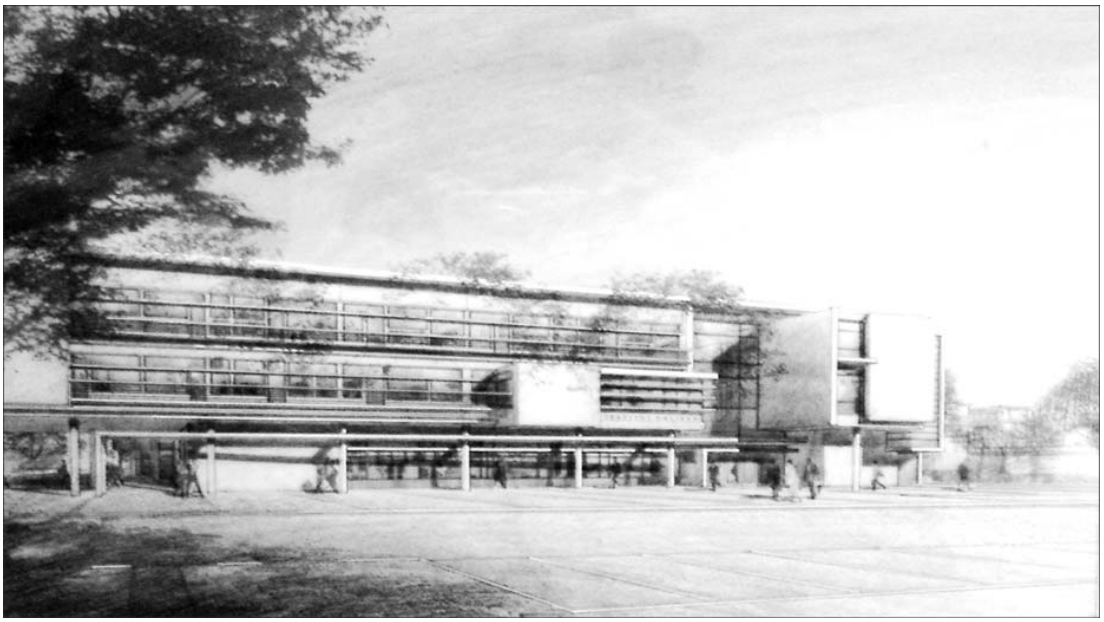


OPTIMISATION

MACS 2, 2019

L. HALPERN

Table des matières

Soit V un espace de Hilbert de dimension finie sur $\mathbb{R}$, K une partie de V , J une fonction définie sur V à valeurs dans $\mathbb{R}$. On dit que u est **minimum local** de J sur K si u appartient à K et s'il existe un voisinage U de u dans K tel que

$$\forall v \in U, J(u) \leq J(v). \tag{1}$$

Si la relation précédente est vraie pour tout v dans K, on dit que u est **minimum global** de J sur K. On parle de minimum strict si l'inégalité est stricte dans (1.1). On définit un problème de minimisation sur K par

$$\begin{cases} u \in K, \\ J(u) = \inf\limits_{v \in K} J(v) \end{cases} \tag{2}$$

K est l'ensemble des **contraintes**. On dit que u est **solution optimale** du problème de minimisation sur K. Le problème de minimisation est dit **sans contrainte** si $V = K$, **avec contraintes** si $V \neq K$.

Bien évidemment, on définit un problème de maximisation, en remplaçant $\leqslant$ par $\geqslant$ dans (1.1) et inf par sup dans (1.2). On parlera en général de problème d'optimisation. On passe de l'un à l'autre en définissant la fonctionnelle opposée. Dans ce cours tous les résultats sont établis sur les problèmes de minimisation.

Les questions que l'on se pose sont :

1. Est-ce que la fonction J est bornée inférieurement sur K ?
2. Existe t-il des extrema locaux ?
3. Existe t-il des extrema globaux c'est-à-dire est-ce que la borne inférieure est atteinte ?
4. Combien y en a t-il ?
5. S'il est unique comment le caractériser ?
6. Comment le calculer ?

Première partie

Résultats théoriques

Chapitre 1

Résultats d'existence

Sommaire

Soit V un espace de Hilbert sur $\mathbb{R}$, K une partie de V , J une fonction définie sur V à valeurs dans $\mathbb{R}$. On dit que u est **minimum local** de J sur K si u appartient à K et s'il existe un voisinage U de u dans K tel que

$$\forall v \in U, J(u) \leq J(v). \tag{1.1}$$

Si la relation précédente est vraie pour tout v dans K, on dit que u est **minimum global** de J sur K. On parle de minimum strict si l'inégalité est stricte dans (1.1). On définit un problème de minimisation sur K par

$$\begin{cases} u \in K, \\ J(u) = \inf\limits_{v \in K} J(v) \end{cases} \tag{1.2}$$

On dit alors que u est **solution optimale** du problème de minimisation sur K. Le problème de minimisation est dit **sans contrainte** si $V = K$, **avec contraintes** si $V \neq K$.

Bien évidemment, on définit un problème de maximisation, en remplaçant $\leqslant$ par $\geqslant$ dans (1.1) et inf par sup dans (1.2). On parlera en général de problème d'optimisation. On passe de l'un à l'autre en définissant la fonctionnelle opposée. Dans ce cours tous les résultats sont établis sur les problèmes de minimisation.

1.1 Théorèmes généraux

Définition 1.1 *Une suite minimisante pour la fonction J sur l'ensemble K est une suite u_n d'éléments de K telle que pour tout ε, il existe N, pour tout $n \geqslant N$,*

$$J(u_n) - \varepsilon \ \leq \ \inf_{v \in K} J(v) \ \leq \ J(u_n).$$

Théorème 1.1 *Soit V un espace vectoriel de dimension finie, K une partie fermée non vide de V. SOit J une fonction continue sur K.*

1. *Si K est compact ou*
2. *si J est infinie à l'infini,*

dans les deux cas, le problème de minimisation (1.2) admet une solution.

PROOF La première partie du théorème relève du L2 : en dimension finie toute fonction continue sur un compact y est bornée et atteint ses bornes (théorème de Weierstrass). Pour la deuxième partie, on suppose seulement que K est fermé et non pas qu'il est borné (les deux ensemble font un compact). Choisissons un point w dans K et définissons

$$\tilde{K} = \{v \in K,\ J(v) \leq J(w)\}.$$

$\tilde{K}$ est fermé puisque défini par une inégalité large, et borné car J est infinie à l'infini. Donc compact. De plus

$$\inf_{v \in K} J(v) = \inf_{v \in \tilde{K}} J(v)$$

et on est ramené au cas 1. ■

Pour les fans de maths : La deuxième partie du théorème est vrai en dimension finie, mais pas en dimension infinie. Pourquoi ? Par contre la convexité va pallier au problème.

On rappelle qu'une partie K de V est convexe si

$$\forall (x, y) \in K, \forall \theta \in [0, 1], \theta x + (1 - \theta) y \in K \tag{1.3}$$

Une fonction J définie sur un convexe K est dite

— convexe si

$$\forall (x, y) \in K, \forall \theta \in [0, 1], J(\theta x + (1 - \theta) y) \leqslant \theta J(x) + (1 - \theta) J(y), \tag{1.4}$$

— strictement convexe si

$$\forall (x, y) \in K, x \neq y, \forall \theta \in]0, 1[, J(\theta x + (1 - \theta) y) < \theta J(x) + (1 - \theta) J(y), \tag{1.5}$$

— α convexe si

$$\begin{aligned} \forall (x, y) \in K, \forall \theta \in [0, 1], J(\theta x + (1 - \theta) y) \leqslant\ & \theta J(x) + (1 - \theta) J(y) - \\ & - \frac{\alpha}{2} \theta (1 - \theta) ||x - y||^2. \end{aligned} \tag{1.6}$$

Théorème 1.2 *Si J est convexe, tout minimum local est global, et l'ensemble des solutions optimales est convexe.*

PROOF Par l'absurde : soit u un minimum local, et supposons qu'il n'est pas global. Alors il existe un w dans K tel que $J(w) < J(u)$. Considérons le segment $[u, w] \subset K$ puisque K est convexe. Pour tout $v = \theta u + (1 - \theta) w$ dans ce segment, $J(v) < J(u)$ par convexité de J. Mais puisque u est minimum local, $J(v)$ doit être au moins égal à $J(u)$ dans un voisinage de u sur ce segment. D'où contradiction.
Soit maintenant deux solutions optimales u_1 et u_2, donc $J(u_1) = J(u_2) = \inf_{v \in K} J(v)$. Mais alors par convexité, le segment qui les joint est dans K, et pour tout $\theta \in [0, 1]$,

$$\inf_{v \in K} J(v) \ \leq\ J(\theta u_1 + (1 - \theta) u_2) \ \leq\ J(u_1) = J(u_2) = \inf_{v \in K} J(v)$$

Donc $\theta u_1 + (1 - \theta) u_2$ est aussi solution optimale et l'ensemble des solutions optimales est convexe. ■

Théorème 1.3 *Si J est strictement convexe, la solution optimale, si elle existe, est unique.*

PROOF Par l'absurde encore, supposons qu'il y en a deux, et reprenons l'inégalité précédente qui devient

$$\inf_{v\in K} J(v) \ \leq \ J(\theta u_1 + (1-\theta)u_2) < J(u_1) = J(u_2) = \inf_{v\in K} J(v)$$

et fournit une contradiction ■

Pour les fans de maths : [Dimension infinie]. Soit K un convexe fermé non vide, J une fonction définie sur K à valeurs dans $\mathbb{R}$ convexe continue. On suppose que J est infinie à l'infini (i.e. $J(v) \to +\infty$ lorsque $||v|| \to +\infty$) ou que K est borné. A lors le problème de minimisation admet une solution.

Rappelons le théorème de projection.

Théorème 1.4 *[Projection sur un convexe fermé]. Soit K une partie convexe fermée non vide d'un espace de Hilbert V, et w un point de V n'appartenant pas à K. Alors il existe un unique point de K, noté $\mathbb{P}_K w$ tel que*

$$\begin{cases} \mathbb{P}_K w \in K, \\ \|w - \mathbb{P}_K w\| = \inf\limits_{v\in K} \|w - v\| \end{cases} \tag{1.7}$$

Il est caractérisé par

$$\forall v \in K, (\mathbb{P}_K w - w, v - \mathbb{P}_K w) \geqslant 0 \tag{1.8}$$

PROOF La démonstration traditionnelle repose sur les suites minimisantes. Ici nous nous appuyons dans le cas de la dimension infinie sur le résultat précédent. Evidemment la fonction définie par $J(v) = \|w - v$ est strictement convexe, infinie à l'infinie. Il existe un minimum, et il est unique. Pour tout $v \in K$ et $\theta \in [0,1]$, puisque K est convexe, $\theta v + (1-\theta) P p_K w \in K$, et on peut donc écrire

$$\|\mathbb{P}_K w - w\|^2 \ \leq \ \|\theta v + (1-\theta) P p_K w - w)\|^2 = \|\theta(v - P p_K w) + (P p_K w - w)\|^2$$

Développons ce dernier terme

$$\|\mathbb{P}_K w - w\|^2 \ \leq \ \|P p_K w - w\|^2 + 2\theta(\mathbb{P}_K w - w, v - \mathbb{P}_K w) + \theta^2 \|v - P p_K w\|^2$$

, ce qui est équivalent à

$$\forall \theta \in [0,1], \quad 2(\mathbb{P}_K w - w, v - \mathbb{P}_K w) + \theta \|v - P p_K w\|^2 \ \geqslant \ 0.$$

Il suffit maintenant de faire tendre θ vers 0 pour obtenir le résultat.

■

Le résultat compact suivant est très utile dans les applications où les fonctions sont souvent $\alpha-$ convexes

Corollaire 1.1 . *Soit K un convexe fermé non vide, J une fonction définie sur K à valeurs dans $\mathbb{R}$, α-convexe continue. Alors le problème de minimisation admet une solution et une seule. De plus toute suite minimisante converge vers u.*

PROOF La démonstration repose sur un lemme technique d'analyse que nous ne démontrons pas ici, et qui dit qu'une fonction $\alpha-$ convexe est infinie à l'infini. Pour les fans, la démonstration de ce fait utilise le théorème de Hahn-Banach de séparation d'un point et d'un convexe. ■

1.2 Rappels de calcul différentiel

Soit J une fonction définie sur V de dimension n à valeurs dans $\mathbb{R}$.

1.2.1 Dérivées premières

Définition 1.2 (Différentiabilité) *J est différentiable en $u \in V$ de différentielle $J' : V \mapsto V$ ou ∇J avec $F'(u) = \nabla J(u)$ si,*

$$\forall w \in V, J(u+w) = J(u) + J'(u) \cdot w + \epsilon(w)\|w\|, \quad \lim_{w \to 0} \epsilon(w) = 0 \tag{1.9}$$

Dans $\mathbb{R}^n$, si J est différentiable en u, elle admet des dérivées partielles $\partial_j J(u) := \frac{\partial J}{\partial x_j}$, et $\nabla J(u) = (\partial_1 J(u), \cdots, \partial_n J(u))$, si bien que $J'(u).v = \sum_{i=1}^n \frac{\partial J}{\partial x_i}(u) v_i$.

Exemples de base

1. Les formes linéaires $J(u) = (c, u)$, où c est un vecteur donné dans V. Alors $J'(u) = \nabla J(u) = c$.
2. Les fonctions $J(u) = a(u, u)$, où a est une forme bilinéaire continue sur V. Alors $J'(u).v = a(u, v) + a(v, u)$, et si a est symétrique $J'(u).v = 2a(u, v)$.

Définition 1.3 (Dérivée directionnelle) *On appelle dérivée de J en u dans la direction v la dérivée en 0 de la fonction d'une variable $t \mapsto J(u + tv)$. On peut alors noter*

$$D_v J(u) = J'(u) \cdot v.$$

1.2.2 Dérivées secondes

Si $J' : V \mapsto V$ admet une différentielle J'' application bilinéaire continue de $V \times V$ dans $\mathbb{R}$. On notera $J"(u) \cdot v \cdot w$.

