

Mathématiques Numériques 1A

MN2

Chapitre 2 - Optimisation non-linéaire

J.-F. Scheid

Télécom Nancy

2025-2026

- 1 Introduction
- 2 Rappels et compléments d'optimisation
 - Existence d'optimum
 - Conditions d'optimalité
 - Ensembles de niveaux et gradients
- 3 Optimisation sans contrainte
 - Méthodes de descente
- 4 Optimisation avec contraintes
 - Contraintes égalités et multiplicateurs de Lagrange
 - Contraintes inégalités et conditions de KKT

I. Introduction

Dans ce chapitre, on s'intéresse au problème suivant : *trouver le minimum d'une fonction de plusieurs variables avec ou sans contraintes sur les variables.*

Pour une fonction $f : V \subset \mathbb{R}^n \rightarrow \mathbb{R}$ où V est un sous-ensemble de $\mathbb{R}^n$, ce problème s'écrit¹ :

$$\boxed{\inf_{x \in V} f(x)} \quad (1)$$

- Si $V = \mathbb{R}^n$ i.e. l'espace tout entier, il n'y a pas de sur les variables : c'est le **cas sans contrainte**.
- Si $V \subsetneq \mathbb{R}^n$ i.e. les variables sont contenues dans un sous-ensemble strict de $\mathbb{R}^n$: c'est le **cas avec contraintes**.

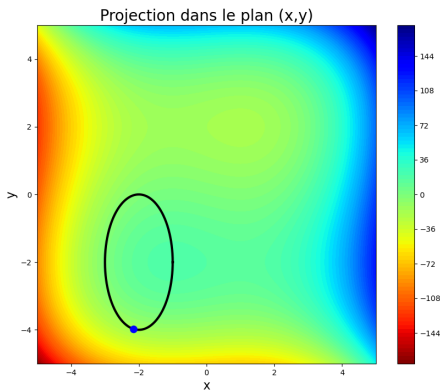
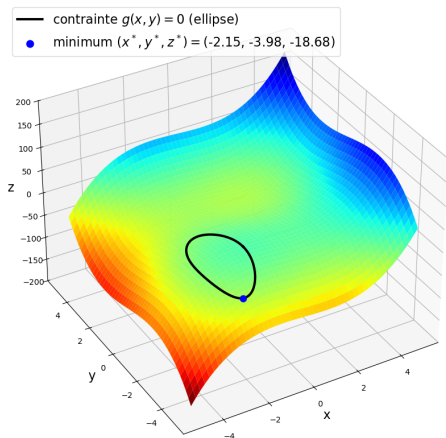
1. "inf" désigne la borne inférieure ou infimum de f c'est à dire le plus grand minorant de f . On ne sait pas a priori si la borne inférieure est atteinte ou pas. Pour indiquer que la valeur de l'infimum est atteint, on écrira $\min_{x \in V} f(x)$.

Exemples.

- Trouver le minimum de $f(x, y) = x^3 - 3x + y^3 - 12y$ avec la contrainte que x et y doivent vérifier

$$g(x, y) := (x + 2)^2 + ((y + 2)/2)^2 - 1 = 0$$

(une ellipse).



- Déterminer le parallélépipède de surface fixée S_0 qui a le volume maximal.

Si a, b, c sont les trois dimensions du parallélépipède, il s'agit de maximiser $V(a, b, c) = abc$ sous les contraintes $a, b, c \geq 0$ et $S(a, b, c) = 2ab + 2bc + 2ca = S_0$.

Il s'agit donc d'un problème d'optimisation dans $\mathbb{R}^3$ sous contrainte (réponse : c'est le cube).

- *Programmation linéaire.* Il s'agit de minimiser une fonction linéaire avec des contraintes linéaires sur les variables :

$$\begin{aligned} \min_{x \in \mathbb{R}^n} c^T x &= c_1 x_1 + \cdots + c_n x_n \\ \begin{cases} Ax \leq b \\ x \geq 0 \end{cases} \end{aligned}$$

où A est une matrice de taille (m, n) (avec $m \leq n$), $b \in \mathbb{R}^m$.

☛ L'étude de ces problèmes est un des objectifs du cours de GRO 2A.

II. Rappels et compléments d'optimisation

On renvoie aussi au Chap. 1 où certains rappels ont déjà été donnés.

1. Existence d'optimum

Théorème 1.

Soit f une fonction continue sur un sous-ensemble V non-vide, fermé^a de $\mathbb{R}^n$. On suppose que

- ❶ ou bien V est borné,
- ❷ ou bien V est non-borné (en particulier si $V = \mathbb{R}^n$) et $\lim_{\|x\| \rightarrow +\infty} f(x) = +\infty$ (on dit alors que f est *coercive*)

alors f possède au moins un minimum sur V .

a. V est fermé si et seulement si pour toute suite convergente de V , sa limite est dans V .

Exemple. La fonction $f(x, y) = x^2 + y^2 - ax - by - c$ avec $a, b, c \in \mathbb{R}$ est coercive sur $\mathbb{R}^2$.

2. Conditions d'optimalité

Soit $f : V \subset \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction dérivable. On note

$$\nabla f(x) \in \mathbb{R}^n$$

le gradient de f en x (vecteur ligne de $\mathbb{R}^n$).

(a) Matrice hessienne.

On dit que f est de classe C^2 sur $\mathbb{R}^n$ si f admet des dérivées partielles jusqu'à l'ordre 2 et si ces dérivées partielles sont continues sur $\mathbb{R}^n$.²

On note alors $D^2f(x)$ la matrice *hessienne* de f en x définie par

$$D^2f(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix} \in \mathbb{R}^{n \times n} \quad (2)$$

2. On peut remplacer l'espace $\mathbb{R}^n$ tout entier par un ouvert $\mathcal{U}$ de $\mathbb{R}^n$.

Autrement dit

$$\boxed{D^2 f(x) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(x) \right)_{1 \leq i, j \leq n} \in \mathbb{R}^{n \times n}} \quad (3)$$

On notera aussi parfois la matrice hessienne

$$f'' := D^2 f$$

Quelques propriétés de la hessienne.

