

---

# Optimisation Non Linéaire Sous Contraintes

---

**A. Bazeniari**

**Enseignant chercheur  
Centre universitaire Abdelhafid Boussouf  
Mila, Algérie**

Se reporter à des manuels de base et à certaines recherches

Septembre 2023

# Chapitre 1

## Conditions d'optimalité

### 1.1 Introduction

L'approche considérée ici pour l'obtention de ces conditions est basée sur les notions de descente et de direction admissible. L'étude de ces conditions a permis de développer les algorithmes de résolution et de vérifier la validité des résultats obtenus.

Dans ce chapitre, les fonctions sont toujours supposées différentiables à tout ordre. Si  $f$  est définie sur  $S \subset \mathbb{R}^n$  et à valeurs réelles, sa différentielle en  $x \in S$  est l'application linéaire notée

$$f(x+h) = f(x) + \langle \nabla f(x), h \rangle + \|h\| \varepsilon(h), \quad \varepsilon(h) \rightarrow 0 \text{ lorsque } h \rightarrow 0. \quad (1.1)$$

Rappelons la formule de Taylor à l'ordre 2

$$f(x+h) = f(x) + \langle \nabla f(x), h \rangle + \frac{1}{2} \langle \nabla^2 f(x), h \rangle h + \|h\|^2 \varepsilon(h), \quad \varepsilon(h) \rightarrow 0 \text{ lorsque } h \rightarrow 0. \quad (1.2)$$

## 1.2 Direction admissible

**Définition 1.2.1.** (*Direction admissible*). Soit  $f : S \subset \mathbb{R}^n \rightarrow \mathbb{R}$  et  $x \in S$ .  $d$  est une direction admissible si

$$\exists \lambda' > 0 \text{ tel que } x + \lambda d \in S, \text{ pour } 0 \leq \lambda \leq \lambda'.$$

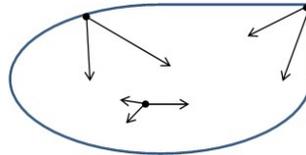


FIGURE 1.1 – Directions admissible

- Remarque 1.2.1.**
1. Cette définition nous montre que tous les vecteurs  $d$  de  $\mathbb{R}^n$  sont des directions admissibles, si le point  $x + \lambda d$  est un point intérieur de  $S$ .
  2. La nature de l'ensemble  $S$  intervient dans le formalisme d'équations ou d'inéquations de la condition d'optimalité, à savoir les points sur la frontière de  $S$ .

### 1.2.1 Direction admissible à l'optimum

**Théorème 1.2.1.** (*condition nécessaire de Peano-Kantorovitch*). Soit  $f : S \subset \mathbb{R}^n \rightarrow \mathbb{R}$  de classe  $C^1$ . Si  $x^*$  est un minimum local (globale) de  $f$  sur  $S$ , alors  $\forall d$  direction admissible en  $x^*$ , on a

$$\nabla^T f(x^*)d \geq 0. \quad (1.3)$$

*Démonstration.* Soit  $d$  une direction admissible en  $x^*$ . Considérons la fonction  $g(\lambda) = f(x^* + \lambda d)$ .

$\forall x \in S$  on a  $f(x^*) \leq f(x)$ . D'où  $\lambda^* = 0$  est un minimum local de  $g(\lambda)$  sur  $[0, \lambda']$ . On a

$$g(\lambda) = g(0) + \lambda g'(0) + o(\lambda),$$

et comme  $x^*$  est un minimum local, alors  $g(\lambda) - g(0) \geq 0$ .

Par conséquent  $g'(\lambda) = [\nabla f(x^* + \lambda d)]^T d \geq 0$ ,

on obtient  $g'(0) = \nabla^T f(x^*)d \geq 0$ . □

## 1.3 Existence et unicité de la solution

**Définition 1.3.1.**

1. Une partie  $U$  de  $S$  est un ouvert de  $S$  si pour tout  $x \in U$  il existe  $\epsilon > 0$  tel que  $B(x, \epsilon) \subset U$ .

2. Une partie  $F$  de  $S$  est un fermé de  $S$  si et seulement si son complémentaire  $F^c$  dans  $S$  est ouvert.

**Remarque 1.3.1.** Dans  $\mathbb{R}$ , les intervalles ouverts sont des ouverts et les intervalles fermés sont des fermés. Plus généralement, toute boule ouverte est une partie ouverte et toute boule fermée est une partie fermée.

**Théorème 1.3.1.** (Existence : Théorème de Weierstrass).

1. Si  $S$  est un sous-ensemble fermé et borné de  $\mathbb{R}^n$  et  $f$  continue, alors  $f$  admet au moins un point extremum.
2. Si  $S$  est un sous-ensemble fermé de  $\mathbb{R}^n$  et si  $f$  est croissante à l'infini ( $f$  est coercive si  $f(x) \rightarrow +\infty$  lorsque  $\|x\| \rightarrow +\infty$ ), alors  $f$  admet un point minimum globale sur  $S$ .

*Démonstration.* 1.  $S$  est fermé et borné de  $\mathbb{R}^n$  alors compact, et comme  $f$  est continue alors elle atteint ses bornes sur  $S$ . Donc  $x^*$  existe.