Exemples de base

1. $J(u) = (c, u)$, $J"(u) = 0$.
2. $J(u) = a(u, u)$, alors $J''(u).v.w = a(v, w) + a(w, v)$, et si a est symétrique $J''(u).v.w = 2a(v, w)$. Si $V = \mathbb{R}^n$, $J(u) = \frac{1}{2}(Au, u)$ où A est une matrice symétrique, alors $J"(u) = A$ pour tout u.
3. Si $V = \mathbb{R}^n$, $J''(u)$ est la matrice des dérivées partielles secondes $\frac{\partial^2 J}{\partial x_i \partial x_j}(u)$.

1.2.3 Formules de Taylor

Pour $u \neq v \in V$,

Taylor Mac-Laurin ordre 1 Si $J : V \mapsto \mathbb{R}$ est définie et continue sur $[u, v]$, différentiable sur $]u, v[$, il existe $\theta \in]0, 1[$ tel que

$$J(v) = J(u) + J'(u + \theta(v - u)) \cdot (v - u) \tag{1.10}$$

On l'appelle aussi la formule de la moyenne, et on peut la formuler ainsi il existe $w \in]u, v[$ tel que

$$J(v) = J(u) + J'(w) \cdot (v - u) \tag{1.11}$$

Taylor Mac-Laurin ordre 2 Si $J : V \mapsto \mathbb{R}$ est définie et continue sur $[u, v]$, 2 fois différentiable sur $]u, v[$, il existe $\theta \in]0, 1[$ tel que

$$J(v) = J(u) + J'(u) \cdot (v - u) + \frac{1}{2} J''(u + \theta(v - u)) \cdot (v - u) \cdot (v - u) \tag{1.12}$$

Taylor Young ordre 2 Si $J : V \mapsto \mathbb{R}^p$ est 2 fois différentiable en u, alors pour tout v

$$J(v) = J(u)+J'(u)\cdot(v-u)+\frac{1}{2}J''(u)\cdot(v-u)\cdot(v-u)+\epsilon(v-u)\|v-u\|^2, \quad \lim_{w\to 0}\epsilon(w) = 0 \tag{1.13}$$

Théorème 1.5 *[caractérisation des fonctions convexes]. J est convexe sur V si et seulement si l'une des conditions équivalentes suivantes est vérifiée :*

(1) Si J est différentiable, le graphe de J est au-dessus de l'hyperplan tangent, i.e.

$$\forall u, v \in V, J(v) \geqslant J(u) + J'(u)\cdot(v-u) \tag{1.14}$$

(2) Si J est différentiable, J' est un opérateur monotone, i.e.

$$\forall u, v \in V, (J'(v) - J'(u))\cdot(v-u) \geqslant 0 \tag{1.15}$$

(3) Si J est deux fois différentiable, J'' est un opérateur non négatif, i.e.

$$\forall u, w \in V, J''(u)w.w \geqslant 0 \tag{1.16}$$

PROOF Nous montrons d'abord que convexe $\iff$ (1), puis que (1) $\implies$ (2) $\implies$ (3) $\implies$ (1)
.

convexe $\implies$ (1)

$$J(\theta v + (1-\theta)u) \leq \theta J(v) + (1-\theta)J(u).$$

Divisons par θ

$$\frac{J(u+\theta(v-u)) - J(u)}{\theta} \leq J(v) - J(u).$$

Passons à la limite en $\theta \to 0$:

$$J'(u)\cdot(v-u) \leq J(v) - J(u).$$

(1) $\implies$ convexe On applique l'inégalité (1.14) successivement aux couples $(\theta u + (1-\theta)v, u)$ et $(\theta u + (1-\theta)v, v)$:

$$\begin{array}{lll} \times\theta & J(u) \geqslant & J(\theta u + (1-\theta)v) + J'(\theta u + (1-\theta)v)\cdot(1-\theta)(u-v) \\ \times(1-\theta) & J(v) \geqslant & J(\theta u + (1-\theta)v) + J'(\theta u + (1-\theta)v)\cdot\theta(v-u) \end{array}$$

Lorsque l'on ajoute ces deux inégalités, les termes de droit se neutralisent, il ne reste plus que l'inégalité de convexité

$$\theta J(u) + (1-\theta)J(v) \geqslant J(\theta u + (1-\theta)v).$$

(1) $\implies$ (2) Appliquons successivement (1.14) aux couples (u, v) et (v, u).

$$\begin{array}{l} J(v) \geqslant J(u) + J'(u)\cdot(v-u) \\ J(u) \geqslant J(v) - J'(v)\cdot(v-u) \end{array}$$

et ajoutons les pour obtenir $0 \geqslant (J'(u) - J'(v))\cdot(v-u)$ qui donne le résultat souhaité en changeant les signes.

(2) $\Longrightarrow$ **(3)** Pour tout w dans V, et $\theta > 0$, posons $v = u + \theta w$, et appliquons (2) au couple (u, v). On obtient

$$(J'(u+\theta w) - J'(u)) \cdot w \geqslant 0 \quad \text{ce qui implique} \quad \frac{(J'(u+\theta w) - J'(u)) \cdot w}{\theta} \geqslant 0$$

Par définition de la différentiabilité de J', $J'(u+\theta w) - J'(u) = \theta J''(u)w + \epsilon(\theta)\|\theta w\|$. Remplaçons dans l'inégalité précédente et divisons par θ :

$$(J''(u)w + \varepsilon(\theta)\|w\|) \cdot w \geqslant 0$$

Faisons tendre maintenant θ vers 0 pour obtenir le résultat.

(3) $\Longrightarrow$ **(1)** On utilise la formule de Taylor Mac-Laurin à l'ordre 2 (1.10) et le résultat vient tout de suite.

■

Pour une fonction strictement convexe, on a :

Théorème 1.6 *[caractérisation des fonctions strictement convexes]. Soit J différentiable sur V. J est strictement convexe et seulement si l'une des conditions équivalentes suivantes est vérifiée :*

$$\forall u, v \in V, u \neq v, \ J(v) > J(u) + J'(u) \cdot (v - u) \tag{1.17}$$

$$\forall u, v \in V, u \neq v, \ (J'(v) - J'(u)) \cdot (v - u) > 0 \tag{1.18}$$

Si J est deux fois différentiable, et si

$$\forall u, w \in V, \ w \neq 0, \ J''(u)w.w > 0, \tag{1.19}$$

alors J est strictement convexe.

PROOF SI on cherche à refaire la démonstration précédente, lorsque l'on utilise un passage à la limite, les inégalités strictes deviennent des inégalités larges, et on ne peut pas conclure. C'est le cas pour strictement convexe $\Longrightarrow$ (1). Reprenons

(1.17) $\Longrightarrow$ **strictement convexe** : même démonstration.

strictement convexe $\Longrightarrow$ (1.17) Il faut être un peu plus subtil. Écrivons pour $u \neq v$, $\theta \in]0,1[$, $\omega \in]0,1[$, $\theta \neq \omega$,

$$u + \theta(v-u) = \frac{\omega - \theta}{\omega} u + \frac{\theta}{\omega}(u + \omega(v-u)).$$

Fixons ω, et choisissons $\theta \leq \omega$. Par convexité on a donc

$$J(u+\theta(v-u)) < \frac{\omega-\theta}{\omega} J(u) + \frac{\theta}{\omega} J(u+\omega(v-u)),$$

inégalité que nous réécrivons en soustrayant $J(u)$ des deux côtés et en divisant par θ :

$$\frac{J(u+\theta(v-u)) - J(u)}{\theta} \leq \frac{J(u+\omega(v-u)) - J(u)}{\omega}.$$

ω est toujours fixé dans $]0,1[$, passons maintenant à la limite en θ,

$$J'(u) \cdot (v-u) \leq \frac{J(u+\omega(v-u)) - J(u)}{\omega},$$

Par stricte convexité de nouveau on a

$$J(u+\omega(v-u)) = J((1-\omega)u + \omega v) < (1-\omega)J(u) + \omega J(v),$$

et en reportant dans l'inégalité du dessus on obtient

$$J'(u) \cdot (v-u) < J(v) - J(u),$$

qui est le résultat souhaité. On a donc maintenant strictement convexe $\Longleftrightarrow$ (1.17).

(1.17) $\Longrightarrow$ (1.18) : même démonstration.

(1.18) $\Longrightarrow$ **strictement convexe** De nouveau il faut être un peu subtil. Soit $u \neq v$ et $\theta \in]0,1[$. Utilisons la formule de la moyenne eqrefeq :TML1b pour $(u, u+\theta(v-u))$ et $(v, u+\theta(v-u))$. Il existe donc $w_1 \in]u, u+\theta(v-u)[$ et $w_2 \in]u+\theta(v-u), v[$ tels que

$$\begin{array}{ll} \times(1-\theta) & J(u+\theta(v-u)) = J(u) + J'(w_1)\cdot\theta(v-u) \\ \times\theta & J(u+\theta(v-u)) = J(v) + J'(w_2)\cdot(1-\theta)(u-v) \end{array}$$

Multiplions la première par $1-\theta$, la deuxième par θ, et ajoutons.

$$J(u+\theta(v-u)) = (1-\theta)J(u) + \theta J(v) + \theta(1-\theta)(J'(w_1) - J'(w_2))\cdot(v-u).$$

Quel est le signe de la quantité en rouge ? Notons que puisque $w_1 \in]u, u+\theta(v-u)[$ et $w_2 \in]u+\theta(v-u), v[$, ils sont distincts, et s'écrivent avec $\alpha \in]0,1[$ et $\beta \in]0,1[$

$$\begin{aligned} w_1 &= \alpha(u+\theta(v-u)) + (1-\alpha)u = u + \alpha\theta(v-u), \\ w_2 &= \beta(u+\theta(v-u)) + (1-\beta)v = v + \beta(1-\theta)(u-v), \end{aligned}$$

si bien que

$$w_1 - w_2 = (1 - \alpha\theta - \beta(1-\theta))(u-v).$$

Il suffit maintenant de remarquer que par convexité $\gamma = 1-\alpha\theta-\beta(1-\theta) \in]0,1[$, et donc

$$J(u+\theta(v-u)) = (1-\theta)J(u) + \theta J(v) - \frac{1}{\gamma}(J'(w_1) - J'(w_2))\cdot(w_1 - w_2).$$

Par hypothèse le terme en rouge est positif, ce qui donne la convexité stricte.

(1.19) $\Longrightarrow$ **strictement convexe** Ici la démonstration repose sur Taylor Mac-Laurin à l'ordre 2, comme dans le cas convexe.

■

Pour une fonction α -convexe, on a :

Théorème 1.7 *[caractérisation des fonctions α -convexes]. J est α- convexe sur V si et seulement si l'une des conditions équivalentes suivantes est vérifiée :*

(1) Si J est différentiable,

$$\forall u, v \in V, J(v) \geqslant J(u) + J'(u)\cdot(v-u) + \frac{\alpha}{2} \parallel v-u \parallel^2, \tag{1.20}$$

(2) Si J est différentiable,

$$\forall u, v \in V, (J'(v) - J'(u))\cdot(v-u) \geqslant \alpha \parallel v-u \parallel^2, \tag{1.21}$$

(3) Si J est deux fois différentiable,

$$\forall u, w \in V, J''(u)w.w \geqslant \alpha \parallel w \parallel^2 . \tag{1.22}$$

PROOF En exercice. ■

En particulier les fonctionnelles de la forme $J(u) = a(u,u)$, où a est une forme bilinéaire symétrique continue sur V sont α-convexes si et seulement si

$$\forall u \in V, 2a(w,w) \geqslant \alpha\|w\|^2$$

Si l'on est dans $\mathbb{R}^n$, avec $J(u) = \frac{1}{2}(Au, u)$, ceci revient à

$$\forall u \in V, (Aw, w) \geqslant \alpha \|w\|^2$$

La matrice A étant symétrique, elle diagonalise en base orthonormée, $A = PDP^T$, où D est la matrice des valeurs propres d_i et P la matrice des vecteurs propres. On a alors

$$(Aw, w) = \sum_{i=1}^{n} d_i((Pw)_i)^2 \geqslant (\min_{1 \leq i \leq n} d_i) \sum_{i=1}^{n} ((Pw)_i)^2$$

$$(Aw, w) \geqslant (\min_{1 \leq i \leq n} d_i)\|Pw\|^2 = (\min_{1 \leq i \leq n} d_i)\|w\|^2$$

car, puisque P est orthogonale, $\|Pw\| = \|w\|$. Si A est définie positive, la fonctionnelle J est $\min_{1 \leq i \leq n} d_i$-convexe.

Chapitre 2

Caractérisation des extrema

Sommaire

2.1 Equation d'Euler, cas général

Théorème 2.1 *[condition nécessaire]. Si u est minimum local de J dans V, alors*

(1) Si J est différentiable en u, $J'(u) = 0$,

(2) Si J est deux fois différentiable en u, on a de plus $\forall w \in V, J''(u)w.w \geqslant 0$.

PROOF Pour tout $v \in V$, définissons la fonction d'une variable réelle

$$\varphi(t) = J(u + tv).$$

(1) Pour t réel assez petit, $u + tv$ est dans un voisinage de u et puisque u est minimum local, $J(u + tv) \geqslant J(u)$ ce qui se traduit par $\varphi(t) \geqslant \varphi(0)$. Par définition φ est dérivable en 0 et $\varphi'(0) = J'(u) \cdot v$. Ecrivons

$$0 \leq \lim_{t \to 0_+} \frac{\varphi(t) - \varphi(0)}{t} = \varphi'(0) = \lim_{t \to 0_-} \frac{\varphi(t) - \varphi(0)}{t} \leq 0.$$

ce qui prouve que $\varphi'(0) = 0$. On a donc pour tout v, $J'(u) \cdot v = 0$, et donc $J'(u) = 0$.