- *Théorème de Schwarz.* Pour $f \in C^2(\mathbb{R}^n)$, on a

$$\boxed{\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}} \quad (4)$$

☞ Ainsi ($f \in C^2$) la matrice hessienne $D^2 f$ est *symétrique* et ses valeurs propres sont donc réelles.

- *Formule de Taylor.* Pour $f \in C^2(\mathbb{R}^n)$ et tout $x \in \mathbb{R}^n$, on a

$$\boxed{f(x + h) = f(x) + \nabla f(x)h + \frac{1}{2}h^\top D^2 f(x)h + o(\|h\|^2)} \quad (5)$$

(b) Conditions nécessaires d'optimalité.

On a des conditions nécessaires d'optimalité portant sur la matrice hessienne.

Théorème 3. Conditions nécessaires d'optimalité

① Condition d'optimalité d'ordre 1.

On suppose que f est dérivable sur $\mathbb{R}^n$. Si $\bar{x}$ est un minimum local^a de f alors

$$\nabla f(\bar{x}) = 0. \quad (6)$$

On dit alors que $\bar{x}$ est un *point critique* de f .

② Condition d'optimalité d'ordre 2.

On suppose que f est de classe C^2 sur $\mathbb{R}^n$. Si $\bar{x}$ est un minimum local de f alors

$$\begin{cases} \nabla f(\bar{x}) = 0 \\ D^2 f(\bar{x}) \text{ est symétrique semi-définie positive}^b. \end{cases} \quad (7)$$

a. cf. Définition 4, Chap. 1

b. cf. Définition 6, Chap. 1

Remarques.

- La matrice $D^2f(x)$ est semi-définie positive (resp. définie positive) si et seulement si toutes ses valeurs propres sont ≥ 0 (resp. > 0).
- Si $\bar{x}$ est un *maximum* local de f alors c'est aussi un point critique i.e. $\nabla f(\bar{x}) = 0$ et si $f \in C^2(\mathbb{R}^n)$, la matrice hessienne est semi définie *négative* cette fois i.e. $(D^2f(\bar{x})v|v) \leq 0, \forall v \in \mathbb{R}^n$.
- On rappelle que les réciproques ne sont pas vraies en général : un point critique n'est pas forcément un minimum ou un maximum local. Il peut s'agir d'un **point selle** i.e. un point critique qui n'est justement ni un minimum local, ni un maximum local. Voir MN2, Chap. 1, §5, p.34.

(c) Conditions suffisantes d'optimalité.

Au Chapitre 1, on a donné une condition *suffisante* de minimum sur $\mathbb{R}^n$. Si la fonction f définie sur $\mathbb{R}^n$ est *convexe* et $\nabla f(\bar{x}) = 0$ alors $\bar{x}$ est un minimum *global* de f (cf. Th. 4 du Chap. 1.).

Théorème 4. Condition suffisante d'optimalité

Soit $f \in C^2(\mathbb{R}^n)$ et $\bar{x} \in \mathbb{R}^n$ tel que

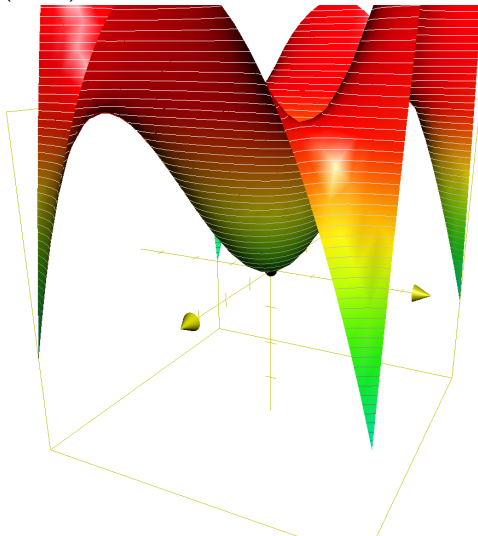
$$\begin{cases} \nabla f(\bar{x}) = 0 \\ D^2 f(\bar{x}) \text{ est symétrique définie positive.} \end{cases} \quad (8)$$

Alors $\bar{x}$ est un minimum local de f .

Exemples.

- Soit $\mathbb{R} \ni x \xrightarrow{f} x^2$ en $\bar{x} = 0$, on $f'(0) = 0$ et $f''(0) = 2 > 0$ donc $\bar{x} = 0$ est un minimum (local) de f .

- La fonction $f(x, y) = x^2(1 - y^2) + y^2$ admet $(\bar{x}, \bar{y}) = (0, 0)$ comme minimum local. En effet, $\nabla f(0, 0) = 0$ et $D^2f(0, 0) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$ qui est définie positive (les v.p sont >0).



(d) Convexité et hessienne.

Un ensemble $E \subseteq \mathbb{R}^n$ est *convexe* si $\forall x, y \in E$, on a

$$z = \alpha x + (1 - \alpha)y \in E$$

pour tout $\alpha \in [0, 1]$ i.e. tout le segment joignant x à y est contenu dans E .

Proposition 2.

Soit $f \in C^2(\mathbb{R}^n)$ et $E \subseteq \mathbb{R}^n$ un ensemble convexe. Alors f est convexe sur E si et seulement si

$$\forall x \in E, D^2f(x) \text{ est semi-définie positive.}$$

Remarque. Si $D^2f(x)$ est *définie positive* pour tout x alors f est *strictement* convexe sur E mais la réciproque est fautive (prendre $f(x) = x^4$ qui est strictement convexe mais $f''(0) = 0$).

