Résumé du cours d'optimisation (recherche d'extrémums)

Généralités

Soit $f : C \subset \mathbb{R}^n \to \mathbb{R}$ et $a \in C$.

D1 : On dit que a est un **minimum global** de f sur C si, pour tout $x \in C$, $f(a) \leq f(x)$. L'ensemble des minimums de f sur C est noté $\mathrm{Arg}_C \min f$.

D2 : On dit que a est un **minimum local** de f sur C s'il existe une boule ouverte B contenant a telle que a soit minimum global de f sur $B \cap C$.

Conditions suffisantes d'existence d'un minimum global si f est continue:

> **TH** : Soit f une application continue de C dans $\mathbb{R}$.
> **1)** Si C est compact, alors $\mathrm{Arg}_C \min f$ est un compact non vide.
> **2)** Si C est fermé et f coercive (c'est-à-dire $\lim\limits_{\|x\| \to +\infty} f(x) = +\infty$), alors $\mathrm{Arg}_C \min f$ est un compact non vide.

Rappels de différentiabilité

D3 : On dit que f est **différentiable en** a s'il existe une forme linéaire $d_a f$ telle que :

$$f(a+h) = f(a) + d_a f(h) + \|h\| \varepsilon(h)$$

si $a + h \in \Omega$, avec $\varepsilon(h) \to 0$ si $\|h\| \to 0$.

> **Condition nécessaire (mais pas suffisante!) pour que f soit différentiable en a:**
> les dérivées partielles $D_i f$ existent en a pour tout $i \in \{1, \cdots, n\}$.

On a alors $d_a f(h) = \langle \nabla f(a), h \rangle$ où le vecteur $\nabla f(a) = (D_i f(a))_{1 \leq i \leq n}$ est le **gradient** de f en a.

D4 : On dit que f est **de classe $\mathcal{C}^1$** en a si les dérivées partielles $D_i f$ de f existent dans un voisinage de a et sont continues en a.

> **Condition suffisante (mais pas nécessaire!) pour que f soit différentiable en a:**
> f est de classe $\mathcal{C}^1$ en a.

D5 : On dit que f est **de classe $\mathcal{C}^2$** en a si les dérivées partielles d'ordre 2 de f existent dans un voisinage de a et sont continues en a. On note $\nabla^2 f(a)$ et on appelle **Hessienne de f en** a, la matrice $(D_i D_j f(a))_{1 \leq i,j \leq n}$.
D'après le théorème de Schwarz, $D_i D_j f(a) = D_j D_i f(a)$ donc $\nabla^2 f(a)$ une matrice symétrique.

Formule de Taylor-Lagrange : $f(a+h) = f(a) + \langle \nabla f(a), h \rangle + \frac{1}{2} \langle \nabla^2 f(a) h, h \rangle + \|h\|^2 \varepsilon(h)$
avec $\varepsilon(h) \to 0$ quand $\|h\| \to 0$, si f est de classe $\mathcal{C}^2$ en a, et si $a + h \in \Omega$.

Optimisation des fonctions différentiables

1) Conditions nécessaires :

> **TH** : Si f admet un minimum local en a en lequel elle est différentiable, alors $\nabla f(a) = 0$.

Les points a solutions de $\nabla f(a) = 0$ sont appelés **points critiques** de f (ou points stationnaires de f).

TH : Si f admet un minimum local en a et si f est de classe $\mathcal{C}^2$ en a, alors $\langle \nabla^2 f(a)h, h\rangle \geq 0$ pour tout $h \in \mathbb{R}^n$, c'est-à-dire que les valeurs propres de la matrice $\nabla^2 f(a)$ sont positives.