(2) Dans le deuxième cas, appliquons la formule de Taylor-Young à l'ordre 2 à la fonction φ :

$$\varphi(t) = \varphi(0) + t\varphi'(0) + \frac{t^2}{2}(\varphi''(0) + \epsilon(t)), \quad \text{avec } \lim_{t \to 0} \epsilon(t) = 0.$$

D'après la première partie, $\varphi'(0) = 0$. supposons que $\varphi''(0) < 0$, alors pour t assez petit, $\frac{1}{2}\varphi''(0) + \epsilon(t)$ est encore strictement négatif, et donc $\varphi(t) < \varphi(0)$ ce qui contredit le fait que u est minimum local. Donc $\varphi''(0) \geqslant 0$ et il est facile de calculer que c'est égal à $J''(u) \cdot v \cdot v$.

■

Théorème 2.2 *[condition suffisante]. Soit J une fonction différentiable dans V et u un point de V tel que $J'(u) = 0$.*

(1) Si J est deux fois différentiable dans un voisinage de u et s'il existe un voisinage Ω de u tel que $\forall v \in \Omega, \forall w \in V, J''(v)w.w \geqslant 0$, alors u est minimum local de J.

(2) Si J est deux fois différentiable,et s'il existe $\alpha > 0$ tel que

$$\forall w \in V, J''(u)w.w \geqslant \alpha \|w\|^2,$$

alors u est minimum local strict pour J.

PROOF

(1) Puisque Ω est un voisinage, il contient une boule ouverte de centre u, donc convexe. Soit v dans cette boule. Alors tout le segment $]u, v[$ est contenu dans cette boule, et d'après la formule de Taylor Mac-Laurin à l'ordre 2,

$$J(v) = J(u) + \underbrace{J'(u) \cdot (v-u)}_{0} + \frac{1}{2} \underbrace{J''(u + \theta(v-u)) \cdot (v-u) \cdot (v-u)}_{\geqslant 0}$$

et donc $J(v) \geqslant J(u)$: u est un minimum local.

(2) Utilisons la formule de Taylor-Young et l'hypothèse :

$$J(v) \ \geqslant \ J(u) + \alpha \|(v-u)\|^2 + \epsilon(v-u)\|v-u\|^2, \quad \lim_{w \to 0} \epsilon(w) = 0$$

Pour v suffisamment proche de u , $\alpha + \epsilon(v-u) > 0$, et donc $J(v) \ > \ J(u)$.

■

2.2 Inéquation d'Euler, ensemble des contraintes convexe

Dans cette section on considère le problème de minimisation globale sur K, soit avec contraintes, On l'écrit

$$\text{Trouver } u \in K \text{ tel que } J(u) = \inf_{v \in K} J(v). \tag{2.1}$$

Si u est solution de (2.1), on dit que u est solution optimale. On suppose que K est un convexe fermé non vide et que J est différentiable.

Théorème 2.3 *Si u est solution optimale, on a l'inéquation d'Euler*

$$\begin{cases} u \in K \\ \forall v \in K, J'(u).(v-u) \geqslant 0. \end{cases} \tag{2.2}$$

Réciproquement si on a l'inéquation d 'Euler en u et si de plus J est convexe, alors u est solution optimale.

PROOF Soit $v \in K$. Puisque K est convexe, le segment $[u, v]$ est contenu dans K. Soit de nouveau la fonction $\varphi(t) = J(u + t(v-u))$ pour $t \in [0, 1]$. Elle est dérivable puisque J est différentiable. Appliquons-lui la définition de la dérivée en 0. Pour $t \in [0, 1]$,

$$\varphi(t) = \varphi(0) + t\varphi'(0) + t\epsilon(t), \quad \lim_{t \to 0} \epsilon(t) = 0,$$

ce qui s'écrit en termes de J, pour tout $t \in [0,1]$, en utilisant le fait que u est solution optimale,

$$J(u + t(v-u) = J(u) + t(J'(u)(v-u) + \epsilon(t)) \geqslant J(u),$$

ce qui implique que $J'(u)(v-u) + \epsilon(t) \geqslant 0$. Faisons maintenant tendre t vers 0, pour obtenir $J'(u)(v-u) \geqslant 0$.

Réciproquement, nous supposons maintenant que J est en plus convexe et satisfait l'inéquation d'Euler. Utilisons la caractérisation de la convexité (1.14). Pour tout v dans K,

$$J(v) \geqslant J(u) + J'(u).(v-u) \geqslant J(u).$$

■

Nous retrouvons en corollaire le théorème de projection sur un convexe fermé (en exo).

Les cas particuliers sont très importants.

1. $K = V$ C'est le cas sans contrainte. On a le

Théorème 2.4 *Si J est convexe différentiable, alors u réalise le minimum de J sur V si et seulement si $J'(u) = 0$.*

Remarque 2.1 *. En particulier si J est α-convexe, il existe une unique solution optimale, caractérisée par $J'(u) = 0$.*

2. K **sous-espace affine** engendré par le sous-espace vectoriel fermé E de V, i.e. $K = \{u_0 + v, v \in E\}$, alors

$$(2.2) \iff \begin{cases} u \in K \\ \forall w \in E,\ J'(u).w = 0 \end{cases} \tag{2.3}$$

Ce qui se traduit par : $\nabla J(u) \in E^{\perp}$. Si E est défini par m contraintes, $E = \{v \in V, (a_i, v) = 0, 1 \leqslant i \leqslant m\}$, alors l'orthogonal de E est l'espace vectoriel engendré par les a_i, et donc

$$(2.2) \iff \begin{cases} u \in K \\ \exists \lambda_1, .., \lambda_m \in \mathbb{R}, \quad \nabla J(u) + \sum_{i=1}^{m} \lambda_i a_i = 0. \end{cases} \tag{2.4}$$

Remarque 2.2 *Si l'on définit les fonctions affines $F_i(w) = (w - u_0, a_i)$, alors $K = \{w \in V, F_i(w) = 0\}$, et (2.4) se réécrit*

$$(2.2) \iff \begin{cases} u \in K \\ \exists \lambda_1, .., \lambda_m, \nabla J(u) + \sum_{i=1}^{m} \lambda_i F_i'(u) = 0. \end{cases} \tag{2.5}$$

3. K **cône convexe fermé** de sommet u_0. On note K_0 le cône de sommet O qui lui est parallèle. (K_0 est un cône si $w \in K_0 \implies tw \in K_0$ pour tout $t \geqslant 0$). Alors

$$(2.2) \iff \begin{cases} u \in K \\ J'(u).(u_0 - u) = 0 \\ \forall w \in K_0,\ J'(u).w \geqslant 0. \end{cases} \tag{2.6}$$

PROOF Pour tout $v \in K$, $u_0+t(v-u_0) \in K$, et donc en particulier pour $v = u$ l'inéquation d'Euler s'écrit

$$\forall t \geqslant 0, \quad J'(u) \cdot (u_0 + t(u - u_0) - u) \geqslant 0,$$

ou encore

$$\forall t \geqslant 0, \quad (1 - t)J'(u) \cdot (u_0 - u) \geqslant 0.$$

Comme $1-t$ peut prendre des valeurs positives ou négatives, ceci implique que $J'(u)\cdot(u_0-u) = 0$. Maintenant pour tout $v = u_0 + w$, $w \in K_0$, écrivons $J'(u) \cdot (u_0 + w - u) \geqslant 0$, et puisque $J'(u) \cdot (u_0 - u) = 0$, il nous reste plus que $\forall w \in K_0,\ J'(u).w \geqslant 0$. ■

Pour M cône convexe fermé de sommet O, on définit le cône dual par

$$M^\star = \{c \in V, \forall v \in M, (c, v) \geqslant 0\}. \tag{2.7}$$

Si M est engendré par un nombre fini de vecteurs, alors on peut décrire $M^\star$:

Théorème 2.5 *[Lemme de Farkas].*
Si $M = \{v \in V, \forall i \in \{1, .., m\}, (v, a_i) \leqslant 0\}$, alors $c \in M^\star$ si et seulement si $-c$ appartient au cône convexe engendré par les a_i, i.e. il existe $\{\lambda_1, .., \lambda_m\}$ tous $\geqslant 0$ tels que $c = -\sum_{i=1}^{m} \lambda_i a_i$.

PROOF Il est d'abord clair que si $c = -\sum_{i=1}^{m} \lambda_i a_i$ avec tous les λ_i positifs ou nuls, alors pour tout v dans M, $(v, c) = -\sum_{i=1}^{m} \lambda_i (a_i, v) \geqslant 0$ et donc $c \in M^*$. La réciproque repose sur le théorème de Hahn-Banach et est hors programme. ■

Intéressons nous maintenant au cas où K_0 est défini par m contraintes, $K_0 = \{w \in V, (a_i, w) \leq 0, 1 \leqslant i \leqslant m\}$. Alors la troisième ligne dans (2.6) exprime que $-J'(u)$ est dans $K_0^\star$, et donc (2.6) se réécrit

$$(2.2) \iff \begin{cases} u \in K \\ J'(u).(u_0 - u) = 0 \\ \exists(\lambda_1, \cdots, \lambda_m) \geqslant 0, \nabla J(u) + \sum_{i=1}^{m} \lambda_i a_i = 0 \end{cases} \tag{2.8}$$

Remarquons comme dans le cas précédent que K se définit ici comme $K = \{w \in V, F_i(w) \leq 0, 1 \leqslant i \leqslant m\}$, et (2.8) s'écrit

$$(2.2) \iff \begin{cases} u \in K \\ J'(u).(u_0 - u) = 0 \\ \exists(\lambda_1, \cdots, \lambda_m) \geqslant 0, \nabla J(u) + \sum_{i=1}^{m} \lambda_i F_i'(u) = 0 \end{cases} \tag{2.9}$$

2.3 Multiplicateurs de Lagrange, cas général

Le lemme de Farkas va nous permettre de trouver des conditions nécessaires d'optimalité dans le cas général.

Pour K fermé non vide, pour tout v dans K, nous définissons le cône des directions admissibles par

$$K(v) = \{0\} \cup \{w \in V, \\ \exists \{v_k\}_{k\in\mathbb{N}} \subset K \lim_{k\to+\infty} v_k = v, v_k \neq v \text{ pour tout } k, \lim_{k\to+\infty} \frac{v_k - v}{||v_k - v||} = \frac{w}{||w||}\} \tag{2.10}$$

Théorème 2.6 *Pour tout v dans K, $K(v)$ est un cône fermé de sommet O.*

Avant de démontrer le théorème, illustrons le en dimension 2.

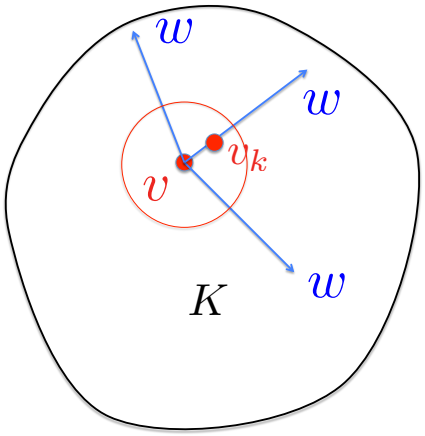


FIGURE 2.1 – Cas 1 : $v \in \mathring{K}$

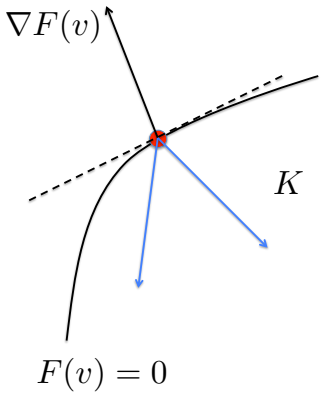


FIGURE 2.2 – Cas 2 : $v \in \text{Fr } K, F(v) = 0$

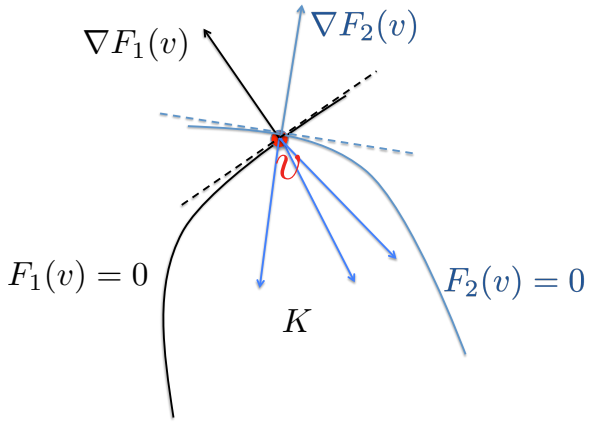


FIGURE 2.3 – Cas 3 : $v \in \text{Fr}\, K, F_1(v) = F_2(v) = 0$, cas convexe

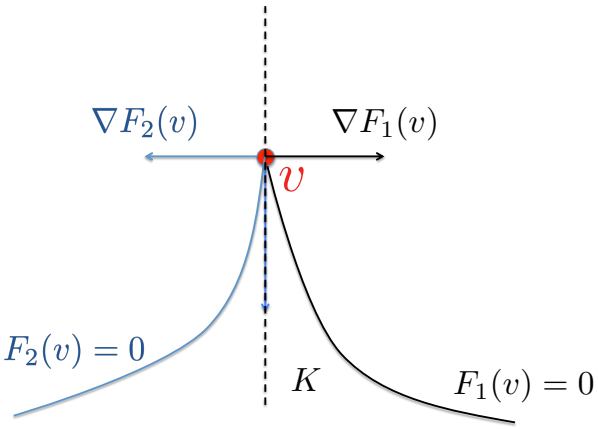


FIGURE 2.4 – Cas 4 : $v \in \text{Fr}\, K, F_1(v) = F_2(v) = 0$, cas rebroussement

PROOF Il est clair que $K(v)$ un cône de sommet O, puisque pour $t > 0$, $\frac{tw}{\|tw\|} = \frac{w}{\|w\|}$. Pour montrer qu'il est fermé, prenons une suite w^n d'éléments de $K(v)$, supposons que la suite tend vers $w \in V$ et montrons que $w \in K(v)$. Si $w = 0$ ou est égal à l'un des w_n, alors on a bien $w \in K(v)$ et c'est fini. Sinon, associons à chaque w^n une suite $(v_k^n)_k$ d'éléments de K tous différents de v, qui converge vers w^n quand k tend vers l'infini et telle que

$$\forall n \geqslant 0, \quad \lim_{k\to+\infty} v_k^n = v, \quad \text{et} \quad \lim_{k\to+\infty} \frac{v_k^n - v}{\|v_k^n - v\|} = \frac{w^n}{\|w^n\|}$$

Nous allons montrer que l'on peut extraire de la double suite v_k^n une suite $u_n = v_{k(n)}^n$ qui tend vers v et telle que

$$\lim_{n\to+\infty} \frac{u_n - v}{\|u_n - v\|} = \frac{w}{\|w\|}$$

Soit ε_n une suite décroissante de nombres réels positifs qui tend vers 0. Pour tout n il existe un entier $k(n)$ tel que

$$\|v^n_{k(n)} - v\| \leq \varepsilon_n, \quad \left\| \frac{v^n_{k(n)} - v}{\|v^n_{k(n)} - v\|} - \frac{w^n}{\|w^n\|} \right\| \leq \varepsilon_n.$$

ce qui prouve que la suite u_n tend vers v, et par l'inégalité triangulaire que

$$\frac{v^n_{k(n)} - v}{\|v^n_{k(n)} - v\|} - \frac{w}{\|w\|} \to 0.$$

■

Le cône des directions admissibles va permettre de donner une condition nécessaire d'optimalité.