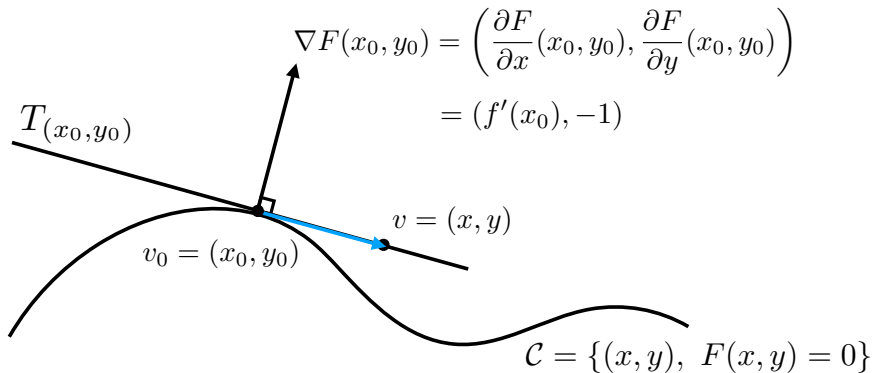
3. Ensembles de niveaux et gradients

(a) *Cas des courbes ($n = 1$)*. Soit $f : \mathbb{R} \mapsto \mathbb{R}$ dérivable en x_0 . On introduit la fonction $F : \mathbb{R}^2 \mapsto \mathbb{R}$ définie par $F(x, y) = f(x) - y$.

☛ La courbe d'équation $y = f(x)$ correspond donc à la courbe de niveau de F par $F(x, y) = 0$.

On note $\mathcal{C}$ la courbe de niveau de F passant par (x_0, y_0) avec $y_0 = f(x_0)$, définie par

$$\mathcal{C} = \{(x, y) \in \mathbb{R}^2 \text{ tel que } F(x, y) = 0\}.$$



La tangente $T_{(x_0, y_0)}$ à la courbe en (x_0, y_0) a pour équation

$$y = y_0 + f'(x_0)(x - x_0) \quad \text{avec } y_0 = f(x_0) \quad (9)$$

Proposition 1.

Soit $v_0 = (x_0, y_0)$. Le gradient $\nabla F(v_0)$ est orthogonal à la courbe de niveau $\mathcal{C}$:

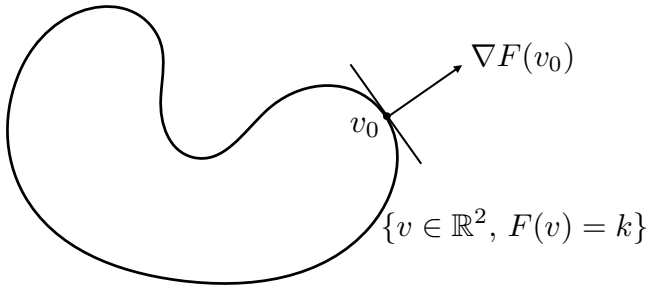
$$(\nabla F(v_0) | v_0 - v) = 0, \quad \forall v \in T_{(x_0, y_0)} \quad (10)$$

Preuve. Pour tout $v = (x, y) \in T_{(x_0, y_0)}$, on a

$$\begin{aligned}(\nabla F(v_0) \mid v_0 - v) &= (f'(x_0), -1) \cdot \begin{pmatrix} x_0 - x \\ y_0 - y \end{pmatrix} \\&= f'(x_0)(x_0 - x) - (y_0 - y) \\&= 0 \quad \text{car } v = (x, y) \in T_{(x_0, y_0)}\end{aligned}$$

□

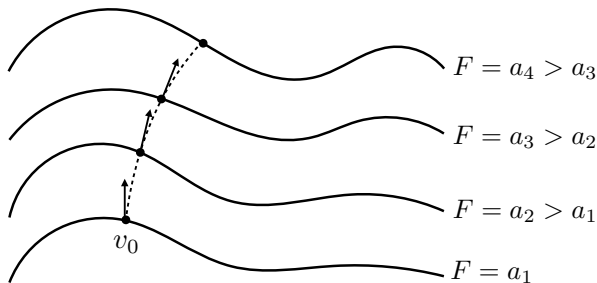
De façon générale, pour une courbe dans $\mathbb{R}^2$ définie par un ensemble de niveau¹ $F(x, y) = k$ d'une fonction $F : \mathbb{R}^2 \mapsto \mathbb{R}$ dérivable en $v_0 = (x_0, y_0)$, le gradient $\nabla F(v_0)$ est orthogonal à la courbe en v_0 .



1. une courbe qui n'est pas donc nécessairement le *graphe d'une fonction*

Proposition 2.

Le gradient $\nabla F(v_0)$ indique la direction de plus forte pente à partir du point v_0 .



Preuve. On a $DF(v_0, d) = \nabla F(v_0)d$ et donc

$$DF(v_0, d) = \|\nabla F(v_0)\| \|d\| \cos(\theta)$$

où θ est l'angle entre le gradient $\nabla F(v_0)$ et la direction d . Le maximum de la dérivée directionnelle $DF(v_0, d)$ dans la direction d est atteint avec $\theta = 0$ i.e. dans la direction du gradient. □

(b) *Cas général des hypersurfaces.* Soit $f : \mathbb{R}^n \mapsto \mathbb{R}$ dérivable en $x_0 \in \mathbb{R}^n$. On note³ $z_0 = f(x_0)$ et $v_0 = (x_0, z_0) \in \mathbb{R}^{n+1}$.

▪ Le plan tangent T_{v_0} de f en x_0 est défini par

$$T_{v_0} = \{v \in \mathbb{R}^{n+1} \text{ tel que } v = (x, z) \text{ avec } x \in \mathbb{R}^n, \\ z = f(x_0) + \nabla f(x_0)(x - x_0)\}. \quad (11)$$

▪ On introduit la fonction $F(x, z) = f(x) - z$ avec $x \in \mathbb{R}^n$, $z \in \mathbb{R}$ et on note $\mathcal{S}$ l'hypersurface correspondant à l'ensemble de niveau

$$F(x, z) = f(x) - z = 0$$

(défini donc aussi par $z = f(x)$).