2) Conditions suffisantes de minimum local :

TH : Soit $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ différentiable en a stationnaire. Dans les 2 cas suivants, f admet un minimum local en a :
1) f est de classe $\mathcal{C}^2$ en a et les valeurs propres de $\nabla^2 f(a)$ sont strictement positives.
2) Il existe $r > 0$ tel que f soit de classe $\mathcal{C}^2$ sur la boule $B(a, r)$ et tel que, pour tout $h \in \mathbb{R}^n$, $\langle \nabla^2 f(x)h, h\rangle \geq 0$ pour tout $x \in B(a, r)$.

En pratique :

- On commence par déterminer le gradient.
- On déterminer ensuite les points où ce gradient s'annule.
- On détermine alors la hessienne de f en ces points.
- On étudie le signe des valeurs propres de la hessienne de f en ces points : si elles sont toutes positives, on a un minimum, si elles sont toutes négatives, on a un maximum, si on a des valeurs propres < 0 et d'autres > 0, il n'y a pas d'extrémums en ces points. (Pour $n = 2$, il suffit de connaître le signe du déterminant et de la trace pour conclure).
- Lorsqu'il y a au moins une valeur propre nulle, et les autres de même signe, il faut faire l'étude "à la main".

<u>**Cas des fonctions convexes**</u>

D6 : Un sous-ensemble C de $\mathbb{R}^n$ est dit **convexe** si, pour tout $(a, b) \in C^2$, $[a, b] \subset C$ (c'est-à-dire pour tout $\lambda \in [0, 1]$, $\lambda a + (1 - \lambda)b \in C$).

D7 : Une fonction réelle f définie sur C convexe est dite **convexe** si, pour tout $(a, b) \in C^2$ et pour tout $\lambda \in [0, 1]$, $f(\lambda a + (1 - \lambda)b) \leq \lambda f(a) + (1 - \lambda)f(b)$.

Propriétés :
- Si C est un convexe de $\mathbb{R}^n$ et $(f_i)_{i\in I}$, une famille quelconque de fonctions convexes alors
  a) $\sup\limits_{i\in I} f_i$ est convexe ;
  b) si I est fini et si $(\lambda_i)_{i\in I}$ est une famille de réels positifs, alors $\sum\limits_{i\in I} \lambda_i f_i$ est convexe.
- Si C est un convexe de $\mathbb{R}^n$, si f est une fonction convexe de C sur $\mathbb{R}$ et si φ est une fonction convexe croissante sur $\mathbb{R}$, alors $\varphi \circ f$ est une fonction convexe.

Caractérisation des fonctions convexes de classe $\mathcal{C}^2$

TH : Soit Ω un ouvert convexe de $\mathbb{R}^n$ et f une fonction de Ω sur $\mathbb{R}$.
Si f est de classe $\mathcal{C}^2$ sur Ω, alors f est convexe sur Ω si et seulement si, pour tout $x \in \Omega$, les valeurs propres de $\nabla^2 f(x)$ sont positives.

Minimum global d'une fonction convexe.

TH : Soit C un convexe de $\mathbb{R}^n$, f une fonction convexe de C sur $\mathbb{R}$ et $a \in C$, alors
1) un minimum local est un minimum global ;
2) si f est de classe $\mathcal{C}^1$ sur C et si C est ouvert, alors $a \in \mathrm{Arg}_C \min f$ si et seulement si $\nabla f(a) = 0$.

Les relations de Kuhn-Tucker.

D8 Soit $X = \{x \in \mathbb{R}^n \ ; \ \varphi_i(x) \leq 0$ pour tout $i \in I\}$. Un arc de courbe Γ défini par $\gamma : \mathbb{R}_+ \to \mathbb{R}^n$ de classe $\mathcal{C}^1$ est dit **admissible** en $\overline{x} \in X$ si $\gamma(0) = \overline{x}$ et si $\gamma(\theta) \in X$ pour θ assez petit. On appelle alors **direction admissible** en $\overline{x}$ le vecteur $v = \gamma'(0)$.
On note $C(\overline{x})$ la fermeture du cône formé par l'ensemble des directions admissibles en $\overline{x}$.