Théorème 2.7 *Si J a un minimum local en $u \in K$ et si J est différentiable en u, alors $J'(u) \in K(u)^\star$.*

PROOF Soit $w \in K(u)$, et u_k une suite telle que

$$\lim_{k\to+\infty} u_k = u, \text{ et } \lim_{k\to+\infty} \frac{u_k - u}{\|u_k - u\|} = \frac{w}{\|w\|}.$$

Puisque u est un point de minimum local, pour k suffisamment grand, $J(u_k) \geqslant J(u)$. Ecrivons d'autre part la formule de Taylor-Young à l'ordre 1 pour u_k et u.

$$J(u_k) - J(u) = J'(u) \cdot (u_k - u) + \epsilon(u_k - u)\|u_k - u\|, \quad \lim_{w\to 0} \epsilon(w) = 0.$$

Divisons par $\|u_k - u\|$:

$$0 \leq \frac{J(u_k) - J(u)}{\|u_k - u\|} = J'(u) \cdot \frac{u_k - u}{\|u_k - u\|} + \epsilon(u_k - u).$$

Passons à la limite en k :

$$0 \leq J'(u) \cdot \frac{w}{\|w\|}.$$

Ceci prouve que $J'(u) \cdot w \geqslant 0$ pour tout $w \in K(u)$, ce qui est juste dire que $J'(u) \in K(u)^*$. ■

Nous allons pouvoir en déduire les mêmes inégalités que dans le cas où K est un cône, à l'aide du lemme de Farkas.

2.3.1 Contraintes égalités

$$K = \{v \in V, F_1(v) = F_2(v) = \cdots = F_m(v) = 0\} \tag{2.11}$$

Les fonctions $F_1, .., F_m$ sont $\mathcal{C}^1 : V \to \mathbb{R}$. On note F l'application de V dans $\mathbb{R}^m$ définie par les F_i, soit $F(v) = (F_1(v), F_2(v), \ldots, F_m(v)) \in \mathbb{R}^m$.

Définition 2.1 *Les contraintes sont régulières en $u \in K$ si les $F_i'(u)$ sont linéairement indépendantes. On dit alors que u est un point régulier.*

On peut alors caractériser le cône des directions admissibles :

Lemme 2.1 *Si les contraintes sont regulières en $u \in K$, alors*

$$K(u) = \{w \in V, F_i'(u).w = 0, 1 \leq i \leq m\} \tag{2.12}$$

Ici en fait le cône est un espace vectoriel, l'orthogonal de l'espace engendré par les $F_i'(u)$. PROOF [of the lemma] Notons $\tilde{K}(u) = \{w \in V, F_i'(u).w = 0, 1 \leq i \leq m\}$. Nous allons montrer que $\tilde{K}(u)$ coincide avec $K(u)$.

$K(u) \subset \tilde{K}(u)$. Soit $w \in K(u)$, u_k la suite associée, et nous appliquons Taylor-Lagrange à l'ordre 1 à chaque F_i :

$$F_i(u_k) = F_i(u) + F_i'(u + \theta(u_k - u)) \cdot (u_k - u)$$

Puisque u et u_k sont dans K, $F_i(u) = F_i(u_k) = 0$, et nous obtenons

$$F_i'(u + \theta(u_k - u)) \cdot (u_k - u) = 0 = F_i'(u + \theta(u_k - u)) \cdot \frac{u_k - u}{\|u_k - u\|}.$$

Il suffit maintenant de passer à la limite en k pour obtenir $F_i'(u) \cdot w = 0$ et donc $w \in \tilde{K}(u)$.

$\tilde{K}(u) \subset K(u)$. Soit $w \in \tilde{K}(u)$, c'est-à-dire orthogonal à l'espace vectoriel engendré par les $F_i'(u)$. Pour tout v dans V on peut écrire

$$v = u + \lambda w + \sum_{i=1}^{m} \alpha_i F_i'(u).$$

Définissons les applications $g_j : \mathbb{R} \times \mathbb{R}^m \to \mathbb{R}$, et $g : \mathbb{R} \times \mathbb{R}^m \to \mathbb{R}^m$, par

$$\alpha = (\alpha_1, \ldots, \alpha_m), \quad g_j(\lambda, \alpha) = F_j(v), \quad g(\lambda, \alpha) = F(v).$$

Nous allons appliquer à la fonction v le théorème des fonctions implicites. D'abord $g(0,0) = F(u) = 0$. Ensuite

$$D_\alpha g(\lambda, \alpha) = \left(\frac{\partial g_j}{\partial \alpha_i}\right)_{1 \leq i,j \leq m} (\lambda, \alpha).$$

Calculons ces quantités :

$$\frac{\partial g_j}{\partial \alpha_i}(\lambda, \alpha) = F_j'(v)\frac{\partial v}{\partial \alpha_i}(\lambda, \alpha) = F_j'(v)F_i'(u), \quad v = u + \lambda w + \sum_{i=1}^{m} \alpha_i F_i'(u).$$

si bien que $\frac{\partial g_j}{\partial \alpha_i}(0,0) = F_j'(u)F_i'(u)$. Puisque les contraintes sont régulières, alors la matrice $D_\alpha g(0,0)$ est inversible. En résumé la fonction g est $\mathcal{C}^1$, $g(0,0) = 0$, la dérivée par rapport à la seconde varaible est inversible. Alors il existe ε_1 et ε_2 et une fonction $\alpha :]-\varepsilon_1, \varepsilon_1[\to \mathbb{R}^m$ tels que pour $(\lambda, \alpha) \in]-\varepsilon_1, \varepsilon_1[\times B_{\varepsilon_2}$,

$$g(\lambda, \alpha) = 0 \iff \alpha = \alpha(\lambda), \text{ de plus } \alpha'(0) = -(D_\alpha g(0,0))^{-1} D_\lambda g(0,0).$$

Calculons

$$D_\lambda g(0,0) = (D_\lambda g_1(0,0), \ldots, D_\lambda g_m(0,0)) = (F_1'(u)w, \ldots, F_m'(u)w) = 0.$$

Donc $\alpha'(0) = 0$. Soit maintenant ε_n une suite décroissante tendant vers 0. Pour tout $\lambda = \varepsilon_n$, définissons

$$v_n = u + \varepsilon_n w + \sum_{i=1}^{m} \alpha_i(\varepsilon_n) F_i'(u).$$

Par construction, $F(v_n) = g(\varepsilon_n, \alpha(\varepsilon_n)) = 0$ pour n assez grand, donc $v \in K$. D'autre part puisque $\alpha(0) = 0$, $v_n \to u$ pour $n \to \infty$. Pour finir

$$\frac{v_n - u}{\|v_n - u\|} = \frac{\varepsilon_n w + \sum_{i=1}^{m} \alpha_i(\varepsilon_n) F_i'(u)}{\|\varepsilon_n w + \sum_{i=1}^{m} \alpha_i(\varepsilon_n) F_i'(u)\|} = \frac{w + \sum_{i=1}^{m} \frac{\alpha_i(\varepsilon_n)}{\varepsilon_n} F_i'(u)}{\|w + \sum_{i=1}^{m} \frac{\alpha_i(\varepsilon_n)}{\varepsilon_n} F_i'(u)\|}$$

Pour tout i, $\frac{\alpha_i(\varepsilon_n)}{\varepsilon_n} \to \alpha_i'(0) = 0$, et donc

$$\frac{v_n - u}{\|v_n - u\|} \to \frac{w}{\|w\|}.$$

Tout ceci montre bien que $w \in K(u)$ et le lemme est démontré.

■

Nous en déduisons l'existence de **multiplicateurs de Lagrange** :

Théorème 2.8 *Si $u \in K$, u régulier, est minimum local pour J, il existe m réels $p_1, .., p_m$ tels que*

$$J'(u) + \sum_{i=1}^{m} p_i F_i'(u) = 0. \tag{2.13}$$

PROOF [of the theorem] Par le théorème 2.7, nous savons que $J'(u) \in K(u)^*$, c'est-à-dire que pour tout w dans $K(u)$, $(J'(u), w) \geqslant 0$. Par le lemme précédent, $K(u)$ est en fait un espace vectoriel : si $w \in K(u)$, $-w \in K(u)$. Donc pour tout w dans $K(u)$, $(J'(u), w) = 0$: $J'(u)$ est dans l'othogonal de $K(u)$ qui est lui même l'orthogonal de $\mathrm{vec}(F_1'(u), \ldots, F_m'(u))$. Donc $J'(u) \in \mathrm{vec}(F_1'(u), \ldots, F_m'(u))$. ■

Remarque 2.3 . *Si K et J sont convexes, alors c'est une condition nécessaire et suffisante.*

Théorème 2.9 *Supposons que J est convexe différentiable et les fonctions F_i sont convexes. Alors $u \in K$, u régulier, est minimum local pour J, si et seulement si il existe m réels $p_1, .., p_m$ tels que*

$$J'(u) + \sum_{i=1}^{m} p_i F_i'(u) = 0. \tag{2.14}$$

PROOF Il ne nous reste plus qu'à prouver le seulement si : si $J'(u) \in \mathrm{vec}(F_1'(u), \ldots, F_m'(u))$, alors u est minimum local. Puisque les F_i sont convexes, K est convexe : si v et w sont dans K, cela signifie que pour tout i, $F_i(v) = F_i(w) = 0$. Alors

$$F_i(\theta v + (1-\theta)w) = \theta F_i(v) + (1_\theta) F_i(w) = 0,$$

et $\theta v + (1-\theta)w \in K$.

Remarquons comme une évidence que pour tout v dans K, $w = v - u \in K(u)$. Donc par le lemme 2.1, w est orthogonal aux vecteurs $F_i'(u)$, et donc à $J'(u)$. On a donc,

$$\forall v \in K, J'(u) \cdot (v - u) = 0,$$

Ce qui est plus fort que l'inéquation d'Euler dans le théorème 2.3.

■

Introduisons le lagrangien défini sur $V \times \mathbb{R}^m$ à valeurs dans $\mathbb{R}$ par

$$\mathcal{L}(v, q) := J(v) + \sum_{i=1}^{m} q_i F_i(v), \tag{2.15}$$

alors

$$\begin{array}{c} D_v \mathcal{L}(v, q) = J'(v) + \sum_{i=1}^{m} q_i F_i'(v) \\ D_q \mathcal{L}(v, q) = F(v) \end{array} \tag{2.16}$$

et

$$\begin{array}{rcl} u \in K & \iff & \forall q \in \mathbb{R}^m,\ D_v \mathcal{L}(u, q) = 0 \\ u \text{ minimum local} & \iff & \exists p \in \mathbb{R}^m,\ D_q \mathcal{L}(u, p) = 0. \end{array} \tag{2.17}$$

p s'appelle la variable duale de u.

2.3.2 Contraintes inégalités

$$K = \{v \in V, F(v) \leq 0\} \tag{2.18}$$

où F est une fonction C^1 de V dans $\mathbb{R}^m$, ses coordonnées sont $F_1, .., F_m$.

Définition 2.2 *Pour $u \in K$, on appelle $I(u)$ l'ensemble des contraintes actives ou saturées, i.e.$F_i(u) = 0$ si $i \in I(u)$, $F_i(u) < 0$ sinon. Les contraintes sont dites qualifiées en u si*

$$\exists \bar{w} \in V, \forall i \in I(u), (F_i'(u), \bar{w}) < 0 \ (\leq 0 \text{ si } F_i \text{ est affine}). \tag{2.19}$$

Remarque 2.4 *SI toutes les contraintes sont affines, alors $\bar{w} = 0$ convient : les contraintes sont qualifiées en tout point.*

On peut encore caractériser le cône des directions admissibles :

Lemme 2.2 *Si les contraintes sont qualifiées en $u \in K$, alors*

$$K(u) = \{w \in V, \forall i \in I(u), F_i'(u).w \leq 0\} \tag{2.20}$$

PROOF Notons $\tilde{K}(u) = \{w \in V, \forall i \in I(u), F_i'(u).w \leq 0\}$.