2. Soit l'ensemble  $S_0 = \{x, x_0 \in S / f(x) \leq f(x_0)\}$ , on voit que  $S_0$  est un compact (fermé car  $f$  est continue, borné car  $f$  est coercive). Donc  $x^*$  existe.

□

**Théorème 1.3.2.** (Unicité). Soit le problème de minimisation (PNL) avec  $S$  convexe et  $S \subset \mathbb{R}^n \rightarrow \mathbb{R}$  convexe. Alors,

1. Tout minimum local de (PNL) est un minimum global de problème.
2. Si  $f$  est strictement convexe, il y a au plus un minimum de problème.

**Exemple 1.3.1.** Soit le problème (PNL) :

$$\begin{cases} \text{Min } f(x) = x^2 \\ g : x^4 - 1 \leq 0 \\ x \in \mathbb{R}^n, \end{cases} .$$

L'ensemble  $g$  est fermé (car  $g^{(-1)}([-1,1])$  est fermé) et  $f$  croissante à l'infinie. Comme  $f$  est continue, elle atteint ses bornes et elle admet un minimum global  $x^*$ .

## 1.4 Conditions nécessaires d'optimalité

L'idée de cette section est basée sur la recherche d'une expression plus pratique de la condition (1.3).

### 1.4.1 Contraintes de type égalité

Soit l'ensemble  $S$  définie par,

$$S = \{x \in \mathbb{R}^n : h_1(x) = 0, \dots, h_p(x) = 0\}.$$

- On dit que  $h$  réalise  $p$  contraintes et  $x$  est dit variable de décision.
- Cette ensemble des contraintes d'qualités est un sous-ensemble de  $\mathbb{R}^n$  et qui représente une hypersurface.

### 1.4.1.1 Plan tangent

**Définition 1.4.1.** On définit le plan tangent au point  $x^*$ , tout vecteur  $y$  qui est orthogonal avec les gradients de  $h_i$ , on écrit

$$T = \{y, \nabla h(x^*)y = 0\}. \quad (1.4)$$

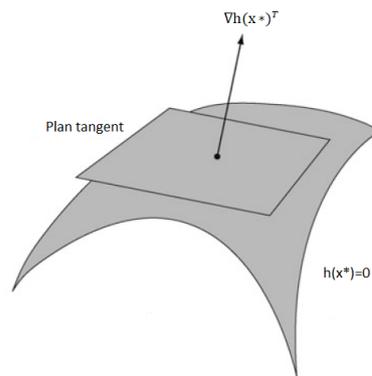


FIGURE 1.2 – Plan tangent.

**Exemple 1.4.1.** Soit la contrainte  $h(x_1, x_2) = x_1$ . Le plan tangent est l'axe des  $x_2$ , car

$$(1, 0)^T (y_1, y_2) = 0 \Rightarrow y_1 = 0.$$

**Définition 1.4.2.** On dit que  $x^*$  est un point régulier pour la contrainte  $h(x) = 0$  si,

- $h(x^*) = 0$ ,
- Les vecteurs  $\nabla h_i(x^*)$  sont linéairement indépendants.

**Théorème 1.4.1.** (CN1) Soit  $f : S \subset \mathbb{R}^n \rightarrow \mathbb{R}$  de classe  $C^1$ . On suppose que  $f$  admet en  $x^* \in S$  un minimum local et que,

$$\text{La famille } (\nabla h_1(x^*), \dots, \nabla h_p(x^*)) \text{ est libre} \quad (*)$$

Alors il existe des multiplicateurs de Lagrange  $\lambda_1, \lambda_2, \dots, \lambda_p \in \mathbb{R}$  tel que,

$$\nabla f(x^*) + \sum_{i=1}^p \lambda_i \nabla h_i(x^*) = 0. \quad (**)$$

(\*\*) Est appelée équation de Lagrange.

(\*) Condition de qualification des contraintes.

**Remarque 1.4.1.** — La recherche des points stationnaires du problème se ramène à la résolution du système (si la solution existe) de l'équation (\*\*).

- Les gradients  $\nabla f$  et  $\nabla h$  sont orthogonaux aux courbes de niveau de  $f$  et  $h$  respectivement.
- Si  $x^*$  est extremum de  $f$  sous la contrainte  $h$  et les gradients  $\nabla f(x^*)$  et  $\nabla h(x^*)$  non nuls, les courbes de niveau de  $f$  et  $h$  sont tangentes en  $x^*$ .

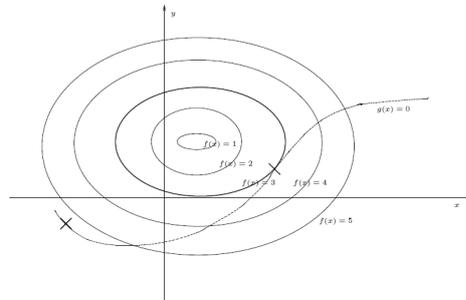


FIGURE 1.3 – Interprétation géométrique des multiplicateurs de Lagrange.

