensemble convexe.

Théorème 2.6 (Formule de Taylor avec reste intégral à l'ordre 1) *Soit U un ouvert de $\mathbb{R}^n$, soient $a \in U$ et $h \in \mathbb{R}^n$ tels que $[a, a+h] \subset U$, et soit $f : U \longrightarrow \mathbb{R}$ une application de classe $\mathcal{C}^2$ sur U. Alors :*

$$f(a+h) = f(a) + \langle \nabla(f)(a), h \rangle + \int_0^1 q_{a+th}(h)(1-t)dt.$$

A noter que ce résultat se démontre à l'aide de la formule de Taylor avec reste intégral à l'ordre 1 bien connue pour les fonctions de $\mathbb{R}$ dans $\mathbb{R}$, appliquée à la fonction $g : t \longmapsto f(a+th)$. On en déduit le :

Théorème 2.7 *Soit U un ouvert convexe de $\mathbb{R}^n$, soit $f : U \longrightarrow \mathbb{R}$ une application de classe $\mathcal{C}^2$ et soit $a \in U$ un point critique de f. Si la matrice hessienne de f en b n'a que des valeurs propres positives (resp. négatives) pour tout $b \in U$, alors f admet un minimum (resp. maximum) global en a.*

A noter que, comme le signe des valeurs propres d'une matrice symétrique (réelle) est directement lié à celui de la forme quadratique qui lui est associée, le théorème 2.7 peut se reformuler de la façon suivante :

> Soit U un ouvert convexe de $\mathbb{R}^n$, soit $f : U \longrightarrow \mathbb{R}$ une application de classe $\mathcal{C}^2$ et soit $a \in U$ un point critique de f. Si la forme quadratique q_b associée à la matrice hessienne de f en b est positive (resp. négative) pour tout $b \in U$, alors f admet un minimum (resp. maximum) global en a.

3 Recherche d'extrema sous contraintes d'égalités linéaires

Définition 3.1 *Un ensemble de contraintes linéaires est l'ensemble $\mathcal{C} \subset \mathbb{R}^n$ des solutions d'un système :*

$$\begin{cases} g_1(x) & = & b_1 \\ \vdots & & \vdots \\ g_p(x) & = & b_p \end{cases},$$

où $g_i(x) = a_{i,1}x_1 + ... + a_{i,n}x_n$ pour tout i, $x = (x_1, ..., x_n)$ et $a_{i,j}, b_j \in \mathbb{R}$ pour tous i, j. On désigne par $\mathcal{H}$ l'ensemble des solutions du système homogène :

$$\begin{cases} g_1(x) & = & 0 \\ \vdots & & \vdots \\ g_p(x) & = & 0 \end{cases}.$$

En particulier, on voit que $\mathcal{H}$ est l'unique sous-espace vectoriel de $\mathbb{R}^n$ parallèle à $\mathcal{C}$.

Définition 3.2 *Soit f une application d'une partie U de $\mathbb{R}^n$ dans $\mathbb{R}$, soit $\mathcal{C}$ un système de contraintes linéaires et soit $a \in U \cap \mathcal{C}$. On dit que f admet un extremum local (resp. global) en a sous la contrainte $\mathcal{C}$ si la restriction $f_{U\cap\mathcal{C}}$ admet un extremum local (resp. global) en a.*

A noter que d'après la méthode du pivot de Gauss, un ensemble de contraintes linéaires peut toujours être décrit à l'aide d'un nombre fini de paramètres linéairement indépendants. Dès lors, en réinjectant ces paramètres dans l'expression de f, on peut écrire la fonction $f_{U\cap\mathcal{C}}$ à l'aide de moins de variables, et cette écriture fait disparaître du coup les conditions données par l'ensemble de contraintes linéaires, ce qui nous ramène à déterminer les extrema d'une fonction de plusieurs variables sans contrainte. Cependant, on procèdera en général sans résoudre le système de contraintes linéaires, à l'aide des résultats suivants :

