

# Chapitre 3

## Recherche linéaire

Dans ce chapitre, nous nous intéressons uniquement au calcul du pas optimal  $t^*$  qui est solution du problème

$$\min_{t>0} \theta(t). \quad (3.1)$$

Avec

- $\theta(t)$ , la fonction test qui représente  $f(x + td)$ , définie pour  $t \geq 0$ ,
- $x$ , le point de départ de la recherche linéaire,
- $d$ , la direction de la recherche linéaire.

Pour la suite,  $d$  est supposée une direction de descente et de ce fait, on aura toujours

$$\theta'(0) < 0.$$

### 3.1 Principe général de résolution

Un algorithme de recherche linéaire résolvant le problème (3.1) génère une suite de valeurs  $\{t^0, t^1, \dots, t^i, \dots\}_{i \in \mathbb{N}}$  est un test à trois sorties qui étant donné un  $t^0 > 0$  (une valeur initial) répond si

- $t^i$  est satisfaisant donc c'est une solution,
- $t^i$  est trop grand,
- $t^i$  est trop petit.

On notera  $t^*$  un pas optimal,  $t_-$  un pas trop petit et  $t_+$  un pas trop grand. Commençons par un exemple simple, mais on verra ensuite qu'il servira de modèle pour des algorithmes plus développés.

**Algorithme 3.1.** Soit  $t^0 > 0$  une valeur initial,  $t_+ = t_- = 0$ ,

- a)  $\theta'(t^i) = 0$  (On peut envisager de stopper la recherche linéaire  $t^i$  à l'air de minimiser  $\theta$ ), donc  $t^* = t^i$ ,
- b)  $\theta'(t^i) > 0$  (alors le minimum de  $\theta$  devrait être plus petit que  $t^i$ ), donc  $t_+ = t^i$ .
- c)  $\theta'(t^i) < 0$  (alors le minimum de  $\theta$  devrait être plus grand que  $t^i$ ), donc  $t_- = t^i$ .

**Remarque 3.1.**

- Pour le choix du pas initial, si  $f$  est de classe  $C^2$ , à hessien définie positive, alors au voisinage de zéro on a

$$\theta(t) = \theta(0) + t\theta'(0) + \frac{1}{2}t^2\theta''(0).$$

Une solution optimale de  $\theta$  au voisinage de zéro implique  $\theta'(t) = 0$ , donc

$$\theta'(t) = \theta'(0) + t\theta''(0) = 0.$$

Et on peut prendre

$$t^1 = -\frac{\theta'(0)}{\theta''(0)} = -\frac{g(x)^\top d}{d^\top H(x)d} > 0. \quad (3.2)$$

- S'il existe un  $N \in \mathbb{N}$  telle que  $t_+^N \neq 0$  alors, il est claire que  $\{t_+^i\}_{i \in \mathbb{N}}$  serait une suite décroissante,  $\{t_-^i\}_{i \in \mathbb{N}}$  est une suite croissante et  $t_-^i < t_+^i, \forall i \geq N$

Dans ces conditions et pour que l'algorithme de recherche linéaire s'arrête après un nombre fini d'itérations, le test doit satisfaire les propriétés suivantes.

### 3.1.1 Propriétés

1) Les ensembles suivants constituent une partition de  $\mathbb{R}_+$

$$\{t \in \mathbb{R}_+ : \theta'(t) = 0\}, \{t \in \mathbb{R}_+ : \theta'(t) > 0\}, \{t \in \mathbb{R}_+ : \theta'(t) < 0\}.$$

- 2)  $\exists N \in \mathbb{N}$  telle que  $t_+^N \neq 0$  ( $\theta'(t_+^N) > 0$ ), sinon  $t_+^i \rightarrow +\infty$ .
- 3) Tout intervalle  $[t_-^i, t_+^i]$  contient un  $t^*$  tel que  $\theta'(t^*) = 0$  pour éviter le cas :  $|t_-^i - t_+^i| \rightarrow 0$ .

### 3.1.2 Remarques sur le calcul de $t^{i+1}$

1) La recherche linéaire est caractérisé par le choix de  $t^{i+1}$ , pour cela considérons les deux cas suivants :

- Cas  $t_+^{i+1} = 0$  (extrapolation)
  - On peut prendre  $t^{i+1} = \alpha t_-^{i+1}$  avec  $\alpha$  fixe par exemple  $\alpha = 10$
- Cas  $t_+^{i+1} \neq 0$  (interpolation)
  - Donc,  $t^{i+1} \in ]t_-^{i+1}, t_+^{i+1}[$ , on peut procéder à une simple interpolation par dichotomie, c'est à dire

$$t^{i+1} = \frac{t_-^{i+1} + t_+^{i+1}}{2}$$

Ce qui nous donne un intervalle de confiance (où d'incertitude) divisé par deux pour chaque itération.