Le gradient de F en $v_0 \in \mathbb{R}^{n+1}$ est un vecteur qui est orthogonal au plan tangent T_{v_0} i.e. :

$$(\nabla F(v_0) | v_0 - v) = 0, \quad \text{pour tout } v \in T_{v_0}. \quad (12)$$

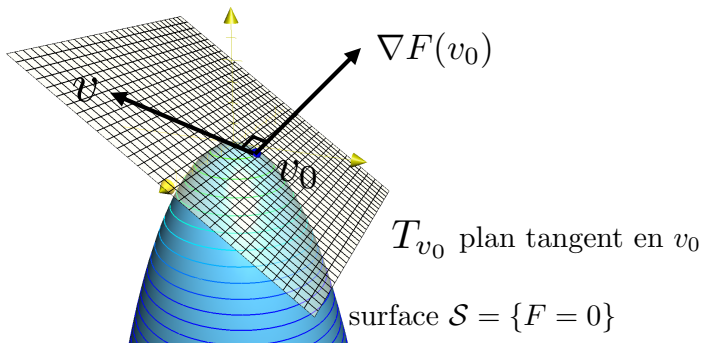
3. il faut comprendre la notation $v = (x, z) \in \mathbb{R}^{n+1}$ avec $x = (x_1, \dots, x_n) \in \mathbb{R}^n$, $z \in \mathbb{R}$, par $v = (x_1, \dots, x_n, z)$ où on a ajouté la composante z à x .

Preuve de (12). On a $\nabla F(v_0) = (\frac{\partial f}{\partial x_1}(x_0), \dots, \frac{\partial f}{\partial x_n}(x_0), -1)$ et avec $v = (x, z)$, on a

$$(\nabla F(v_0) | v_0 - v) = (\nabla f(x_0) | x_0 - x) - (f(x_0) - z) = 0$$

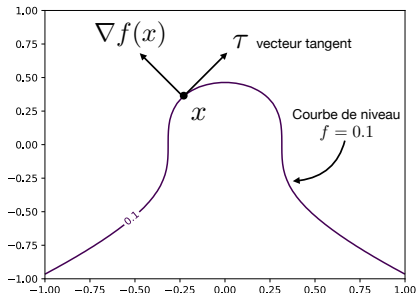
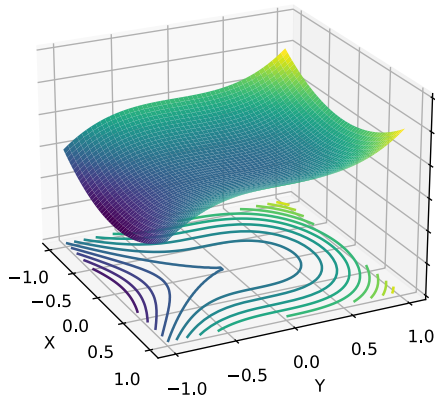
avec $z = f(x_0) + \nabla f(x_0)(x - x_0)$ pour $v \in T_{v_0}$. □

Exemple avec $n = 2$ (surface).



Exemple.

Surface $z = f(x, y) = x^2 + y^3$ et ses courbes de niveaux.



Le gradient de f en x est orthogonal à $\mathcal{C} = \{(x, y) \mid f(x, y) = 0.1\}$, il est donc dans la direction normale à $\mathcal{C}$.

III. Optimisation sans contrainte

Méthodes de descente.

Les méthodes classiques pour calculer numériquement l'extrémum $x^* \in \mathbb{R}^n$ du problème sans contrainte

$$\inf_{x \in \mathbb{R}^n} f(x)$$

consiste à construire par récurrence une suite de vecteurs $x_k \in \mathbb{R}^n$ avec :

$$\begin{cases} x_0 \in \mathbb{R}^n, \text{ donné} \\ x_{k+1} = x_k + \rho_k d_k, \quad k \geq 0 \end{cases} \quad (13)$$

où

- $d_k \in \mathbb{R}^n$ est la *direction de descente*
- $\rho_k > 0$ est le pas de la méthode (un scalaire positif)

choisis pour avoir

$$f(x_{k+1}) \leq f(x_k)$$

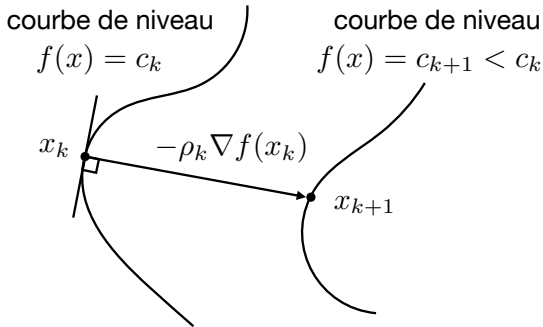
La détermination du pas ρ_k s'appelle la *recherche linéaire*.

1. Choix de la direction de descente : méthode de gradient.

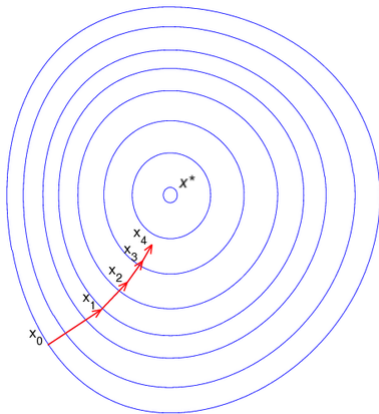
Le gradient $\nabla f(x)$ indique la direction de plus forte pente à partir du point x (cf. Proposition 2). On choisit comme direction de descente dans (13) :

$$d_k = -\nabla f(x_k) \quad (14)$$

i.e. la direction opposée au gradient (on cherche un minimum).



Le gradient en x_k est toujours orthogonal à la courbe de niveau passant par x_k .



2. Choix du pas : la recherche linéaire.

On considère la fonction d'une variable

$$q(\rho) = f(x_k + \rho d_k)$$

On a $q'(0) = \nabla f(x_k) \cdot d_k = -\|\nabla f(x_k)\|^2 < 0$. On voit donc que si on prend $\rho > 0$ "petit", la fonction q va décroître.