Théorème 3.3 *Soit $f : U \longrightarrow \mathbb{R}$ une application de classe $\mathcal{C}^1$ sur un ouvert U de $\mathbb{R}^n$, soit $\mathcal{C}$ un système de contraintes linéaires et soit $a \in U \cap \mathcal{C}$. Si f admet un extremum local en a sous la contrainte $\mathcal{C}$, alors $f'_h(a) = 0$ pour tout $h \in \mathcal{H}$.*

Théorème 3.4 *Soit $f : U \longrightarrow \mathbb{R}$ une application de classe $\mathcal{C}^1$ sur un ouvert U de $\mathbb{R}^n$, soit $\mathcal{C}$ un système de contraintes linéaires et soit $a \in U \cap \mathcal{C}$. Si f admet un extremum local en a sous la contrainte $\mathcal{C}$, alors :*

$$\nabla(f)(a) \perp \mathcal{H}.$$

Définition 3.5 *Soit $f : U \longrightarrow \mathbb{R}$ une application de classe $\mathcal{C}^1$ sur un ouvert U de $\mathbb{R}^n$, soit $\mathcal{C}$ un système de contraintes linéaires et soit $a \in U \cap \mathcal{C}$. On dit que a est un point critique de f (pour l'optimisation) sous la contrainte $\mathcal{C}$ si $\nabla(f)(a) \perp \mathcal{H}$.*

En d'autres termes, si f admet un extremum local en a sous la contrainte $\mathcal{C}$, alors a est un point critique de f sous la contrainte $\mathcal{C}$. A noter que ce résultat n'est pas une équivalence, mais seulement une condition nécessaire d'existence d'un extremum local sous contrainte. En effet, il existe des fonctions admettant des points critiques sous contraintes qui ne sont pas des extrema locaux sous contraintes. Remarquons qu'avec les notations de la définition 3.1, si l'on pose $u_i = (a_{i,1}, ..., a_{i,n})$ pour tout $i \in \{1, ..., p\}$, alors on voit que $g_i(x) = \langle u_i, x\rangle$ pour tout $i \in \{1, ..., p\}$, et donc $\mathcal{H}$ est l'ensemble des vecteurs de $\mathbb{R}^n$ orthogonaux à tous les vecteurs u_i. En particulier, il s'ensuit que $\mathcal{H}^\perp = \mathrm{Vect}(u_1, ..., u_p)$. On montre alors par des calculs simples que $u_i = \nabla(g_i)$ pour tout $i \in \{1, ..., p\}$, ce qui entraine que :

$$\boxed{\mathcal{H}^\perp = \mathrm{Vect}(\nabla(g_1), ..., \nabla(g_p)).}$$

En particulier, on voit par le calcul que $\nabla(g_i)$ ne dépend pas du point d'application x. Dès lors, la notion de point critique sous contraintes peut se retraduire de la façon suivante :

Théorème 3.6 *Soit $f : U \longrightarrow \mathbb{R}$ une application de classe $\mathcal{C}^1$ sur un ouvert U de $\mathbb{R}^n$, soit $\mathcal{C}$ un système de contraintes linéaires et soit $a \in U \cap \mathcal{C}$. Alors a est un point critique de f sous la contrainte $\mathcal{C}$ si et seulement s'il existe des réels $\lambda_1, ..., \lambda_p$ tels que $\nabla(f)(a) = \lambda_1\nabla(g_1) + ... + \lambda_p\nabla(g_p)$.*

A noter que, dans la littérature mathématique, les constantes $\lambda_1, ..., \lambda_p$ qui apparaissent (si elles sont uniques) sont appelées *les multiplicateurs de Lagrange*. De façon générale, une fois détecté un point critique a de f sous contraintes, on doit déterminer s'il s'agit d'un extremum local ou global de f sous la contrainte $\mathcal{C}$, ce qui se fait en étudiant le signe de $f(x) - f(a)$ pour tout $x \in U\cap\mathcal{C}$. Pour ce faire, on pourra notamment utiliser la formule de Taylor avec reste intégral à l'ordre 1 telle qu'elle apparaît dans le théorème 2.6.