2) De coutume, il est préférable d'orienter le choix de  $t^{i+1}$ . Pour cela l'idée générale est de construire un polynôme d'interpolation à partir des points déjà testés puis minimiser ce dernier.

**Théorème 3.1.** *La recherche linéaire définie par l'algorithme 3.1 et satisfaisant les propriétés 3.1.1 se termine après un nombre fini d'itérations.*

**Démonstration.** En utilisant le fait que si la recherche linéaire est infinie,  $|t_-^i - t_+^i|$  tendrait vers 0 et on aurait une contradiction avec la propriété 3) de 3.1.1.  $\square$

**Remarque 3.2.** *Pour assurer la convergence de l'algorithme 3.1 dans le cas par exemple d'une interpolation par un polynôme, une fois calculé, le pas  $t^{i+1}$  doit être ramené à l'intérieur de  $[t_-^{i+1}, t_+^{i+1}]$  en remplaçant successivement  $t^{i+1}$  par*

$$\min \left\{ t^{i+1}, t_+^{i+1} - \alpha(t_+^{i+1} - t_-^{i+1}) \right\},$$

puis par

$$\max \left\{ t_+^{i+1}, t_-^{i+1} + \alpha(t_+^{i+1} - t_-^{i+1}) \right\},$$

où  $\alpha = \frac{1}{10}$  (par exemple).

La question du choix de  $t^{i+1}$  étant réglé, nous allons se pencher sur les tests a), b) et c) de l'algorithme 3.1.

## 3.2 Recherche linéaire exacte

Historiquement (Cauchy 1847 [6]), On cherché une valeur  $t^*$  tel que

$$\theta(t^*) < \theta(0) \text{ et } \theta'(t^*) = 0. \quad (3.3)$$

Sous ces conditions, on peut dresser l'algorithme suivant

### Algorithme 3.2.

#### Etape 0 (Initialisation) :

$i = 0, t^0 > 0, t_-^0 = 0, t_+^0 = 0, \varepsilon > 0$  (suffisamment petit),

#### Etape 1 :

si  $|\theta'(t^i)| \leq \varepsilon$ , stop  $t^* = t^i$

si  $\theta(t^i) \geq \theta(0)$  où  $\theta'(t^i) > 0$  alors,

$$t_+^{i+1} = t^i, t_-^{i+1} = t_+^i, \text{ on va à l'étape 2}$$

si  $\theta(t^i) < \theta(0)$  et  $\theta'(t^i) < -\varepsilon$  alors,

$$t_-^{i+1} = t^i, t_+^{i+1} = t_+^i, \text{ on va à l'étape 2}$$

#### Etape 2 :

si  $t_+^{i+1} = 0$  déterminer  $t^{i+1} \in ]t_-^{i+1}, +\infty[$

si  $t_+^{i+1} \neq 0$  déterminer  $t^{i+1} \in ]t_-^{i+1}, t_+^{i+1}[$

remplacer  $i$  par  $i + 1$  et aller à l'étape 1

On peut noter que ce test est plus raffiné que l'algorithme 3.2 (ce dernier n'assurait même pas la décroissance de  $\theta$ ).

**Théorème 3.2.** Supposant que  $\theta$  est de classe  $C^1$ ,  $\theta'(0) < 0$  et  $\exists t^* \in \mathbb{R}_+$  tel que,  $\theta(t) \geq \theta(t^*)$ ,  $\forall t \in \mathbb{R}_+$  alors l'algorithme 3.1 est finie.

*Démonstration.*

Puisque  $\exists t^* \in \mathbb{R}_+$  tel que  $\theta(t) \geq \theta(t^*)$ ,  $\forall t \in \mathbb{R}_+$  alors  $\exists N \in \mathbb{N}^*$ ,  $t_+^N \neq 0$  donc  $\{t_-^i\}$  et  $\{t_+^i\}$  sont des suites adjacentes avec

$$\lim t_+^i = \lim t_-^i = t^*$$

$\theta$  est continue et  $\theta(t_+^i) < \theta(0)$  donc,

$$\theta(t_-^i) \rightarrow \theta(t^*) \leq \theta(0)$$

la même chose pour  $\theta'$  donne

$$\theta'(t^*) \leq -\varepsilon \leq -\frac{\varepsilon}{2} < 0.$$

Le développement de Taylor pour  $t_+^i - t^*$  petit :

$$\theta(t_+^i) = \theta(t^*) + \theta'(t^*)(t_+^i - t^*) \leq \theta(t^*) - (\varepsilon/2)(t_+^i - t^*) < \theta(t^*) < \theta(0)$$