On peut alors choisir :

- ρ_k constant (petit) : c'est la méthode de gradient à *pas fixe*.
- ρ_k qui minimise la fonction q : c'est la méthode de gradient à *pas optimal*.

☞ La méthode du gradient à pas fixe est simple (et converge sous certaines conditions sur f et des bornes sur le pas) mais attention car il y a deux éléments contradictoires :

- si ρ est trop grand, la fonction q risque de ne pas décroître ou son comportement peut être oscillant,
- si ρ est trop petit, la convergence sera très lente.

La méthode de gradient à pas fixe est peu utilisée en pratique.

☞ D'un autre côté, la détermination du pas optimal est généralement coûteuse, c'est pourquoi on préfère généralement calculer des approximations du pas optimal. Il existe des règles classiques pour calculer une approximation du pas optimal : La règle d'Armijo, la règle de Goldstein (1967) et celle de Wolfe (1969).

Remarque. Il existe aussi d'autres méthodes de descente : gradient conjugué, Newton, quasi-Newton, gradient stochastique (cf. Chapitre suivant pour cette dernière).

IV. Optimisation avec contraintes

1. Contraintes égalités.

On considère le problème sous contraintes

$$\inf_{x \in V} f(x)$$

où l'ensemble V des contraintes (contraintes égalités) est défini par

$$V = \{x \in \mathbb{R}^n, g(x) = 0\}$$

avec $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ une fonction vectorielle, c'est-à-dire qu'il y a m contraintes

$$\boxed{g_1(x) = g_2(x) = \cdots = g_m(x) = 0.} \quad (15)$$

On suppose par la suite que les fonctions g_i sont de classe C^1 sur $\mathbb{R}^n$.

Une première remarque. Avec les contraintes (15), les extrema vérifiant ces contraintes ne sont plus nécessairement les points critiques de f i.e. les points qui annulent le gradient ∇f .

Exemple. Soit $f(x, y) = x^2 + y^2$ et on veut minimiser f sous la contrainte $g(x, y) := 1 - x - y = 0$.

On peut montrer qu'avec la contrainte $g(x, y) = 0$, la fonction f atteint son minimum⁴ en $(x^*, y^*) = (\frac{1}{2}, \frac{1}{2})$. Par ailleurs, on a $\nabla f(x, y) = (2x, 2y)$ et donc $\nabla f(x^*, y^*) = (1, 1) \neq 0$. Le minimum sous la contrainte $g(x, y) = 0$ n'est donc pas un point critique de f .

4. Substituer $y = 1 - x$ dans l'expression de f pour obtenir une fonction de la seule variable x (un polynôme de degré 2).

Le résultat suivant donne une condition nécessaire (l'équivalent du " $\nabla f(x) = 0$ " dans le cas sans contrainte) lorsque les variables sont assujetties à des contraintes égalités.

Théorème des extrema liés : multiplicateurs de Lagrange

On suppose que $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est dérivable et $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ est de classe C^1 . Si f admet un minimum local x^* sur l'ensemble $V = \{x \in \mathbb{R}^n, g(x) = 0\}$ et que les vecteurs $\nabla g_1(x^*), \dots, \nabla g_m(x^*)$ sont linéairement indépendants, alors il existe des réels $\lambda_1, \dots, \lambda_m$ tels que

$$\nabla f(x^*) = \lambda_1 \nabla g_1(x^*) + \dots + \lambda_m \nabla g_m(x^*) \quad (16)$$

Le nombre réel λ_i est appelé *multiplicateur de Lagrange* associé à la contrainte $g_i = 0$.

Remarque importantes. Dans le cas d'une seule contrainte $g(x) = 0$ (i.e. $m = 1$), la condition du théorème devient

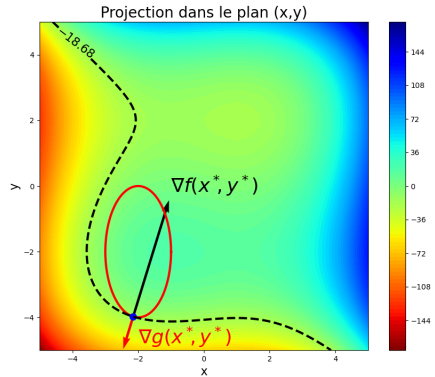
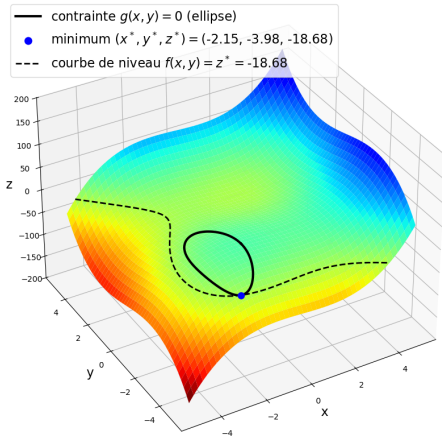
$$\nabla g(x^*) \neq 0$$

Exemple. Trouver le minimum de $f(x, y) = x^3 - 3x + y^3 - 12y$ avec la contrainte que x et y doivent vérifier
 $g(x, y) := (x + 2)^2 + ((y + 2)/2)^2 - 1 = 0$ (une ellipse)

- Le gradient de f en un point est toujours orthogonal à la courbe de niveau de f passant par ce point.
- De même, le gradient de g en un point de la courbe de niveau $g = 0$ est orthogonal à cette courbe de niveau en ce point.
- Le théorème des extrema liés dit qu'en un minimum (x^*, y^*) , le gradient de f est colinéaire au gradient de g : il existe $\lambda \in \mathbb{R}$ tel que

$$\nabla f(x^*, y^*) = \lambda \nabla g(x^*, y^*)$$

(avec $\nabla g(x^*, y^*) \neq 0$)