**Remarque 1.4.2.** Les multiplicateurs de Lagrange mesurent combien la fonction objectif doit être modifiée pour obtenir la solution optimale tout en respectant les contraintes.

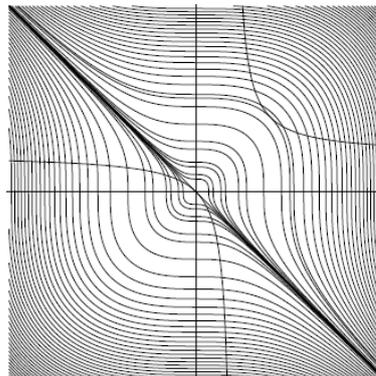


FIGURE 1.4 – Les courbes de niveau de  $f$  et  $h$  sont tangentes.

*Démonstration.* On se contente de l'ensemble  $S = \{x \in \mathbb{R}^2 : h(x) = 0\}$ . La généralisation sur  $\mathbb{R}^n$  reste la même.

La condition (\*) indique que  $\nabla h(x^*) = 0$ . On suppose que  $\partial h / \partial x_2 \neq 0$ .

Ceci nous permet d'appliquer le théorème des fonctions implicite,  $h(x_1, x_2) = 0 \Rightarrow x_2 = \alpha(x_1)$ ,  $\alpha$  une fonction continue et dérivable.

Maintenant, appliquant les équations d'Euler sur  $f$  et  $h$ , pour  $i = \{1, \dots, n - 1\}$ , on aura

$$\partial f / \partial x_1(x_1^*, \alpha(x_1^*)) + \partial f / \partial y(x_1^*, \alpha(x_1^*)) \alpha'(x_1^*) = 0 \tag{1.5}$$

$$\partial h / \partial x_1(x_1^*, \alpha(x_1^*)) + \partial h / \partial x_2(x_1^*, \alpha(x_1^*)) \alpha'(x_1^*) = 0 \tag{1.6}$$

De (1.5) et (1.6) on trouve,

$$\nabla f(x^*) + \lambda \nabla h(x^*) = 0,$$

Tel que,

$$\lambda = [\partial h / \partial x_2(x^*)]^{-1} \times \partial h / \partial x_2(x^*).$$

□

**Exemple 1.4.2.**  $f(x, y, z) = (x - 2)^2 + y^2 + z^2$  sous la contrainte  $g : x^2 + 2y^2 + 3z^2 = 1$ .

Le gradient est,  $\nabla g(x, y, z) = (2x, 4y, 6z)^T$ ,

$\nabla g(x, y, z) \neq 0$  car la contrainte  $g$  n'est pas active au point  $(0, 0, 0)$ . La contrainte est qualifiée.

Les équations de Lagrange sont,

$$\begin{cases} x^2 + 2y^2 + 3z^2 = 1 \\ x - 2 = \lambda x \\ y = 2\lambda y \\ z = 3\lambda z \end{cases}$$

### 1.4.2 Interprétation de la valeur du multiplicateur $\lambda$ (analyse de sensibilité)

Soit le problème  $(p)$ , ( $f$  dif.)

$$\begin{cases} \text{Min } f(x) \\ \text{s.c } h(x) = b \\ x \in \mathbb{R}^n \end{cases}$$

Considérons la fonction  $V(b)$ , qui donne la valeur optimale du problème  $(p)$  pour une valeur  $b$  donnée. Si on varie la valeur de  $b$ , alors la valeur de la solution optimale varie. A partir de ça, on peut définir la fonction  $x(b)$  comme,

$$V(b) = f(x(b)) \text{ et } h(x(b)) = b.$$

Maintenant, Mesurant la variation de  $V(b)$ ,

$$\begin{aligned} V'(b) &= \frac{dV}{db}, \\ &= \frac{\partial f}{\partial x_1} \frac{dx_1}{db} + \dots + \frac{\nabla f}{\nabla x_n} \frac{dx_n}{db}, \\ &= \lambda \left( \frac{\partial h}{\partial x_1} \frac{dx_1}{db} + \dots + \frac{\nabla h}{\nabla x_n} \frac{dx_n}{db} \right), \\ &= \lambda \frac{db}{db}, \\ &= \lambda. \end{aligned}$$

**Conséquence :** le multiplicateur  $\lambda$  est sensible à la variation de la valeur de  $b$  et il donne l'information sur le taux de variation de la valeur de la fonction objective lorsque  $b$  augmente. Un multiplicateur élevé indique que la contrainte est contraignante (l'ajustement des ressources est coûteux), tandis qu'un multiplicateur faible indique que la contrainte est moins importante (l'ajustement est peu coûteux).

**Calcul directe :** On peut approcher la fonction  $V$  par son polynôme autour de  $b'$ ,

$$V(b) = V(b') + V'(b')(b - b') = V(b') + \lambda(b - b').$$

**Remarque 1.4.3.** L'interprétation géométrique des multiplicateurs les associe à la tangence des contraintes, tandis que leur interprétation économique les considère comme des coûts marginaux ou des prix d'ajustement.