$K(u) \subset \tilde{K}(u)$ Soit $w \in K(u)$, u_n une suite d'éléments de K convergeant vers u. Soit $i \in I(u)$. Alors $F_i(u) = 0$ et $F_i(u_n) \leq 0$. Ecrivons la formule de Taylor -Young à l'ordre 1 pour F_i au voisinage de u :

$$F_i(u_n) = \underbrace{F_i(u)}_{0} + F_i'(u) \cdot (u_n - u) + \epsilon(u_n - u)\|u_n - u\| \ \leq 0$$

On divise maintenant par $\|u_n - u\|$:

$$F_i'(u) \cdot \frac{u_n - u}{\|u_n - u\|} + \epsilon(u_n - u) \ \leq 0$$

et on passe à la limite, ce qui donne $F_i'(u) \cdot w \leq 0$. Donc $w \in \tilde{K}(u)$.

$\tilde{K}(u) \subset K(u)$ Soit $w \in \tilde{K}(u)$. Nous allons montrer que pour tout $\delta \geqslant 0$, $w + \delta\bar{w} \in K(u)$.

- Si $w + \delta\bar{w} = 0$, puisque $0 \in K(u)$, c'est vrai.
- Si $w + \delta\bar{w} \neq 0$, pour ε_n une suite décroissante tendant vers 0, posons

$$u_n = u + \varepsilon_n(w + \delta\bar{w}).$$

Il est clair que $u_n \to u$. Montrons que $u_n \in K$.

- Soit $i \in I(u)$, donc $F_i(u) = 0$.
  - Si F_i est affine, alors la formule de Taylor à l'ordre 1 est exacte

$$F_i(u_n) = F_i(u) + \varepsilon_n F_i'(u)(w + \delta\bar{w}) = \varepsilon_n F_i'(u)(w + \delta\bar{w}) \leq 0.$$

  - Si F_i n'est pas affine,

$$F_i(u_n) = F_i(u) + \varepsilon_n(F_i'(u)(w + \delta\bar{w}) + \epsilon(\varepsilon_n)\|w + \delta\bar{w}\|)$$

Ici $F_i'(u)(w + \delta\bar{w}) < 0$, et donc pour n assez grand le terme en facteur de ε_n est ≤ 0.

Donc pour $i \in I(u)$, $F_i(u_n) \leq 0$ pour n suffisamment grand.

- $\diamond$ Soit $i \notin I(u)$, alors $F_i(u) < 0$. Par la formule de Taylor ci-dessus, $F_i(u_n) \leq 0$ pour n suffisamment grand.

Nous avons donc montré que $u_n \in K$ pour n assez grand.
Maintenant

$$\frac{u_n - u}{\|u_n - u\|} = \frac{w + \delta\bar{w}}{\|w + \delta\bar{w}\|}$$

Tous ces éléments montrent que $w + \delta\bar{w} \in K(u)$. En faisant tendre δ vers 0, puisque $K(u)$ est fermé, nous obtenons que $w \in K(u)$.

■

Le lemme de Farkas permet alors d'établir l'existence de multiplicateurs de Lagrange :

Théorème 2.10 *Si $u \in K$, où les contraintes sont qualifiées, est minimum local pour J, il existe m réels $p_1, .., p_m \geqslant 0$ tels que*

$$\begin{gathered} J'(u) + \sum_{i=1}^{m} p_i F_i'(u) = 0 \\ \sum_{i=1}^{m} p_i F_i(u) = 0 \text{ ou encore } F_i(u) < 0 \implies p_i = 0. \end{gathered} \tag{2.21}$$

PROOF D'après le théorème 2.7, pour tout v dans K, $J(u) \in K(u)^*$. D'après le lemme de Farkas,

$$J'(u) = -\sum_{i \in I(u)} \lambda_i F_i'(u), \quad \lambda_i \geqslant 0.$$

ce qui est équivalent à (2.21). ■

Remarque 2.5 . *Le lagrangien est maintenant défini sur $V \times \mathbb{R}_+^m$, et l'on peut écrire*

$$\begin{gathered} u \in K \text{ solution optimale } \Rightarrow \exists p \in \mathbb{R}_+^m, \\ \mathcal{L}_v'(u,p) = \mathcal{L}_q'(u,p).p = 0. \end{gathered} \tag{2.22}$$

Attention, contrairement au cas des contraintes égalités, on n'a qu'une condition nécessaire. Le développement d'une condition nécéssaire et suffisante est l'objet du chapitre suivant.

Chapitre 3

Lagrangien et point selle

Sommaire

3.1 Point selle

Soient V et M deux espaces de Hilbert, U une partie de V et P une partie de M. On définit le lagrangien comme une application de $U \times P$ dans $\mathbb{R}$ et on le note $\mathcal{L}$.

Exemple 3.1 *au problème d'optimisation du chapitre précédent,*

$$\begin{cases} u \in K, \\ J(u) = \inf\limits_{v \in K} J(v) \end{cases} \tag{3.1}$$

nous avons associé de façon naturelle un lagrangien dans les cas suivants :

$$\begin{array}{l} K = \{v, F(v) \leq 0\} \; ; \; \mathcal{L} : K \times \mathbb{R}^m_+ \to \mathbb{R} \\ K = \{v, F(v) = 0\} \; ; \; \mathcal{L} : K \times \mathbb{R}^m \to \mathbb{R} \end{array} \tag{3.2}$$

où $F : V \to \mathbb{R}^m$, et

$$\mathcal{L}(v, q) = J(v) + (F(v), q) \tag{3.3}$$

(.,.) désigne le produit scalaire dans $\mathbb{R}^m$.

Lemme 3.1

$$\sup_{q \in P} \inf_{v \in U} \mathcal{L}(v, q) \leq \inf_{v \in U} \sup_{q \in P} \mathcal{L}(v, q) \tag{3.4}$$

Remarquons que l'on n'interdit pas les valeurs $+\infty$ et $-\infty$.

PROOF Pour tout $(u, p) \in U \times P$,

$$G(p) := \inf_{v \in U} \mathcal{L}(v, p) \; \leq \; \mathcal{L}(u, p) \; \leq \; \sup_{q \in P} \mathcal{L}(u, q) := J(u). \tag{3.5}$$

Oublions le terme du milieu, et écrivons,

$$\forall u \in U, \forall p \in P, \quad G(p) \leq \; J(u).$$

ce qui implique que

$$\forall u \in U, \sup_{p\in P} G(p) \leq J(u),$$

puis que

$$\sup_{p\in P} G(p) \leq \inf_{u\in U} J(u),$$

Reportons ensuite les définitions de G et J dans les formules pour obtenir (3.4). Attention il est très important de mettre toutes ces définitions dans le bon ordre. ■

Définition 3.1 *(u, p) est point selle du lagrangien si*

$$\sup_{q\in P} \mathcal{L}(u, q) = \mathcal{L}(u, p) = \inf_{v\in U} \mathcal{L}(v, p) \tag{3.6}$$

Lemme 3.2 *Si (u, p) est point selle du lagrangien, alors*

$$\sup_{q\in P} \inf_{v\in U} \mathcal{L}(v, q) = \mathcal{L}(u, p) = \inf_{v\in U} \sup_{q\in P} \mathcal{L}(v, q) \tag{3.7}$$

PROOF Il est très facile d'obtenir à partir de la définition que

$$\inf_{v\in U} \sup_{q\in P} \mathcal{L}(v, q) \leq \mathcal{L}(u, p) \leq \sup_{q\in P} \inf_{v\in U} \mathcal{L}(v, q).$$

Le lemme 3.1 donne

$$\sup_{q\in P} \inf_{v\in U} \mathcal{L}(v, q) \leq \inf_{v\in U} \sup_{q\in P} \mathcal{L}(v, q).$$

ce qui prouve l'égalité. ■

On associe maintenant au lagrangien un problème primal et un problème dual. On définit d'une part K et J par

$$K = \{v \in U, \sup_{q\in P} \mathcal{L}(v, q) < +\infty\},$$

et pour v dans K,

$$J(v) = \sup_{q\in P} \mathcal{L}(v, q).$$

Le problème primal associé s'écrit :

$$(\mathcal{P}) \text{ Trouver } u \in K \text{ tel que } J(u) = \inf_{v\in K} J(v).$$

On définit également K^* et G par $K^* = \{q \in P, \inf_{v\in U} \mathcal{L}(v, q) > -\infty\}$, et pour q dans K^*, $G(q) = \inf_{v\in U} \mathcal{L}(v, q)$. Le problème dual associé s'écrit :

$$(\mathcal{P}^*) \text{ Trouver } p \in K^* \text{ tel que } G(p) = \sup_{q\in K^*} G(q)$$

Théorème 3.1 *(u, p) est point selle du lagrangien si et seulement si u est solution de $(\mathcal{P})$, p est solution de $(\mathcal{P}^*)$, et $J(u) = G(p)$.*

PROOF

$\Longrightarrow$ Introduisons les fonctions J et G dans la définition du point selle et le lemme 3.2

$$\mathcal{L}(u, p) = J(u) = G(p),$$

$$\mathcal{L}(u, p) = \sup_{q \in P} G(q) \ = \ \inf_{v \in U} J(v)$$

$\Longleftarrow$ Réécrivons (3.5)

$$G(p) := \inf_{v \in U} \mathcal{L}(v, p) \ \leq \ \mathcal{L}(u, p) \ \leq \ \sup_{q \in P} \mathcal{L}(u, q) := J(u).$$

Si $G(p) = J(u)$, alors on a l'égalité

$$\inf_{v \in U} \mathcal{L}(v, p) \ = \ \mathcal{L}(u, p) \ = \ \sup_{q \in P} \mathcal{L}(u, q),$$

ce qui est juste la définition du point selle.

■

3.2 Théorie de Kuhn et Tucker

On considère maintenant le problème de minimisation convexe avec contraintes inégalité :

$$K = \{v \in V, F(v) \leq 0\} \tag{3.8}$$

où F est une fonction convexe C^1 de V dans $\mathbb{R}^m$, ses coordonnées sont $F_1, .., F_m$. On suppose J convexe et on définit le lagrangien sur $V \times \mathbb{R}^m_+$ par

$$\mathcal{L}(v, q) = J(v) + (F(v), q) \tag{3.9}$$

On a vu au chapitre précédent une condition nécessaire de minimum local, au moyen des multiplicateurs de Lagrange. On va maintenant établir une réciproque, avec une autre définition de la qualification des contraintes.

Définition 3.2 *Les contraintes sont qualifiées si*

$$\exists \bar{v} \in V, \forall i, 1 \leq i \leq m, F_i(\bar{v}) < 0 \textit{ (resp. } \leq 0 \textit{ si } F_i \textit{ est affine)}. \tag{3.10}$$

Si les contraintes sont qualifiées en ce sens, elles sont qualifiées en tout point au sens de la définition 2.2 du chapitre 2. En effet soit $u \in K$.

— Si $u = \bar{v}$, Pour toute contrainte saturée, $F_i(u) = 0$, soit $F_i(\bar{v}) = 0$, et donc F_i est affine. Toutes les contraintes saturées sont affines, et $\bar{w} = 0$ convient.

— Si $u \neq \bar{v}$, Pour toute contrainte saturée, $F_i(u) = 0$, et par la caractérisation des fonctions convexes au théorème 1.5, on peut écrire

$$F_i(\bar{v}) \geqslant \underbrace{F_i(u)}_{=0} + F_i'(u) \cdot (\bar{v} - u).$$

— Si $F_i(\bar{v}) < 0$, cela implique que $F_i'(u) \cdot (\bar{v} - u) < 0$.

— Si $F_i(\bar{v}) = 0$, donc si F_i est affine, cela implique que $F_i'(u) \cdot (\bar{v} - u) \leq 0$.

Donc $\bar{w} = \bar{v} - u$ convient.

Remarque 3.1 . *Si aucune des F_i n'est affine, la définition 3.2 se résume à $\overset{\circ}{K} \neq \emptyset$. Si toutes les F_i sont affines, elle signifie que $K \neq \emptyset$.*

Théorème 3.2 *Sous les hypothèses de qualification de la définition 3.2, si u est solution de $(\mathcal{P})$, il existe p dans $\mathbb{R}_+^m$ tel que (u, p) soit point selle du lagrangien.*

PROOF Si les contraintes sont qualifiées au sens de la définition 3.2, elles le sont au sens de la définition 2.2, et on peut appliquer le théorème 2.10. Il existe $u \in V$ et $p \in \mathrm{Pr}_+^m$ tel que

$$\begin{aligned}(F(u), p) = 0 = \mathcal{L}_q'(u, p) \cdot p,\\ J'(u) + \sum_{i=1}^{m} \lambda_i F_i'(u) = 0 = \mathcal{L}_v'(u, p).\end{aligned}$$

Nous devons maintenant montrer que (u, p) est point selle du Lagrangien.

La fonction $v \mapsto \mathcal{L}(v, p)$ est convexe sur V, appliquons lui la caractérisation par le plan tangent :

$$\forall v \in K, \mathcal{L}(v, p) \geqslant \mathcal{L}(u, p) + \underbrace{\mathcal{L}_v'(u, p)}_{=0} \cdot (v - u).$$

et donc $\mathcal{L}(u, p) = \inf_{v \in K} \mathcal{L}(v, p)$.