Remarque sur le lagrangien. On peut intégrer la contrainte $g(x) = 0$ dans une fonction objectif. On définit ainsi le *lagrangien* de f sous la contrainte $g = 0$ avec

$$\begin{aligned} L : \mathbb{R}^n \times \mathbb{R}^m &\rightarrow \mathbb{R} \\ L(x, \lambda) &= f(x) + \lambda^\top \cdot g(x) = f(x) + \sum_{j=1}^m \lambda_j g_j(x) \end{aligned} \tag{17}$$

Les conditions d'optimalité (16) et les contraintes égalité $g = 0$ correspondent aux points critiques (x^*, λ^*) de L :

$$\begin{aligned} \frac{\partial L}{\partial x_i}(x^*, \lambda^*) &= 0, \quad \forall i \in \llbracket 1, n \rrbracket \\ \frac{\partial L}{\partial \lambda_j}(x^*, \lambda^*) &= 0, \quad \forall j \in \llbracket 1, m \rrbracket \end{aligned} \tag{18}$$

Exemple.⁵ Considérons l'exemple suivant :

$$\inf_{(x,y) \in V} f(x,y) := x^4 + y^4$$

avec

$$V = \{(x,y) \in \mathbb{R}^2 \text{ tel que } x^2 + y^2 = 1\}.$$

▪ **Calcul direct.** On peut se ramener à la minimisation d'une fonction d'une variable réelle. En effet, posons $x = \cos \theta$ et $y = \sin \theta$ avec $\theta \in [0, 2\pi[$. On trouve alors que

$$f(\cos \theta, \sin \theta) = (\cos^2 \theta + \sin^2 \theta)^2 - 2 \cos^2 \theta \sin^2 \theta = 1 - \frac{1}{2} \sin^2(2\theta).$$

Ainsi, $\theta \mapsto f(\cos \theta, \sin \theta)$ est minimale lorsque $\sin^2(2\theta) = 1$, c'est-à-dire pour $\theta \in \{\frac{\pi}{4}, \frac{3\pi}{4}, \frac{5\pi}{4}, \frac{7\pi}{4}\}$ ou encore que f est minimale pour $(x,y) = \left(\pm \frac{\sqrt{2}}{2}, \pm \frac{\sqrt{2}}{2}\right)$.

5. Exemple tiré du cours d'optimisation de Y. Privat (merci à lui), Ecole de Mines de Nancy, 2026.

▪ Extrema liés.

On retrouve ce résultat en appliquant le théorème des extrema liés. En effet, il existe $\lambda \in \mathbb{R}$ tel qu'au point (x, y) minimisant f sur V (le cercle unité), on a

$$\nabla f(x, y) = \lambda \nabla h(x, y)$$

avec $h(x, y) = x^2 + y^2 - 1$, autrement dit

$$\begin{cases} x(2x^2 - \lambda) = 0 \\ y(2y^2 - \lambda) = 0 \end{cases}$$

Dans l'ensemble des contraintes V les solutions sont ⁶

$$(i) \quad (x, y) = (0, \pm\sqrt{\lambda/2})$$

$$(ii) \quad (x, y) = (\pm\sqrt{\lambda/2}, 0)$$

$$(iii) \quad (x, y) = (\pm\sqrt{\lambda/2}, \pm\sqrt{\lambda/2})$$

En tenant compte de la contrainte égalité dans V , on obtient

$\lambda = 2$ pour les deux premiers cas (i) et (ii)

$\lambda = 1$ pour (iii)

6. $(x, y) = (0, 0)$ n'est pas solution sur V

A ce stade, il est nécessaire de faire le tri parmi les points critiques, entre maxima locaux, minima locaux et points selle.

La meilleure solution est d'évaluer f aux points obtenus.

☞ On montre alors aisément que f est maximale dans les deux premiers cas ($f = 1$) et minimale dans le dernier ($f = 1/2$). On retrouve ainsi le résultat obtenu par le calcul direct précédent.

2. Contraintes inégalités. On considère le problème sous contraintes

$$\inf_{x \in V} f(x)$$

où cette fois l'ensemble V des contraintes comporte des inégalités :

$$V = \{x \in \mathbb{R}^n, g(x) \leq 0\}$$

avec $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ une fonction vectorielle, c'est-à-dire avec m contraintes d'inégalités

$$\boxed{\begin{array}{l} g_1(x) \leq 0 \\ g_2(x) \leq 0 \\ \vdots \\ g_m(x) \leq 0 \end{array}} \quad (19)$$

☞ Il existe des conditions dues à Karush-Kuhn-Tucker (KKT, 1951) qui généralisent les conditions nécessaires des multiplicateurs de Lagrange dans le cas de contraintes égalités.

Conditions de complémentarité de KKT.

On transforme les contraintes inégalités $g_i(x) \leq 0$ en contraintes égalités en introduisant autant de variables supplémentaires, appelées *variables d'écarts* qu'il y a de contraintes inégalités :

$$g_i(x) \leq 0 \Leftrightarrow g_i(x) + e_i^2 = 0$$

On obtient ainsi les variables d'écarts $e_1, \dots, e_m$ et on peut alors appliquer le théorème des extrema liés pour des contraintes égalités avec

$$h_i(x, e) := g_i(x) + e_i^2 = 0, \quad \forall i \in \llbracket 1, m \rrbracket \quad (20)$$

Ainsi, si x^* est un minimum de f avec $g_i(x^*) \leq 0$ et $e_i^* = \sqrt{-g_i(x^*)}$ pour tout $i \in \llbracket 1, n \rrbracket$ et sous les conditions du théorème des extrema liés⁷, il existe des nombres $\lambda_1, \dots, \lambda_m$ tels que

$$\nabla_{(x,e)} f(x^*) = \sum_{i=1}^m \lambda_i \nabla_{(x,e)} h_i(x^*, e^*)$$

7. cela donne des conditions appelées *conditions de contraintes qualifiées* (voir Annexe A).