D9 : Si $x \in X$, on note $I(x) = \{i \in I \ ; \ \varphi_i(x) = 0\}$ et on appelle $I(x)$ l'ensemble des **contraintes saturées** en x. On pose $C^*(x) = \{w \in \mathbb{R}^n \ ; \ \langle \nabla\varphi_i(x), w\rangle \ \leq \ 0$ pour tout $i \in I(x)\}$.

TH : On a $C(x) \subset C^*(x)$ mais la réciproque est fausse en général.

D10 : On dit que les contraintes sont **qualifiées** en x si $C(x) = C^*(x)$.

Lemme admis : Pour que (QC) soit vérifiée en tout point x de X, il suffit que les φ_i soient linéaires, ou convexes si X est d'intérieur non vide.
Pour que (QC) soit vérifiée en $\overline{x}$, il suffit que la famille $(\nabla\varphi_i(\overline{x}))_{i\in I(\overline{x})}$ soit libre.

TH : Kuhn-Tucker
Si f admet un minimum en $\overline{x} \in X = \{x \ ; \ \varphi_i(x) \leq 0$ pour tout $i \in I\}$, et si les contraintes sont qualifiées en $\overline{x}$, alors il existe une famille $(\lambda_i(\overline{x}))_{i\in I}$ de réels positifs ou nuls tels que :

$$\begin{cases} \nabla f(\overline{x}) + \sum\limits_{i\in I} \lambda_i(\overline{x})\nabla\varphi_i(\overline{x}) = 0 \\ \lambda_i(\overline{x})\varphi_i(\overline{x}) = 0 \text{ pour tout } i \in I \\ \varphi_i(\overline{x}) \leq 0 \text{ pour tout } i \in I \end{cases}$$

Interprétation : les contraintes déterminent les “bords” du domaine. $\varphi(x) = 0$ signifie que x est sur le “bord” déterminé par φ. Si $\varphi(x) \neq 0$, le λ correspondant est nul et le gradient n'intervient pas dans la première relation.

En pratique : Soit aucune contrainte n'est saturée (tous les λ_i sont alors nuls et on a un point critique), soit le point est sur un bord. Si une seule contrainte est saturée, on peut utiliser la première relation ; sinon, il est souvent plus simple de déterminer directement l'intersection de deux bords...

Extension à des problèmes avec en plus des contraintes d'égalités

On note ici $X = \{x \in \mathbb{R}^n \ ; \ \varphi_i(x) \leq 0$ pour tout $i \in I$ et $\psi_j(x) = 0$ pour tout $j \in J\}$.

TH : Si f admet sur X un minimum en $\overline{x}$ en lequel les contraintes sont qualifiées, alors il existe des nombres positifs ou nuls $\lambda_i(\overline{x})$ pour $i \in I$ et des nombres $\mu_j(\overline{x})$ (non nécessairement positifs) pour $j \in J$ tels que :

$$\begin{cases} \nabla f(\overline{x}) + \sum\limits_{i\in I} \lambda_i(\overline{x})\nabla\varphi_i(\overline{x}) + \sum\limits_{j\in J} \mu_j(\overline{x})\nabla\psi_j(\overline{x}) = 0 \\ \lambda_i(\overline{x})\varphi_i(\overline{x}) = 0 \text{ pour tout } i \in I \\ \varphi_i(\overline{x}) \leq 0 \text{ et } \psi_j(\overline{x}) = 0 \text{ pour tout } i \in I \text{ et pour tout } j \in J. \end{cases}$$

On note ici $X = \{x \in \mathbb{R}^n \ ; \ \psi_j(x) = 0$ pour tout $j \in J\}$.

TH : Si f admet sur X un minimum en $\overline{x}$ en lequel les contraintes sont qualifiées, alors il existe une famille de nombres $(\mu_j(\overline{x}))_{j\in J}$ (non nécessairement positifs) telle que :

$$\nabla f(\overline{x}) + \sum_{j\in J} \mu_j(\overline{x})\nabla\psi_j(\overline{x}) = 0.$$