La fonction $q \mapsto \mathcal{L}(u, q)$ est affine sur $\mathbb{R}_+^m$, donc

$$\begin{aligned}\forall q \in \mathbb{R}_+^m, \mathcal{L}(u, q) &= \mathcal{L}(u, p) + \mathcal{L}_q'(u, p) \cdot (q - p),\\ &= \mathcal{L}(u, p) + \mathcal{L}_q'(u, p) \cdot q,\\ &= \mathcal{L}(u, p) + \underbrace{(F(u), q)}_{\leq 0},\end{aligned}$$

et donc $\mathcal{L}(u, p) = \sup_{q \in \mathbb{R}_+^m} \mathcal{L}(u, q)$, ce qui est la définition du point selle. ■

Donc **dans le cas convexe**, avec l'hypothèse de **qualification des contraintes de la définition 3.2**, on a le schéma suivant :

$$u \text{ solution optimale} \overset{\text{(Th ??)}}{\Longrightarrow} \exists p \in \mathbb{R}_+^m \begin{cases} J'(u) + \sum_{i=1}^{m} p_i F_i'(u) = 0 \\ \sum_{i=1}^{m} p_i F_i(u) = 0 \end{cases}$$

$$\overset{\text{(Th 3.2)}}{\Longrightarrow} (u, p) \text{ point selle du lagrangien } \overset{\text{(Th 3.1)}}{\Longrightarrow} u \text{ solution optimale de (1.2).}$$

La forme opérationnelle est la suivante.

Théorème 3.3 *[Karush, Kuhn et Tucker].*

On suppose que les fonctions J et $\{F_i\}_{1\leq i\leq m}$ sont convexes différentiables et que (3.10) est vérifiée. Soit

$$K = \{v, F_i(v) \leq 0, 1 \leq i \leq m\}.$$

Alors u est minimum de J sur K si et seulement si il existe p dans $\mathbb{R}^m_+$ tel que

$$\begin{cases} J'(u) + \displaystyle\sum_{i=1}^{m} p_i F_i'(u) = 0 \\ \displaystyle\sum_{i=1}^{m} p_i F_i(u) = 0 \end{cases} \tag{3.11}$$

De plus p est solution du problème dual $(\mathcal{P}^*)$.

Retour sur le problème dual. Soit $\mathcal{L}(v, q) = J(v) + (F(v), q)$. On a posé

$$G(q) = \inf_{v\in U} \mathcal{L}(v, q) = \mathcal{L}(u_q, q).$$

Alors $G'(q) = F(u_q)$. En effet soit $\delta q \in \mathbb{R}^m$. Alors

$$G(q + \delta q) = \mathcal{L}(u_{q+\delta q}, q + \delta q) \ \leq \ \mathcal{L}(u_q, q + \delta q),$$

$$G(q) = \mathcal{L}(u_q, q) \ \leq \ \mathcal{L}(u_{q+\delta q}, q).$$

Donc

$$\mathcal{L}(u_{q+\delta q}, q + \delta q) - \mathcal{L}(u_{q+\delta q}, q) \ \leq \ G(q + \delta q) - G(q)| \ \leq \ \mathcal{L}(u_q, q + \delta q) - \mathcal{L}(u_q, q).$$

Remplaçons $\mathcal{L}$ par sa valeur

$$(F(u_{q+\delta q}), \delta q) \leq \ G(q + \delta q) - G(q)| \ \leq \ (F(u_q), \delta q).$$

Si l'on a pris la précaution de montrer que $q \to u_q$ est une application continue, en faisant tendre δq vers 0 on obtient

$$G'(q) \cdot \delta q = (F(u_q), \delta q) \ \implies \ G'(q) = F(u_q).$$

Cela nous sera utile pour l'algorithme d'Uzawa.

Deuxième partie

Algorithmes

Chapitre 4

Méthodes de descente. Problèmes sans contraintes

Sommaire

4.1 Principe

On se place dans un espace de Hilbert V, et on cherche à calculer numériquement un x (qui n'est pas forcément unique) tel que

$$\forall y \in V, J(x) \leqslant J(y) \tag{4.1}$$

Le principe est de construire un algorithme itératif de la forme

$$x^{k+1} = x^k - \rho_k d^k \tag{4.2}$$

d^k est la **direction de descente**, ρ_k est le **pas**. Il est, soit fixé, éventuellement le même pour toutes les étapes (on parle alors de **méthode à pas variable**), soit calculé à chaque étape de façon à minimiser J dans la direction d^k (on parle alors de **méthode à pas optimal**).

4.2 Méthode de relaxation

On se place en dimension finie, *i.e.* $V = \mathbb{R}^n$. Pour passer de x^k à x^{k+1}, on minimise successivement dans les n directions de la base canonique.

1. $x^{k,1}$ est défini par
$$J(x^{k,1}) = \inf_{\rho \in \mathbb{R}} J(x^k - \rho e_1)$$
ou encore
$$x^{k,1} = (x_1^k - \rho_1, x_2^k, .., x_n^k)$$
On note $x_1^{k+1} = x_1^k - \rho_1$
2. à l'étape i on a
$$x^{k,i} = (x_1^{k+1}, .., x_i^{k+1}, x_i^k, .., x_n^k)$$
$x^{k,i+1}$ est maintenant défini par
$$J(x^{k,i+1}) = \inf_{\rho} J(x^{k,i} - \rho e_{i+1})$$
3. $x^{k+1} = x^{k,n}$

Théorème 4.1 . *Si J est α-convexe différentiable sur $\mathbb{R}^n$, si J' est uniformément lipscitzienne sur les bornés, l'algorithme de relaxation est bien défini et converge vers la solution optimale.*

Remarque 4.1 . *Dans le cas où J est quadratique, i.e. $J(v) = \frac{1}{2}(Av, v) - (b, v)$, on retrouve l'algoritme de Gauss-Seidel ou S.O.R. pour la résolution du système linéaire $Ax = b$.*

4.3 Méthode du gradient

Ici on choisit à chaque étape $d^k = \nabla J(x^k)$.

4.3.1 Méthode à pas variable

On se donne le pas ρ_k, il peut être différent d'une étape à l'autre.

Théorème 4.2 . *Supposons J α-convexe dérivable sur V, et ∇J uniformément lipschitzien de constante de Lipschitz M. Si il existe deux constantes a et b telles que pour tout $k > 0$, $0 < a \leq \rho_k \leq b < \frac{2\alpha}{M^2}$, l'algorithme de gradient à pas variable converge vers la solution optimale.*

Remarque 4.2 . *Si J est 2 fois différentiable, l'hypothèse est*
$$\sup_{v \in V} ||D^2 J(v)|| \leq M$$

4.3.2 Méthode à pas optimal

Ici on choisit à chaque étape ρ_k de façon que
$$J(x^k - \rho_k \nabla J(x^k)) = \inf_{\rho \in \mathbb{R}} J(x^k - \rho \nabla J(x^k)) \tag{4.3}$$

Théorème 4.3 . *Si J est α-convexe dérivable sur V, si ∇J est uniformément lipschitzien de constante de Lipschitz M, l'algorithme de gradient à pas optimal est bien défini et converge vers la solution optimale.*

Remarque 4.3 . *Les directions de descente sont orthogonales, i.e.*
$$\nabla J(x^k).\nabla J(x^{k+1}) = 0.$$

4.4 Estimations et convergence dans le cas quadratique

Ici la fonctionnelle J est quadratique sur $\mathbb{R}^n$:

$$J(v) = \frac{1}{2}(Av, v) - (b, v)$$

où la matrice A est symétrique définie positive. La solution x du problème de minimisation vérifie $Ax = b$. On appellera **résidu** à l'étape k la quantité $r^k = b - Ax^k$.

4.4.1 Méthode à pas optimal

On prend ici une direction de descente d^k quelconque dans $\mathbb{R}^n$, non orthogonale à r^k. A chaque étape, la valeur du paramètre optimal ρ_k est donnée par

$$\rho_k = -\frac{(r^k, d^k)}{(Ad^k, d^k)} \tag{4.4}$$

et l'on a $(r^{k+1}, d^k) = 0$.
Notons $E(v) = \frac{1}{2}(A(v-u), v-u)$, on a alors

$$E(x^{k+1}) = (1 - \gamma_k)E(x^k) \tag{4.5}$$

avec

$$\gamma_k = \frac{(r^k, d^k)^2}{(Ad^k, d^k)(A^{-1}r^k, r^k)}. \tag{4.6}$$

Par construction $0 \leq \gamma_k \leq 1$. De plus, notant λ_j les valeurs propres de A, ordonnées, $\lambda_1 \leq \lambda_2 \leq \ldots \lambda_n$, on a l'estimation suivante :

$$\gamma_k \geqslant \frac{\lambda_1}{\lambda_n}\left(\frac{r^k}{\|r^k\|}, \frac{r^k}{\|r^k\|}\right)^2$$

si la direction de descente est telle que

$$\left(\frac{r^k}{||r^k||}, \frac{d^k}{||d^k||}\right)^2 \geqslant \mu > 0 \tag{4.7}$$

alors $\gamma_k \geqslant \gamma = \dfrac{\mu}{K(A)}$ (où $K(A)$ est le conditionnement de A pour la norme euclidienne, c'est-à-dire le rapport de la plus grande à la plus petite valeur propre), et donc

$$E(x^{k+1}) \leq (1 - \gamma)E(x^k) \tag{4.8}$$

On dit que la méthode **converge linéairement**.

Dans le cas particulier de la méthode du gradient, $d^k = \nabla J(x^k)$, grâce à **l'inégalité de Kantorovitch** on peut écrire

$$\tau = \frac{K(A)-1}{K(A)+1}, \quad E(x) = \frac{1}{2}(A(x - E(x^{k+1}) \leq \tau^2 E(x^k),\ E(x^k) \leq \left(\frac{K(A)-1}{K(A)+1}\right)^{2k} E(x^0) \tag{4.9}$$

Remarque 4.4 . *Plus la matrice est bien conditionnée (i.e. $K(A)$ proche de 1), plus la convergence est rapide. Plus la matrice est mal conditionnée (i.e. $K(A)$ »1), plus la convergence est lente.*

Lemme 4.1 (Lemme de Kantorovitch) *Pour toute matrice A symétrique définie positive, on a l'inégalité :*

$$\forall y \in \mathbb{R}^m \backslash 0, \quad \frac{(Ay,y)(A^{-1}y,y)}{(y,y)^2} \leq \frac{(\lambda_1+\lambda_n)^2}{4\lambda_1\lambda_n} = \frac{((K(A)^{\frac{1}{2}}+K(A)^{-\frac{1}{2}})^2}{4}.$$

4.4.2 Méthode de gradient à pas constant

On choisit à chaque étape $\rho_k = \rho$. On a alors l'estimation

$$||x^k - x||_2 \leq \left[\max_{1\leq i\leq n} |1-\rho\lambda_i|\right]^k ||x^0 - x||_2 \tag{4.10}$$

On en déduit que la méthode converge si et seulement si $\rho < \dfrac{2}{\lambda_n}$ où λ_n est la plus grande valeur propre de A. Ici encore, la convergence est linéaire.

Le meilleur ρ est égal à $\frac{2}{\lambda_1+\lambda_b}$.

4.5 Méthode du gradient conjugué

On se place ici dans le cas où la fonctionnelle J est quadratique sur $\mathbb{R}^N$: $J(v) = \frac{1}{2}(Av,v) - (b,v)$, la matrice A étant symétrique définie positive. La solution x du problème de minimisation vérifie $Ax = b$.

4.5.1 Principe de la méthode

Les $(k+1)$ premières itérées $x^0, .., x^k$ étant données, on cherche x^{k+1}, non plus dans la direction du gradient, mais dans l'espace vectoriel engendré par tous les gradients précédents. On note

$$\mathcal{L}_k = vect\{\nabla J(x^0), .., \nabla J(x^k)\} \tag{4.11}$$

et on définit x^{k+1} par :

$$J(x^{k+1}) = \inf_{\Delta \in L_k} J(x^k + \Delta) \tag{4.12}$$

Ceci définit $x^{k+1} = x^k + \Delta^k$ de manière unique (*cf* Corollaire 1.1, Partie I) et

Théorème 4.4 . *On a les propriétés suivantes :*

1. *Les $\nabla J(x^k)$ forment un système orthogonal (donc libre),*
2. *l'algorithme converge en au plus N itérations.*

La première propriété traduit l'équation d'Euler (2.3, Partie I). Ce théorème nous dit que la méthode du gradient conjugué est en fait une méthode directe. La forme (4.12) n'est pas pratique, aussi allons nous réécrire l'algorithme sous forme d'un algorithme de descente.

4.5.2 Ecriture comme algorithme de descente

Lemme 4.2 *On a les propriétés suivantes*

1. *Pour tout $\ell \neq 0$, $\Delta_\ell \neq 0$,*
2. *Les directions Δ_ℓ sont conjuguées par rapport à A, i.e.*

$$\forall \ell, m, \quad \ell \neq m, \quad (A\Delta_\ell, \Delta_m) = 0.$$

On développe maintenant les directions Δ_ℓ dans la base des résidus

$$\begin{aligned}\Delta^0 &= \delta_0^0 r^0,\\ &\vdots\\ \Delta^k &= \delta_k^k r^k + \ldots + \delta_k^0 r^0,\\ &\vdots\end{aligned}$$

et on calcule les coefficients

Théorème 4.5 . *L'algorithme du gradient conjugué s'écrit sous la forme*

$$\begin{cases} x^{k+1} = x^k - \rho_k d^k \\ d^k = \nabla J(x^k) + \dfrac{||\nabla J(x^k)||^2}{||\nabla J(x^{k-1})||^2} d^{k-1} \\ \rho_k = \dfrac{||\nabla J(x^k)||^2}{(Ad^k, d^k)} \\ (r^{k+1}, d^k) = 0 \end{cases} \tag{4.13}$$

Il suffit de se donner $d^0 = \nabla J(x^0)$.
N peut être très grand, on peut alors compter le nombre d'opérations nécessaires pour réaliser l'algorithme : une itération nécessite $2cN$ opérations élémentaires, où c est le nombre moyen de coefficients non nuls par ligne de A. Si bien que pour une matrice pleine, le nombre d'opérations élémentaires pour N itérations est $2N^3$. Cela disqualifie la méthode par rapport à Cholewski ($\frac{N^3}{3}$opérations élémentaires), si on la considère seulement comme une méthode directe.

Remarque 4.5 *On a aussi* $\rho_k = -\dfrac{(r^k, d^k)}{(Ad^k, d^k)}$*, ce qui correspond à à un gradient à pas optimal dans la direction* d^k.