### 1.4.3 Contraintes de type inégalité

Soit l'ensemble des contraintes  $S$  définit par,

$$S = \{x \in \mathbb{R}^n : g_1(x) \leq 0, \dots, g_m(x) \leq 0\}.$$

**Définition 1.4.3.** On dit qu'une contrainte  $g$  est active en  $x^*$  si,

- $g(x^*) = 0$ ,
- L'ensemble des indices des contraintes actives est,

$$I(x^*) = \{i / g_i(x^*) = 0\}.$$

- L'ensemble des vecteurs  $E(x)$  qui active  $g$  est,

$$E(x) = \{x \in \mathbb{R}^n, g_j(x) = 0, j \in I\}.$$

**Théorème 1.4.2.** (CN1) Soit  $f : S \subset \mathbb{R}^n \rightarrow \mathbb{R}$  de classe  $C^1$ . On suppose que  $f$  admet en  $x^* \in S$  un minimum local et qu'il existe un vecteur  $v \in \mathbb{R}^n$  tel que,

$$\forall j g_j(x^*) = 0 \Rightarrow \langle \nabla g_j(x^*), v \rangle \leq 0 \quad (*)$$

Alors il existe des multiplicateurs de Kuhn et Tucker  $\mu_1, \mu_2, \dots, \mu_m \in \mathbb{R}$  tel que,

$$\nabla f(x^*) + \sum_{j=1}^m \mu_j \nabla g_j(x^*) = 0. \quad (**)$$

Avec

$$\begin{cases} \mu_j \geq 0, & \text{(Positivité)} \\ \mu_j g_j(x^*) = 0, & \text{(Relation d'exclusion).} \end{cases}$$

(\*\*) Est appelée équation de Lagrange.

(\*) Condition de qualification des contraintes.

*Démonstration.* 1. Les deux relations d'exclusion et (\*\*) sont une conséquence directe du théorème 1.4.1 ( car il suffit de prendre  $\mu_j = 0$  pour  $j \notin I$ ).

2. Maintenant, on montre la positivité par absurde :

Supposons qu'il existe  $\mu_k < 0, k \in I$ .

D'autre part, on définit  $y \in \mathbb{R}^n$  tel que :

$$\nabla g_j(x^*)y = 0, \quad j \in I,$$

$$\nabla g_k(x^*)y = -1, \quad j \neq k.$$

Alors,  $y$  est une direction admissible,

$$\begin{aligned} \nabla f(x^*)^T y &= - \sum_{j=1}^m \mu_j \nabla g_j(x^*)^T y, \\ &= -\mu_k \nabla g_k(x^*)^T y, \\ &= \mu_k < 0. \end{aligned}$$

Ceci est impossible car  $x^*$  est un point minimum de  $f$ .

□

**Exemple 1.4.3.** Soit la fonction  $f(x, y) = -xy$  à minimiser sur le domaine  $S$

$$x \geq 0, y \geq 0 \text{ et } x + y \leq 6.$$

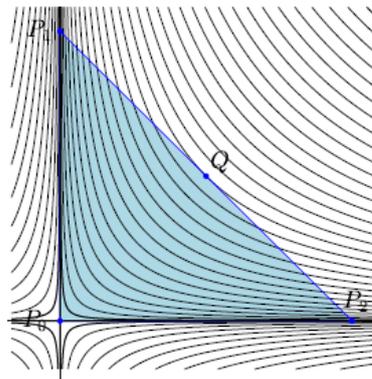


FIGURE 1.5 – Les lignes de niveau de  $-xy$  sur le domaine  $S$

### 1.4.4 Le théorème de Kuhn-Tucker avec lagrangien généralisé

Dans le cadre général du théorème de Kuhn et Tucker, la contrainte  $S$  est de la forme,

$$S = \{x \in \mathbb{R}^n, g_j(x) \leq 0, h_i(x) = 0\}.$$

On suppose que  $f$  admet en  $x^* \in X$  un minimum local et qu'il existe un vecteur  $v \in \mathbb{R}^n$  tel que,

$$\langle \nabla h_i(x^*), v \rangle = 0,$$

$$\forall j \in I(x^*) \langle \nabla g_j(x^*), v \rangle \leq 0.$$

Où  $I(x^*)$  l'ensemble des indices des contraintes actives au point  $x^*$ .

Alors il existe des multiplicateurs  $\lambda_1, \lambda_2, \dots, \lambda_p \in \mathbb{R}$  et  $\mu_1, \mu_2, \dots, \mu_m \in \mathbb{R}$  tel que,

$$\nabla f(x^*) + \sum_{i=1}^p \lambda_i \nabla h_i(x^*) + \sum_{j=1}^m \mu_j \nabla g_j(x^*) = 0.$$

Sous les conditions des deux théorèmes 2.3.1 et 2.3.2.