Les fonctions f et h_i sont vues désormais comme des fonctions des variables x et e et les gradients sont calculées par rapports à toutes ces variables. On a

$$\nabla_{(x,e)} h_i(x, e) = (\nabla_x g_i(x^*), \underbrace{0, \dots, 0}_{i-1}, 2e_i, 0, \dots, 0) \quad (21)$$

On obtient ainsi

$$\begin{aligned} (\nabla_x f(x^*), \underbrace{0, \dots, 0}_m) &= \sum_{i=1}^m \lambda_i (\nabla_x g_i(x^*), \underbrace{0, \dots, 0}_{i-1}, 2e_i, 0, \dots, 0), \\ \Leftrightarrow \begin{cases} \nabla f(x^*) = \sum_{i=1}^m \lambda_i \nabla g_i(x^*) \\ 0 = 2\lambda_i e_i^*, \forall i \in \llbracket 1, m \rrbracket \end{cases} &\Leftrightarrow \begin{cases} \nabla f(x^*) = \sum_{i=1}^m \lambda_i \nabla g_i(x^*) \\ 0 = \lambda_i (e_i^*)^2, \forall i \in \llbracket 1, m \rrbracket \end{cases} \end{aligned}$$

En utilisant la relation (20) sur les variables d'écart, on obtient les conditions

$$\boxed{\begin{cases} \nabla f(x^*) = \sum_{i=1}^m \lambda_i \nabla g_i(x^*) \\ 0 = \lambda_i g_i(x^*), \forall i \in \llbracket 1, m \rrbracket \end{cases}} \quad (22)$$

Les relations

$$\boxed{\lambda_i g_i(x^*) = 0, \quad \forall i \in \llbracket 1, m \rrbracket} \quad (23)$$

portent le nom de *conditions de complémentarité*. Elles indiquent que pour chaque contrainte, soit la contrainte est "saturée" ou encore dite "active" c'est-à-dire $g_i(x^*) = 0$, soit son multiplicateur est nul $\lambda_i = 0$.

Remarques.

- Il y a $n + m$ inconnues et autant d'équations. On peut résoudre ce système non-linéaire par des techniques spécifiques mais dans la pratique les relations de KKT sont peu utilisées car le système est difficile à résoudre et présente souvent beaucoup d'autres solutions qu'un minimum (tous les points critiques).
- On peut montrer aussi que les multiplicateurs de Lagrange dans (22) sont nécessairement *négatifs* (voir Annexe B).

Annexe A. Qualification des contraintes.

Pour que les conditions de KKT soient valables au point x^* , il faut que les contraintes soient “qualifiées”. Pour pouvoir appliquer le théorème des extrema liés avec la contrainte égalité $h(x, e) = 0$, il faut que les gradients $\nabla h_1(x^*, e^*), \dots, \nabla h_m(x^*, e^*)$ soient *linéairement indépendants* c'est-à-dire :

$$\sum_{i=1}^m \alpha_i \nabla h_i(x^*, e^*) = 0 \Rightarrow \alpha_i = 0, \forall i \in \llbracket 1, m \rrbracket \quad (24)$$

Soit les réels $\alpha_1, \dots, \alpha_m$ tels que

$$\sum_{i=1}^m \alpha_i \nabla h_i(x^*, e^*) = 0$$

Avec l'expression (21) du gradient de h_i , on obtient

$$\left(\sum_{i=1}^m \alpha_i \nabla g_i(x^*), 2\alpha_1 e_1^*, 2\alpha_2 e_2^*, \dots, 2\alpha_m e_m^* \right) = 0$$

On en déduit ainsi

$$\begin{cases} \sum_{i=1}^m \alpha_i \nabla g_i(x^*) = 0 \\ 2\alpha_i e_i^* = 0, \forall i \in \llbracket 1, m \rrbracket \end{cases} \Leftrightarrow \begin{cases} \sum_{i=1}^m \alpha_i \nabla g_i(x^*) = 0 \\ \alpha_i g_i(x^*) = 0, \forall i \in \llbracket 1, m \rrbracket \end{cases} \quad (25)$$

On introduit l'ensemble d'indices

$$\boxed{I(x^*) = \{i \in \llbracket 1, m \rrbracket \text{ tel que } g_i(x^*) = 0\}} \quad (26)$$

qui correspond aux contraintes "actives" ou "saturées".

D'après (25), on a $\alpha_i = 0, \forall i \notin I(x^*)$ i.e. tels que $g_i(x^*) < 0$. On a alors

$$0 = \sum_{i=1}^m \alpha_i \nabla g_i(x^*) = \sum_{i \in I(x^*)} \alpha_i \nabla g_i(x^*)$$

On obtient ainsi la propriété suivante.

Contraintes qualifiées

Les gradients $\nabla h_1(x^*, e^*), \dots, \nabla h_m(x^*, e^*)$ sont *linéairement indépendants* si et seulement si

$$\boxed{\text{la famille } \{\nabla g_i(x^*)\}_{i \in I(x^*)} \text{ est libre}} \quad (27)$$

i.e. les gradients des contraintes qui sont “actives” (celles où $g_i(x^*) = 0$) sont linéairement indépendants.

Si les contraintes vérifient (27), on dit que *les contraintes sont qualifiées* en x^* .

Remarque. En fait, la définition usuelle dans la littérature de la qualification des contraintes est la suivante. Les contraintes sont dites qualifiées en x^* s'il existe une direction $d_0 \in \mathbb{R}^n$ (appelée direction rentrante) telle que

$$(\nabla g_i(x^*) \mid d_0) < 0, \quad \forall i \in I(x^*). \quad (28)$$

On peut montrer que si l'hypothèse (27) est vérifiée i.e. si les gradients $\nabla g_i(x^*)$ pour $i \in I(x^*)$ sont linéairement indépendants alors il existe d_0 vérifiant (30).

Cas (important) d'une seule contrainte $g(x) \leq 0$.