4.5.3 Analyse de convergence

On introduit l'espace de Krylov

$$\mathcal{K}_k = vect\{r^0, Ar^0, .., A^k r^0\} \tag{4.14}$$

et on a le

Théorème 4.6 . *Si* $r^j \neq 0$ *pour* $j \leq k$*, alors* $\mathcal{K}_k \equiv \mathcal{L}_k$.

ON en déduit que $\dim \mathcal{K}_k = k+1$ et donc que $r^0, Ar^0, \ldots, A^k r^0$ forment un système libre. On a donc

$$J(x^{k+1}) = \inf_{v \in u^0 + \mathcal{K}_k} J(v)$$

On en déduit une première estimation de l'erreur

Théorème 4.7

$$E(x^k) = \inf_{P \in \mathbb{P}_{k-1}} \frac{1}{2} \left(A(I + AP(A))e^0, (I + AP(A))e^0 \right) \tag{4.15}$$

avec $e^0 = u^0 - u$ *et* $E(v) = \frac{1}{2}(A(v-u), v-u)$.

Par diagonalisation de A on en déduit

Corollaire 4.1

$$E(x^k) = \inf_{P\in\mathbb{P}_{k-1}} \max_{1\le i\le N} [1+\lambda_i P(\lambda_i)]^2 E(x^0) \tag{4.16}$$

où les λ_i sont les valeurs propres de A.

et par un calcul assez long sur les polynômes de Tchebycheff,

Corollaire 4.2 . *On a l'estimation d'erreur*

$$E(x^k) \le 4\Big(\frac{\sqrt{K(A)}-1}{\sqrt{K(A)}+1}\Big)^{2k} E(x^0) \tag{4.17}$$

De nouveau, la convergence est linéaire. Cette estimation est à comparer avec l'estimation d'erreur (4.9) pour l'algorithme du gradient à pas optimal . Par exemple, d'après ces estimations pour $K(A) = 100$, pour obtenir une erreur relative de 10^{-6} sur l'énergie, il faudrait 340 itérations du gradient à pas optimal et seulement 34 itérations du gradient conjugué ! Comme les itérations sont comparables, ces performances font de cet algoritme le favori de tous les gens qui font des calculs de grande taille. De nombreuses extensions ont été proposées : BiCGSTAB, GMRES, etc, pour des problèmes non symétriques, à coefficients complexes, etc..

4.6 Calcul du pas pour les méthodes de descente

Référence 1 : Bonnans, Joseph-Frédéric, Jean Charles Gilbert, Claude Lemaréchal, and Claudia A. Sagastizábal. Numerical optimization : theoretical and practical aspects. Springer Science & Business Media, 2006.
Référence 2 : Gander, W., Gander, M.J. and Kwok, F., 2014. Scientific computing-An introduction using Maple and MATLAB (Vol. 11). Springer Science & Business, 2014. Les codes utilisés sont tirés de la référence 2. Les codes peuvent être trouvés là `https://www.unige.ch/~gander/book.php`

Voir aussi `https://optimization.mccormick.northwestern.edu/index.php/Line_search_methods`

Le choix de la direction de descente et du pas sont très importants. Dans l'exemple suivant, tiré de la référence 2, on minimise la fonction x^2 en partant de $x_0 = 2$. A gauche on fixe la direction $d^k = (-1)^{k+1}$ et $\rho_k = 2 + \frac{3}{2^k}$, à droite $d^k = 1$ et $\rho_k = \frac{1}{2^k}$. Les itérées sont tracées Figure 4.6, sur la fonction $x \to x^2$.

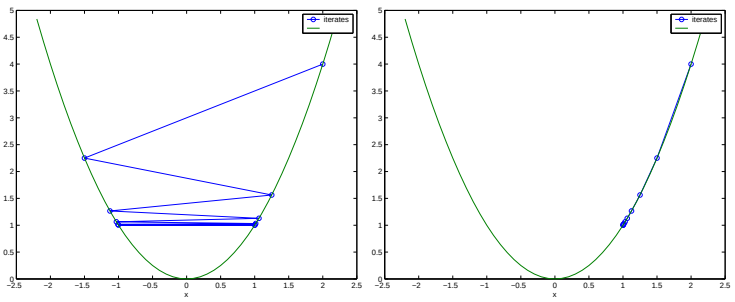
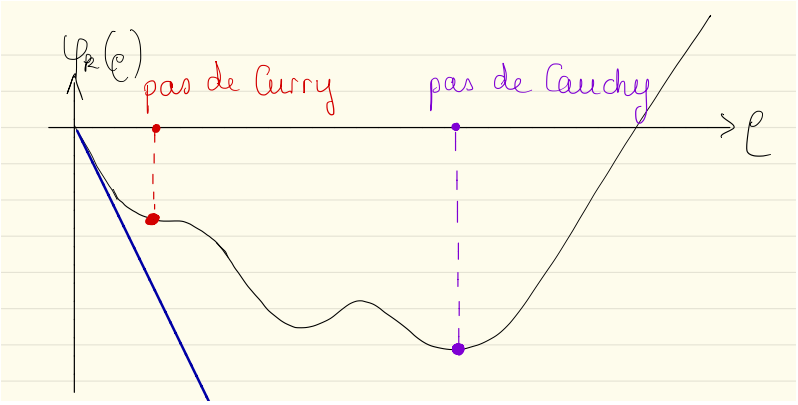


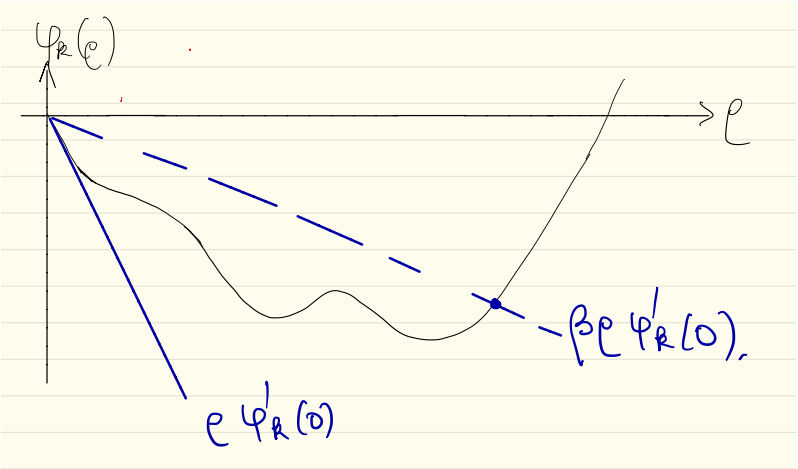
FIGURE 12.12.
On the left a case where the step length in the line search is too long, and on the right one where the step length is too short. In both cases, the line search methods do not converge to a minimum.

Supposons à l'étape k un point x^k et une direction de descente d^k déterminés (relaxation, gradient, gradient conjugué, Newton, cf paragraphe suivant). Rappelons que si $\varphi_k(\rho) = J(x^k - \rho d^k)$, une direction de descente dans le cas différentiable satisfait forcément $\varphi_k'(0) < 0$. Dans le cas quadratique on peut déterminer facilement le pas optimal par la formule (4.4) en écrivant que le minimum de J dans la direction d^k est déterminé de manière unique par $\varphi_k'(\rho_k) := J'(x^k - \rho_k d^k).d^k = 0$. On parle alors de *pas de Cauchy*. Sinon, il peut y avoir plusieurs valeurs de ρ_k où la dérivée de φ_k s'annule. On définit le pas de Curry comme le premier $\rho > 0$, cf Figure 4.6.



Mais il n'est pas facile de le calculer numériquement. On va choisir le pas de façon à avoir une fraction du modèle linéaire. Soit $\beta \in]0,1[$, on dira que le pas ρ est admissible si

$$\varphi_k(\rho) \leq \varphi_k(0) + \beta\rho\varphi_k'(0) \tag{4.18}$$



La technique de rebroussement (ou backtracking) d'Armijo consiste à se donner un ρ_{init}, regarder s'il convient (*i.e.* si (4.18) est vérifié) et sinon à le multiplier par τ donné à l'avance ($\tau = \frac{1}{2}$ par exemple) jusqu'à ce que (4.18) soit vérifié.

Données (x, d) ;
Initialisation $\rho = \rho_{init}$;
while $J(x - \rho d > J(x) - \rho\beta J'(x) \cdot d$;
  $\rho = \tau\rho$;
end

Algorithme 1 : Armijo, recherche de pas linéaire

Théorème 4.8 *Soit J une fonction continue différentiable, avec J' localement lipqchitzienne pour tout x de constante de Lipschitz $L(y)$:*

$$\|J'(x+y) - J'(x)\| \leq L(x)\|y\|$$

Si $\beta \in]0,1[$, si d^k est une direction de descente en x^k, alors la condition d'Armijo

$$\varphi_k(\rho) \ \leq \ \varphi_k(0) + \beta\rho\varphi_k'(0)$$

est satisfaite pour tout $\rho \in [0, \rho_{max}]$ où

$$\rho_{max} = \frac{2(\beta-1)J'(x^k)\cdot d^k}{L(x^k)\|d^k\|^2}$$

On en déduit

Corollaire 4.3 *Sous les hypothèses du précédent théorème, la recherche du pas d'Armijo avec rebroussement se termine avec*

$$\rho_k \ \geqslant \ \min(\rho_{init}, \frac{2\tau(\beta-1)J'(x^k)\cdot d^k}{L(x^k)\|d^k\|^2})$$

En général on ne sait pas calculer la constante de Lipschitz, et l'algorithme calcule le pas optimal sans avoir cette information.

Initialisation $x = x_0,\ k = 0$;
while not converged ;
  trouver une direction de descente d;
  $\rho = \rho_{init}$;
  while $J(x - \rho d > J(x) - \rho\beta J'(x)\cdot d$;
    $\rho = \tau\rho$;
  end;
  $x = x - \rho d$;
  $k = k + 1$;
end

Algorithme 2 : Algorithme de descente générique basé sur Armijo

Théorème 4.9 *Soit J une fonction continue différentiable, avec J' localement lipqchitzienne pour tout x de constante de Lipschitz $L(y)$:*

$$\|J'(x+y) - J'(x)\| \leq L(x)\|y\|$$

Alors l'algorithme générique de descente 2 produit l'un des résultats suivants :

1. *Il existe un $k \geqslant 0$ tel que $J'(x^k) = 0$.*
2. $\lim_{k\to+\infty} J(x^k) = -\infty$.
3. $\lim_{k\to+\infty} \min(|J'(x^k)\cdot d^k|, \frac{|J'(x^k)\cdot d^k|}{\|d^k\|^2}) = 0$.

Malheureusement ce résultat ne prouve pas que l'algorithme converge vers un point critique, puisque il se peut que le gradient soit de plus en plus orthogonal à la direction de descente. Il faut alors préciser la direction de descente pour avoir un meilleur résultat.

4.6.1 Méthode du gradient

Ici $d^k = J'(x^k)$, et on a

Corollaire 4.4 *Soit J une fonction continue différentiable, avec J' localement lipschitzienne pour tout x de constante de Lipschitz $L(y)$:*

$$\|J'(x+y) - J'(x)\| \leq L(x)\|y\|$$

Alors l'algorithme générique de descente 2 avec direction de descente égale au gradient produit l'un des résultats suivants :

1. *Il existe un $k \geqslant 0$ tel que $J'(x^k) = 0$.*
2. $\lim_{k\to+\infty} J(x^k) = -\infty$.
3. $\lim_{k\to+\infty} J'(x^k) = 0$.

```
function [x,xk]=SteepestDescent(f,fp,x0,tol,maxiter,tau,be,alinit)
% STEEPESTDESCENT steepest descent minimum search with Armijo line search
%   [x,xk]=SteepestDescent(f,fp,x0,tol,maxiter,tau,be,alinit) finds an
%   approximate minimum of the function f with gradient fp, starting
%   at the initial guess x0. The remaining parameters are optional and
%   default values are used if they are omited. xk contains all the
%   iterates of the method.

if nargin<8, alinit=1; end;
if nargin<7, be=0.1; end;
if nargin<6, tau=0.5; end;
if nargin<5, maxiter=50; end;
if nargin<4, tol=1e-6; end;

x=x0;
xk=x0;
p=-feval(fp,x);
k=0;
while norm(p)>tol & k<maxiter
  al=alinit
  while feval(f,x+al*p)>feval(f,x)-al*be*p'*p
    al=tau*al;
  end;
  x=x+al*p;
  p=-feval(fp,x);
  k=k+1;
  xk(:,k+1)=x;
end;
```

Application à un problème quadratique

```
f=inline('x(1)^2+4*x(2)^2+x(1)*x(2)','x');
fp=inline('[2*x(1)+x(2);8*x(2)+x(1)]','x');
F=inline('x.^2+4*y.^2+x.*y','x','y');   % for plotting purposes
[X,Y]=meshgrid(-1.5:0.1:1,-0.5:0.1:1.5);
contour(X,Y,F(X,Y),50);
hold on
```

```
% steepest descent Armijo
 [x,xk]=SteepestDescent(f,fp,[-1.2;1])
for i=1:size(xk,2)
  plot(xk(1,1:i),xk(2,1:i),'-om','Linewidth',2);
  %pause
end;
% steepest decent optimal quadratic case
A=[2 1;1 8];
x0=[-1.2;1];
x=x0;
xkopt=x0;
d=A*x;
k=0;
tol=1e-6;
maxiter=12;
while  norm(d)>tol & k<maxiter
  ar=A*d;
  rho=(d'*d)/(d'*ar);
  x=x-rho*d;
  k=k+1;
  xkopt(:,k+1)=x;
  d= A*x;
end;
for i=1:size(xkopt,2)
  plot(xkopt(1,1:i),xkopt(2,1:i),'-*b','Linewidth',2);
  pause
end;
hold off
```