## 1.5 Conditions suffisantes d'optimalité

Rappelant les deux cas suivants :

Le Lagrangien associé au (PCE) est :

$$L(x, \lambda) = f(x) + \sum_{i=1}^p \lambda_i h_i(x).$$

Le Lagrangien associé au (PCI) est :

$$L(x, \lambda) = f(x) + \sum_{j=1}^m \lambda_j g_j(x).$$

**Théorème 1.5.1.** Soit  $x^* \in \mathbb{R}^n$  un point régulier. Alors il existe  $\lambda^*$  tel que,

- $\nabla L(x^*, \lambda^*) = 0$
- $d^T \nabla_{xx}^2 L(x^*, \lambda^*) d \geq 0, d \neq 0$  direction admissible

*Démonstration.* Puisque  $x^*$  est minimum, on a  $\nabla_{xx}^2 f(x^*) \geq 0$  Et donc par les équations d'Euler,

$$d^T \nabla_{xx}^2 f(x^*) d + \nabla f(x^*)^T d' \geq 0. \quad (1.7)$$

D'autre part,

$$d^T \nabla_{xx}^2 h(x^*) d + \nabla h(x^*)^T d' = 0. \quad (1.8)$$

En multipliant par  $\lambda_i$  la relation 1.8 puis on fait la somme avec la relation 1.7, on aura

$$d^T (\nabla_{xx}^2 (f(x^*) + \sum_{i=1}^p \nabla^2 h_i(x^*) \lambda_i) d + (\nabla f(x^*) + \sum_{i=1}^p \lambda_i \nabla h_i(x^*)^T) d' \geq 0.$$

Et comme  $\nabla f(x^*) + \sum_{i=1}^p \lambda_i \nabla h_i(x^*)^T = 0$ . L'inégalité est vérifiée.

□

**Théorème 1.5.2.** (CS) pour (PCE)

Soit  $x^* \in \mathbb{R}^n$  et  $\lambda^* \in \mathbb{R}^p$  vérifiant,

- $\nabla f(x^*) + \sum_{i=1}^p \lambda_i^* \nabla h_i(x^*) = 0,$
- $h_i(x^*) = 0,$
- $d^T \nabla_{xx}^2 L(x^*, \lambda^*) d \geq 0, d \neq 0.$

Alors  $x^*$  est un point minimum du problème.

**Théorème 1.5.3.** (CS) pour (PCI)

Soit  $x^* \in \mathbb{R}^n$  et  $\mu^* \in \mathbb{R}^m$  vérifiant,

- $\nabla f(x^*) + \sum_{j=1}^m \mu_j^* \nabla g_j(x^*) = 0,$
- $g_j(x^*) \leq 0,$
- $\mu_j \geq 0,$
- $\mu_j g_j(x^*) = 0,$
- $d^T \nabla_{xx}^2 L(x^*, \lambda^*) d \geq 0, d \neq 0.$

Alors  $x^*$  est un point minimum du problème.

## Chapitre 2

# Méthodes de résolution

### 2.1 Introduction

Les méthodes de résolution de programmes mathématiques non linéaires s'est développée rapidement depuis 1955. Beaucoup de méthodes proposées pour la résolution des problèmes avec contraintes conduisent à remplacer le problème donné par une suite de problèmes sans contraintes, dont les résolutions utilisent des méthodes classiques. Dans le reste du chapitre on définit le problème d'optimisation comme,

$$\begin{cases} \text{Min } f(x) \\ \text{S.C. } x \in S \end{cases}$$

Avec  $f$  une fonction convexe et  $S$  convexe et fermé.

## 2.2 Notions algorithmiques utiles

**Définition 2.2.1.** Soit l'application vectoriel  $M : \mathbb{R}^{\mathbb{N}} \rightarrow \mathbb{R}^{\mathbb{N}}$ .  $(x^n)_{n \in \mathbb{N}}$  est la suite générée par cette application comme suit,

$$x^{n+1} = M(x^n), \quad x^0 \text{ est la phase initiale.} \quad (2.1)$$

**Définition 2.2.2.** On dit qu'un algorithme est convergent s'il atteint son minimiseur  $x$ . Leur efficacité est mesurée par,

- **vitesse de convergence** : qui mesure la rapidité d'atteindre  $x$  cherché.
- **complexité** : qui mesure le coût par chaque itération afin d'obtenir la solution  $x$  souhaitée, et ça avec une précision  $\epsilon$  fixée.

**Remarque 2.2.1.** 1. Le critère d'arrêt de l'algorithme peut être exprimé de la forme suivante,

$$\| \nabla f(x^{(n)}) \| \leq \epsilon. \quad (2.2)$$

2. La convergence est linéaire si,

$$\exists C \in [0, 1[, \forall x^{(0)} \in \mathbb{R}^{\mathbb{N}}, \text{err}^{(n+1)} \leq C \text{err}^{(n)}, \quad (2.3)$$

avec  $\text{err}^{(n)} = \| x - x^{(n)} \|$ , l'erreur à l'iteration  $n$ .

3. La convergence est d'ordre  $p$  si elle satisfait la relation,

$$\exists C \in [0, 1[, \forall x^{(0)} \in \mathbb{R}^{\mathbb{N}}, \text{err}^{(n+1)} \leq C (\text{err}^{(n)})^p, \quad (2.4)$$

si  $p = 2$  la vitesse est dite quadratique.