Pour appliquer le théorème des extrema liés avec la contrainte $h(x^*, e^*) := g(x^*) + (e^*)^2 = 0$, il faut que $\nabla h(x^*, e^*) \neq 0$. On a

$$\nabla h(x^*, e^*) \neq 0 \Leftrightarrow \boxed{\nabla g(x^*) \neq 0 \text{ ou } g(x^*) \neq 0} \quad (29)$$

C'est la condition de *contrainte qualifiée* dans le cas d'une seule contrainte.

Remarques.

- Si tous les gradients $\nabla g_1(x^*), \dots, \nabla g_m(x^*)$ sont linéairement indépendants, les conditions (25) sont vérifiées et les contraintes sont qualifiées.

☞ En résumé, on a obtenu le résultat suivant pour les contraintes inégalités.

Contraintes inégalités : conditions de KKT.

Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ et $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ de classe C^1 . On suppose que f admet un minimum local x^* sur $V = \{x \in \mathbb{R}^n, g(x) \leq 0\}$. Si les contraintes dans V sont *qualifiées* i.e. si les gradients $\nabla g_i(x^*)$ pour $i \in I = \{i \in \llbracket 1, m \rrbracket \text{ tel que } g_i(x^*) = 0\}$ sont linéairement indépendants^a, alors il existe des réels $\lambda_1, \dots, \lambda_m$ tels que

$$\begin{cases} \nabla f(x^*) = \lambda_1 \nabla g_1(x^*) + \dots + \lambda_m \nabla g_m(x^*) \\ \lambda_i g_i(x^*) = 0, \forall i \in \llbracket 1, m \rrbracket \end{cases}$$

a. ou bien si la condition $(\nabla g(x^*) \neq 0 \text{ ou } g(x^*) \neq 0)$ est vérifiée dans le cas d'une seule contrainte ($m = 1$).

Annexe B. *Contraintes inégalités : négativité des multiplicateurs de Lagrange.*

On établit la négativité du multiplicateur de Lagrange dans le cas d'une seule contrainte inégalité $g(x) \leq 0$ ¹.

Soit x^* un minimum de f avec $g(x^*) \leq 0$. On suppose que les contraintes sont qualifiées i.e.

$$\nabla g(x^*) \neq 0 \text{ ou } g(x^*) \neq 0. \quad (30)$$

Alors (cf. conditions KKT) il existe $\lambda \in \mathbb{R}$ tel que

$$\begin{cases} \nabla f(x^*) = \lambda \nabla g(x^*) \\ \lambda g(x^*) = 0 \text{ (condition de complémentarité)} \end{cases} \quad (31)$$

On suppose que $\lambda \neq 0$. D'après les conditions de complémentarité, on a

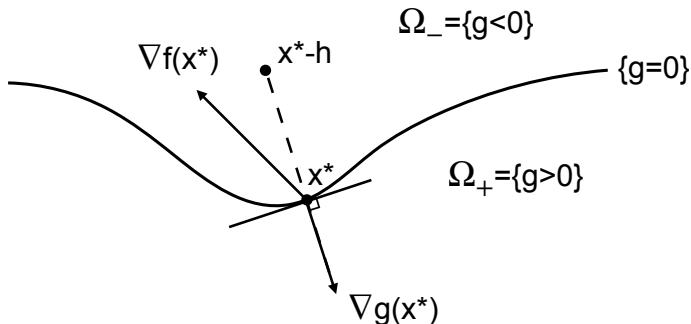
$$g(x^*) = 0. \quad (32)$$

1. le cas de plusieurs contraintes est un peu plus délicat à démontrer...

Avec (30), on a nécessairement

$$\nabla g(x^*) \neq 0. \quad (33)$$

☛ On rappelle que le gradient $\nabla g(x^*)$ est perpendiculaire à l'ensemble de niveau $\{g = 0\}$ qui sépare l'espace entier $\mathbb{R}^n$ en deux ensembles $\Omega_+ = \{g > 0\}$ et $\Omega_- = \{g < 0\}$. De plus, le gradient $\nabla g(x^*)$ est dirigé vers $\Omega_+ = \{g > 0\}$ (direction de plus forte pente).



On montre que

$$\boxed{(\nabla f(x^*) \mid \nabla g(x^*)) \leq 0} \quad (34)$$

d'où, avec (31) et (33) :

$$\boxed{\lambda \leq 0}$$

Preuve de (34).

Supposons que $(\nabla f(x^*) \mid \nabla g(x^*)) > 0$.

Soit $h = \tau \nabla g(x^*) \neq 0$ avec $\tau > 0$.

▪ On a $g(x^* - h) = \underbrace{g(x^*)}_{=0} - (\nabla g(x^*) \mid h) + \|h\|\varepsilon(h)$ où $\varepsilon(h) \rightarrow 0$

quand $\|h\| \rightarrow 0$, soit

$$g(x^* - h) = -\tau \|\nabla g(x^*)\|^2 + \tau \varepsilon(\tau)$$

où $\varepsilon(\tau) \rightarrow 0$ quand $\tau \rightarrow 0$. Puisque $\nabla g(x^*) \neq 0$, pour $\tau > 0$ suffisamment petit, on obtient

$$g(x^* - h) < 0 \quad \text{i.e. } x^* - h \in \Omega_-.$$

■ On a aussi

$$f(x^* - h) = f(x^*) - (\nabla f(x^*) | h) + \|h\|\varepsilon(h)$$

soit

$$f(x^* - h) - f(x^*) = -\tau(\nabla f(x^*) | \nabla g(x^*)) + \tau\varepsilon(\tau)$$

Compte tenu de l'hypothèse initiale $(\nabla f(x^*) | \nabla g(x^*)) > 0$, pour τ suffisamment petit, on a

$$f(x^* - h) < f(x^*) \quad \text{avec} \quad g(x^* - h) < 0$$

ce qui contredit le fait que x^* est le minimum de f sous la contrainte $g(x^*) \leq 0$.