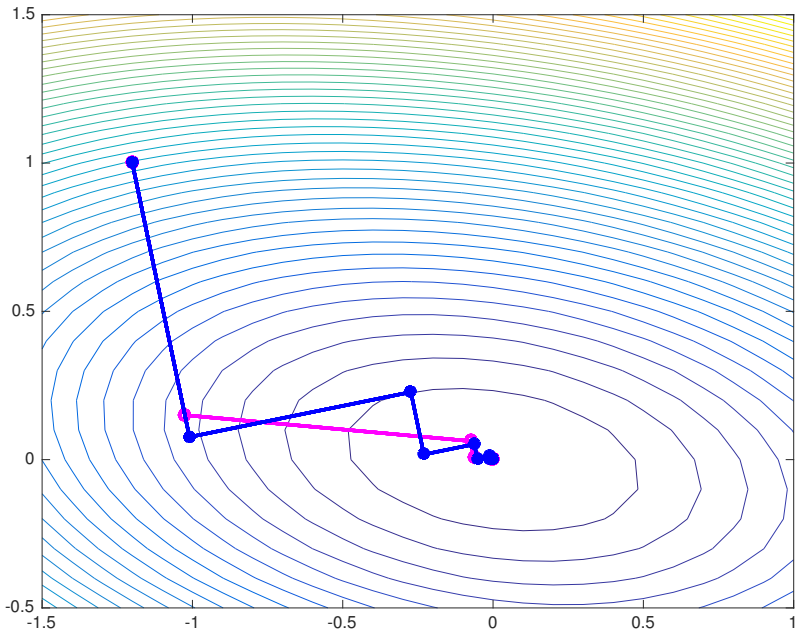


FIGURE 4.1 – un exemple quadratique, Armijo (magenta), gradient à pas optimal (bleu)

Un cas non quadratique

```
f=@(x) 10*(x(2)-x(1)^2)^2+(x(1)-1)^2;
fp=@(x) [-40*(x(2)-x(1)^2)*x(1)+2*(x(1)-1); 20*(x(2)-x(1)^2)];
[x,xk]=SteepestDescent(f,fp,[-1.2;1])
F=@(x,y) 10*(y-x.^2).^2+(x-1).^2;
Fp1=@(x,y) -40*(y-x.^2).*x+2*(x-1);
Fp2=@(x,y) 20*(y-x.^2);
[X,Y]=meshgrid(-1.5:0.1:1,-0.5:0.1:1.5);
contour(X,Y,F(X,Y),50)
hold on
quiver(X,Y,Fp1(X,Y),Fp2(X,Y),5)
plot(xk(1,:),xk(2,:),'?-o?')
hold off
```

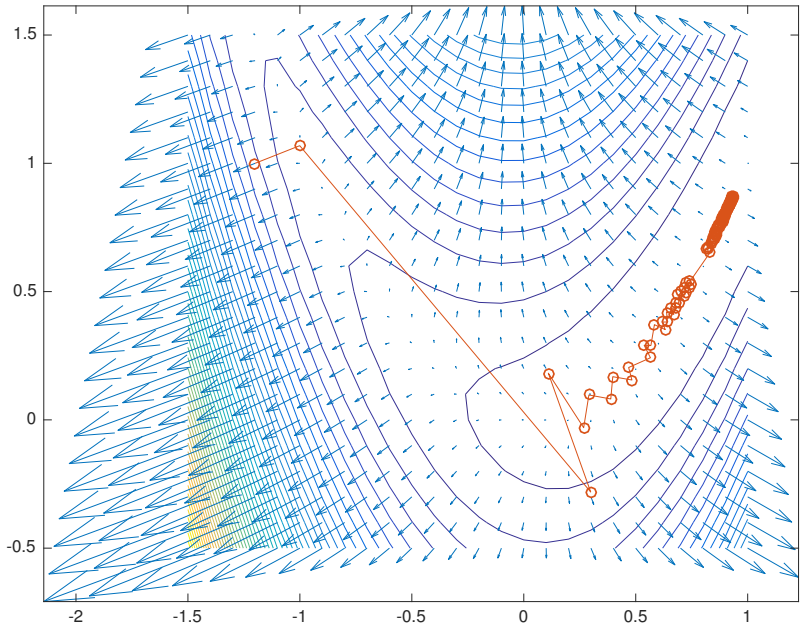


FIGURE 4.2 – Armijo-gradient, un exemple non quadratique

Un autre exemple

$$r(x, y) = (x - 1)^2 + p(x^2 - y^2)^2, p = 10.$$

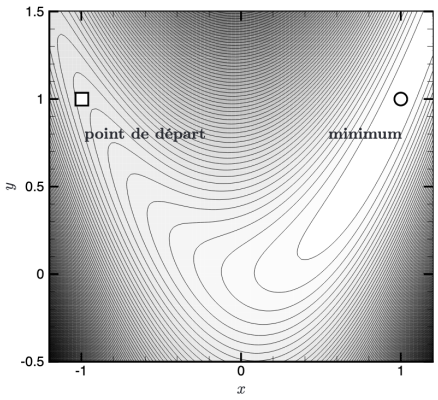


FIGURE C.1 – *Isovaleurs de la fonction de Rosenbrock banana. Le point de départ des algorithmes déterministes est* $(x\,;y) = (-1\,;1)$ *et le minimum est localisé au point* $(x\,;y) = (1\,;1)$.

FIGURE 4.3 – Fonction de Rosenbrook

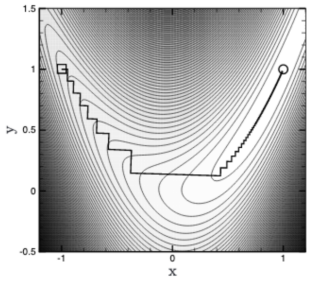


FIGURE C.2 – *Algorithme du gradient à pas optimal appliqué à la fonction Rosenbrock.*

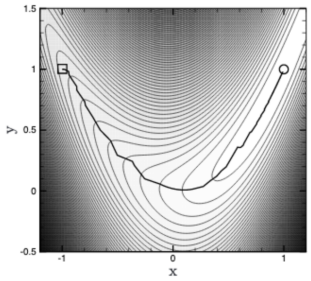


FIGURE C.3 – *Algorithme du gradient à pas d'Armijo appliqué à la fonction Rosenbrock.*

FIGURE 4.4 – Algorithmes pour la fonction de Rosenbrook

4.7 Méthodes de Newton et quasi-Newton

Soit J une fonctionnelle de V dans $\mathbb{R}$ où V est un Hilbert. On cherche les points critiques de J c'est-à-dire les zéros de $G = J'$, $V \to V$. Pour cela on définit une méthode itérative en faisant un développement de Taylor-Young en x^k :

$$G(x^{k+1}) = G(x^k + (x^{k+1} - x^k) = G(x^k) + G'(x^k).(x^{k+1} - x^k) + \mathcal{O}((x^{k+1} - x^k)^2)$$

SI on veut que $G(x^{k+1})$ approche 0, on doit résoudre à chaque étape

$$G'(x^k).(x^{k+1} - x^k) := J''(x^k) \cdot (x^{k+1} - x^k) = -J'(x^k). \tag{4.19}$$

On retrouve la méthode de Newton de la dimension 1. Avantage : on converge plus vite. Inconvénient : on n'est pas sûr de converger, et il faut calculer le Hessien $H(x^k) = J''(x^k)$ et résoudre un système linéaire à chaque étape.

Théorème 4.10 (Méthode de Newton, convergence quadratique) *On suppose J deux fois différentiable, on pose $G = J'$, et on suppose de plus qu'il existe 3 constantes $r, M > 0$, et $0 < \beta < 1$ tels que*

1. $\sup_{x \in B(x^0,r)} \|G'(x)^{-1}\| \leq M$,
2. $\sup_{(x,y) \in B^2 \in B(x^0,r)} \|G'(x) - G'(y)\| \leq \frac{\beta}{M}$,
3. $\|G(x^0)\| \leq \frac{r}{M}(1-\beta)$,

Alors (4.19) *définit à partir de x^0 une suite contenue dans $B(x^0, r)$ qui converge vers un zéro a de G qui est le seul zéro de G dans $B(x^0, r)$. De plus si G est deux fois différentiable la convergence est quadratique :*

$$\|x^{k+1} - a\| \leq \frac{1}{2} M (\sup_{x \in B(x^0,r)} \|G''(x)\|) \|x^k - a\|^2.$$

Comment appliquer la méthode de Newton pour l'optimisation ? On constate d'abord que si B_k est une matrice symétrique définie positive, la direction

$$d^k = B_k^{-1} J'(x^k)$$

est une direction de descente ($D^k \cdot J'(x^k) > 0$). De plus cette direction de descente mène directement à un point stationnaire de la fonction quadratique

$$g(x^k + d) := J(x^k) - (J'(x^k), d) + \frac{1}{2}(B_k d, d)$$

qui peut être vue comme une approximation quadratique de J dans un voisinage de x^k. De plus si $B_k = H(x^k) := J''(x^k)$, c'est la méthode de Newton.

Théorème 4.11 *Avec les mêmes hypothèses sur la fonction J que dans les théorèmes précédents. L'algorithme générique de descente 2 avec direction de descente $d^k = B_k^{-1} J'(x^k)$ où les B^k sont symétriques définies positives avec $\lambda_{max}(B_k) \leq \overline{\lambda_{max}} < +\infty$ et $\lambda_{min}(B_k) \geqslant \overline{\lambda_{min}} > -\infty$ produit l'un des résultats suivants :*

1. *Il existe un $k \geqslant 0$ tel que $J'(x^k) = 0$.*
2. $\lim_{k \to +\infty} J(x^k) = -\infty$.
3. $\lim_{k \to +\infty} J'(x^k) = 0$.

```
function [x,xk]=Newton(f,fp,fpp,x0,tol,maxiter,tau,be,alinit)
% NEWTON Minimization with Newton descent and Armijo line search
%   [x,xk]=Newton(f,fp,fpp,x0,tol,maxiter,tau,be,alinit) finds an
%   approximate minimum of the function f with gradient fp and Hessian
%   fpp, starting at the initial guess x0. The remaining parameters are
%   optional and default values are used if they are omited. xk
%   contains all the iterates of the method.

if nargin<9, alinit=1; end;
if nargin<8, be=0.1; end;
if nargin<7, tau=0.5; end;
if nargin<6, maxiter=100; end;
if nargin<5, tol=1e-6; end;

x=x0;
```

```
xk=x0;
p=-feval(fpp,x)\feval(fp,x);
k=0;
while norm(feval(fp,x))>tol & k<maxiter
  al=alinit;
    while feval(f,x+al*p)>feval(f,x)-al*be*p'*p
     al=tau*al;
   end;
  x=x+al*p;
  p=-feval(fpp,x)\feval(fp,x);
  k=k+1;
  xk(:,k+1)=x;
end;
```

Théorème 4.12 *Avec l'hypothèse supplémentaire que J est deux fois différentiable et que son Hessien est globalement lipschitz de constante de Lipschitz L. On considère la séquence d'itérées générées par l'algorithme générique de descente 2 avec direction de descente $d^k = B_k^{-1} J'(x^k)$ où les B^k sont soit $H(x^k)$ si elle est définie positive, soit comme dans le théorème précédent. Soit $\rho_{init} = 1$ et $\beta \in]0, \frac{1}{2}[$. Si la suite x^k a un point d'accumulation x^*, alors*

1. *$\rho_k = 1$ pour k grand,*
2. *toute la suite x^k converge vers x^*,*
3. *la convergence est quadratique, $\|x^{k+1} - x\| \leq C\|x^k - x\|^2$.*

Le même exemple non quadratique que précédemment montre que Newton converge et bien mieux que le gradient, même si l'on part loin de la solution.

```
f=inline('10*(x(2)-x(1)^2)^2+(x(1)-1)^2','x');
fp=inline('[-40*(x(2)-x(1)^2)*x(1)+2*(x(1)-1); 20*(x(2)-x(1)^2)]','x');
H=inline('[-40*(x(2)-x(1)^2)+80*x(1)^2+2 -40*x(1); -40*x(1) 20]','x');
x0=[-1.2;1];
[x,xkn]=Newton(f,fp,H,x0);
```

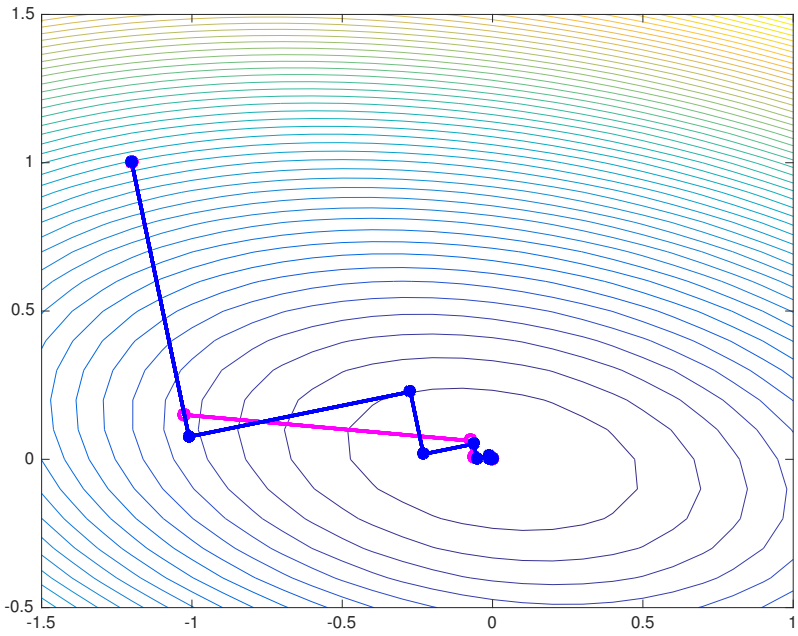


FIGURE 4.5 – Le même exemple que précédemment, Armijo (magenta), Newton (bleu)