**Algorithmes de descente.** Les algorithmes que nous allons considérer pour les problèmes d'optimisation ont la forme générale suivante :

$$x^{(0)} \text{ étant donné, calculer } x^{(n+1)} = x^{(n)} + t_n d^{(n)}. \quad (2.5)$$

Le vecteur  $d^{(n)}$  s'appelle la direction de descente, et le réel  $t_k > 0$  le pas de la méthode à la  $n$ -ième itération. On pratique, on choisira la direction et le pas afin que l'inégalité suivante soit satisfaite :

$$f(x^{(n+1)}) \leq f(x^{(n)}).$$

De tels algorithmes sont appelés **algorithmes de descente**.

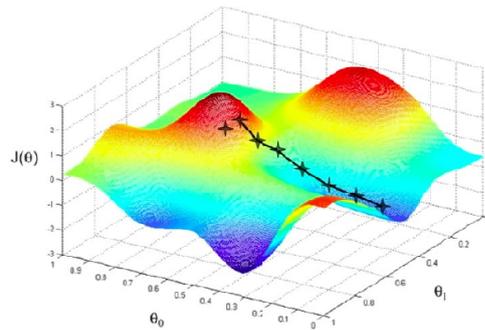


FIGURE 2.1 – Descente en gradient

**Proposition 2.2.1.** (Interprétation géométrique du gradient)  $f : S \subset \mathbb{R}^N \rightarrow \mathbb{R}^N$  de classe  $C^1$ .

- L'opposé du gradient,  $-\nabla f(x)$ , est une direction de descente en  $x$  (si  $\nabla f(x) \neq 0$ ).
- La direction du gradient est la direction de plus forte pente (si  $\|d\| = 1$ , alors la pente dans la direction  $d$  correspond à la dérivée de  $h$  en 0).
- Le gradient est orthogonal à la tangente à la courbe en  $x$ .

**Algorithmes de gradient.** Appliquant l'algorithme de descente décrit au dessus. La formule de Taylor en  $x^{(n)}$  est,

$$f(x^{(n+1)}) = f(x^{(n)}) + \langle \nabla f(x^{(n)}), t_n d^{(n)} \rangle + o(t_n d^{(n)}).$$

Pour s'assurer que  $f(x^{(n+1)}) - f(x^{(n)}) \leq 0$ ,  
le choix de  $d^{(n)}$  est donc,

$$d^{(n)} = -\nabla f(x^{(n)}). \quad (2.6)$$

**Remarque 2.2.2.** Le pas  $t_k$  peut être choisi en tant que pas fixe ou pas optimal (vitesse de convergence).

## 2.3 Méthodes de Linéarisation

Parmi les méthodes de Linéarisation on trouve,

- Méthodes Tangentielle (centres linéarisés, Frank Wolf),
- Méthodes Barycentriques.

Dans cette section on choisit de développer la méthode de Frank Wolf.

### 2.3.1 Méthode de Frank Wolf (1956)

On applique la méthode pour la résolution des programmes à fonction objective quadratique et à contraintes linéaires.

### 2.3.1.1 Principe de la méthode

Le choix de la direction de descente consiste à linéariser en  $x^{(k)}$  la fonction  $f(x)$  en construisant une fonction linéaire tel que,

$$f(x^{(k+1)}) = f(x^{(k)}) + \nabla f(x^{(k)})^T (x^{(k+1)} - x^{(k)}). \quad (2.7)$$

Puis de minimiser cette fonction sur le domaine réalisable  $S$ .

### 2.3.1.2 Construction de la Méthode

- Soit  $x^{(0)}$  un point arbitraire et  $x^{(k)} \in S$ . Pour construire  $x^{(k+1)}$  on a,

$$\begin{aligned} x^{(k+1)} &= x^{(k)} + t_k d^{(k)}, \\ &= x^{(k)} + t_k (y^{(k)} - x^{(k)}), \\ &= (1 - t_k) x^{(k)} + t_k y^{(k)}. \end{aligned}$$

$x^{(k+1)}$  est en combinaison linéaire avec  $x^{(k)}$  et  $y^{(k)}$ .

- Maintenant, soit le polynôme de Taylor de degré 1 autour de  $x^{(k)}$ ,

$$\begin{aligned} f(y) &= f(x^{(k)}) + \nabla f(x^{(k)})^T (y - x^{(k)}), \\ &= f(x^{(k)}) + \nabla f(x^{(k)})^T y - \nabla f(x^{(k)})^T x^{(k)}. \end{aligned}$$

Comme  $x^{(k)}$  est fixe, alors minimiser  $f(y)$  est équivalent à minimiser,

$$\nabla f(x^{(k)})^T y. \quad (2.8)$$

**Conséquence :**  $y^{(k)}$  est la solution optimale du problème  $P_k$  suivant,

$$\begin{cases} \text{Min } \nabla f(x^{(k)})^T y \\ \text{S.C. } y \in S \end{cases}$$

**Question :**  $d^{(k)}$  est-elle une direction de descente ?

$$\begin{aligned} \nabla f(x^{(k)})^T d^{(k)} &= \nabla f(x^{(k)})^T (y^{(k)} - x^{(k)}), \\ &= \nabla f(x^{(k)})^T y^{(k)} - \nabla f(x^{(k)})^T x^{(k)}, \\ &< 0. \end{aligned}$$

Car  $y^{(k)}$  est améliorante.

- On détermine Le pas  $t_k^*$  qu'est une solution du problème,

$$f((1 - t_k^*)x^{(k)} + t_k^*y^{(k)}) = \min_{0 \leq t_k \leq 1} \{f((1 - t_k)x^{(k)} + t_k y^{(k)})\}. \quad (2.9)$$

- Le critère d'arrêt de la méthode est que  $x^{(k)}$  soit solution du problème  $P_k$  i.e.

$$\nabla f(x^{(k)})^T (y^{(k)} - x^{(k)}) \geq 0. \quad (2.10)$$

### 2.3.1.3 Algorithme de Frank Wolfe

1. Soit  $x^{(0)} \in S$ ; point arbitraire.
2. **Pour**  $k = 0$  à  $N$  **faire**  
 $r^{(k)} := \nabla f(x^{(k)})$ .  
 $y^{(k)} := \arg \min_{y \in S} \langle y, r^{(k)} \rangle$ .  
 $g_k := \langle x^{(k)} - y^{(k)}, r^{(k)} \rangle$ .
3. **Si**  $g_k < \epsilon$  **alors** renvoyer  $x^{(k)}$ .
4.  $t_k^* := \arg \min_{0 \leq t_k \leq 1} f((1 - t_k)x^{(k)} + t_k y^{(k)})$ .
5.  $x^{(k+1)} := (1 - t_k^*)x^{(k)} + t_k^* y^{(k)}$ .
6. **Fin pour.**

**Exemple 2.3.1.** Soit le problème suivant,

$$\begin{cases} \text{Min } f(x) = 1/2(x_1^2 + x_2^2) \\ -x_1 + x_2 \leq 7 \\ x_1 + x_2 \leq 5 \\ -x_2 \leq -2 \end{cases}$$

Avec  $x^{(0)} = (-2, 3)$ .

**Question :** calculer  $x^{(2)}$ .

**Solution :**

$$y^{(0)} = (3, 2), \quad t_0 = 0.5 \text{ et } x^{(1)} = (0.5, 2.5),$$

$$y^{(1)} = (-5, 2), \quad t_1 = 0.13 \text{ et } x^{(2)} = (-0.215, 2.435).$$

## 2.4 Méthode de gradient projeté

Pour rechercher la direction admissible en  $x^{(k)}$  l'idée est de projeter le point  $x^{(k+1)}$  sur l'ensemble  $S$ . Si  $S$  est particulier (un parallélépipède, une boule), on peut savoir calculer  $S(x)$  et l'algorithme de gradient projeté est envisageable. Dans le cas général, on ne sait pas calculer  $S(x)$  et il faut recourir à d'autres méthodes.

### 2.4.1 Construction de la méthode (Idée de Rosen)

**Théorème 2.4.1.** (théorème de projection sur un convexe) Soit  $S$  un convexe fermé, et soit  $v \in \mathbb{R}^n$ , alors il existe un élément unique  $P_S(v)$  tel que ,

$$\|v - P_S(v)\| = \min_{u \in S} \|v - u\|. \quad (2.11)$$

On peut définir la suite des approximations  $x^{(k)} \in S$  par,

$$x^{(k+1)} = P_S(x^{(k)} + t_k d^{(k)}) = P_S(x^{(k)} - t_k \nabla^T f(x^{(k)})). \quad (2.12)$$

La projection essaye de garder  $x^{(k+1)}$  dans  $S$ .

- **Déterminer  $d^{(k)}$**

Soit l'ensemble des indices des contraintes actives,

$$I_k = \{i, a_i x^{(k)} = b_i\}.$$

On sait qu'une direction  $d$  est admissible si,

$$A d = 0$$

Alors on définit le sous-espace,

$$P_0 = \{x \in \mathbb{R}^n, a_i^T x^{(k)} = 0\}, \quad i \in I_k$$

Dénotons  $A_k$  la matrice du rang  $|I_k|$  dont les lignes sont les vecteurs  $a_i^T$ .

Le sous espace  $P_v$  orthogonale à  $P_0$  est,

$$P_v = \{x \in \mathbb{R}^n, x = A_k^T v\}.$$

Puisque  $d^{(k)}$  est la projection de  $-\nabla^T f(x^{(k)})$  sur  $P_0$  alors,

$$-\nabla^T f(x^{(k)}) = d^{(k)} + A_k^T v^k$$

$$-A_k \nabla^T f(x^{(k)}) = A_k d^{(k)} + A_k A_k^T v^k$$

Alors,

$$v^k = -A_k (A_k A_k^T)^{-1} \nabla^T f(x^{(k)})$$

Donc,

$$d_k = -\nabla^T f(x^{(k)}) - A_k^T v^k = -[I - A_k (A_k A_k^T)^{-1} A_k] \nabla^T f(x^{(k)}).$$

- **Détermine  $t_k$  (pas optimal)**

Une fois déterminer  $d^{(k)}$  alors  $t_k$  est le pas optimal, c.-à-d.

$$t_k^* = \{t_k, f(x^{(k)} + t_k d^{(k)})\},$$

Tel que,

$$t_k = \min_{i \notin I_k} \left\{ \frac{b_i - a_i x^k}{a_i^T d^{(k)}}, a_i^T d^{(k)} > 0 \right\}.$$

Le critère d'arrêt de la méthode est que  $d^{(k)} = 0$  et  $v_k \geq 0$  (Le critère KKT).

## 2.4.2 Convergence de la méthode

On dit que  $f$  est de gradient supposé Lipchitzien de constante  $L > 0$ , si

$$\forall x, y \in \mathbb{R}^n, \|\nabla f(y) - \nabla f(x)\| \leq L \|y - x\|.$$

**Proposition 2.4.1.** Si  $0 < t_k < \frac{2}{L}$  alors, la suite  $(x_k)_k$  générée par l'algorithme du gradient projeté converge vers un point qui minimise  $f$ .

**Note :** Les méthodes en dessous ont été partagées comme des exposés au étudiants.

## 2.5 Algorithme d'Uzawa :

est une méthode itérative pour résoudre des problèmes d'optimisation avec contraintes non linéaires, en particulier des problèmes d'optimisation avec des contraintes d'inégalité. Il est souvent utilisé pour les problèmes d'optimisation convexes.

## 2.6 Méthode de Newton :

Utilise la matrice Hessienne pour trouver un minimum local en se basant sur une approximation quadratique de la fonction.

## 2.7 Méthode de Gauss-Newton :

Cette méthode est spécialement conçue pour résoudre des problèmes d'optimisation non linéaire liés à des modèles de moindres carrés non linéaires.

## 2.8 Méthode de Quasi-Newton :

Variante de la méthode de Newton qui approxime la Hessienne sans la calculer exactement. Exemple : BFGS.

## 2.9 Méthode de Gradient Conjugué :

Utilise le gradient pour trouver un minimum en exploitant la conjugaison de directions de recherche (fait en cours).

## 2.10 Algorithme de Levenberg-Marquardt :

Utilisé pour résoudre des problèmes d'ajustement de courbe, il combine des éléments de la méthode de Gauss-Newton et de la méthode de gradient stochastique.

## 2.11 Méthode de Marquardt non linéaire :

Une variante de l'algorithme de Levenberg-Marquardt utilisée pour minimiser la somme des carrés des résidus.

## 2.12 Méthode de Descente de Gradient :

Utilise le gradient de la fonction objectif pour trouver un minimum local en suivant la direction la plus raide.

## 2.13 Méthode de Descente de Gradient Stochastique (SGD) :

Utilisée pour l'optimisation de fonctions coût en apprentissage automatique et en apprentissage profond. Elle suit un échantillon aléatoire des données.

## 2.14 Méthode de Recuit Simulé (Simulated Annealing) :

Utilisée pour l'optimisation globale, elle simule le refroidissement d'un système physique pour éviter les minima locaux.

## 2.15 Méthode de l'Évolution Différentielle :

Utilise une population de solutions pour explorer l'espace de recherche et trouver un minimum global.

## 2.16 Méthode de l'Essaim de Particules (Particle Swarm Optimization) :

Inspirée du comportement social des oiseaux ou des poissons, elle guide un ensemble de "particules" vers la recherche d'un optimum.

## 2.17 Algorithme génétique :

Utilisé pour résoudre des problèmes d'optimisation en s'inspirant de la sélection naturelle et de l'évolution.

## 2.18 Méthode de la Fonction de Valeur Quadratique (Quadratic Approximation) :

Utilise une approximation quadratique de la fonction objectif autour d'un point.

## 2.19 Méthode de la Recherche Linéaire :

Trouve le minimum d'une fonction en effectuant une recherche le long d'une ligne dans l'espace des variables.

## 2.20 Méthode de la Région de Confiance (Trust Region) :

Utilise une région de confiance pour gérer la taille des pas lors de l'optimisation.

## 2.21 Méthode de la Pénalité Extérieure (Exterior Penalty Function) :

Utilisée pour résoudre des problèmes d'optimisation avec contraintes en convertissant les contraintes en termes de pénalités dans la fonction objectif.

## 2.22 Méthode de la Barrière (Barrier Method) :

Utilisée pour résoudre des problèmes d'optimisation avec contraintes en ajoutant une fonction de barrière dans la fonction objectif.

## 2.23 Méthode de l'Intérieur (Interior Point Method) :

Une approche pour les problèmes d'optimisation linéaire et non linéaire avec contraintes en utilisant des points intérieurs.

## 2.24 Méthode de Points Intérieurs Primal-Dual :

Cette méthode combine les méthodes duales et primales en utilisant une approche de point intérieur. Elle suit une trajectoire à l'intérieur de la région admissible tout en optimisant les contraintes duales.

## 2.25 Méthode de Pénalisation :

Cette méthode convertit les contraintes non linéaires en termes de pénalités dans la fonction objectif, créant ainsi un problème d'optimisation avec des contraintes pénalisées.

## 2.26 Méthode de Gradient Augmenté :

Cette méthode combine des itérations de gradient avec la résolution du problème dual pour trouver la solution optimale.