et pour que b) soit vérifié il faut que  $\theta'(t_+^i) > \varepsilon$ , donc

$$\begin{cases} \theta'(t^*) \leq -\varepsilon \\ \theta'(t^*) \geq \varepsilon \end{cases}$$

d'où une contradiction avec la continuité de  $\theta'$ .  $\square$

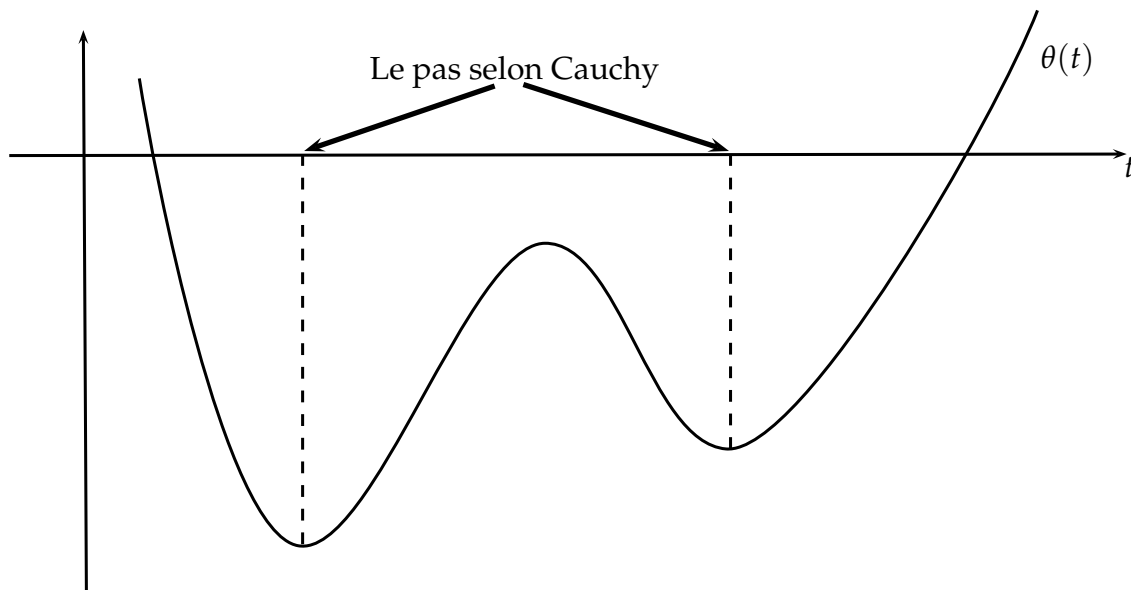


Figure 3.1: Recherche d'un pas optimal.

### 3.3 Recherche linéaire inexacte

Jusqu'à maintenant on s'efforce d'identifier un minimum local de  $\theta(t)$  pour  $t > 0$ . Cela conduit à un long algorithme, actuellement il est reconnu que ceci est pratiquement inutile. On préfère utiliser une recherche linéaire dite inexacte ou économique, où l'objectif est de

- 1) Faire décroître  $\theta$  suffisamment.
- 2) Empêcher le pas optimal  $t^*$  d'être trop petit.

Le premier objectif se traduit le plus souvent en forçant la réalisation d'une inégalité de la forme

$$\theta(t^*) \leq \theta(0) + \nu,$$

où  $\nu$  est un terme négatif.

Le second objectif est motivé par le fait que la décroissance est assurée pour des valeurs proche de zéro (vue la continuité de  $\theta'$ ).

### 3.3.1 La recherche linéaire de Goldstein

Cette méthode est commode lorsqu'on veut pas calculer  $\theta'$ . On choisit deux coefficients  $0 < m_1 < m_2 < 1$ , le pas  $t^*$  "optimal" doit satisfaire aux conditions

$$\begin{cases} \theta(t^*) \leq m_1\theta'(0)t^* + \theta(0), \\ \theta(t^*) \geq m_2\theta'(0)t^* + \theta(0), \end{cases} \quad (3.4)$$

dites conditions de Goldstein [14].

#### Algorithme 3.3.

**Etape 0 (Initialisation) :**  $i = 0, t^0 > 0, t_-^0 = 0, t_+^0 = 0, 0 < m_1 < m_2 < 1$

**Etape 1 :** si  $m_2\theta'(0)t^i + \theta(0) \leq \theta(t^i) \leq m_1\theta'(0)t^i + \theta(0)$ , stop.  $t^* = t^i$

si  $\theta(t^i) > m_1\theta'(0)t^i + \theta(0)$ , alors

$$t_+^{i+1} = t^i, t_-^{i+1} = t_-^i, \text{ on va à l'étape 2}$$

si  $\theta(t^i) < m_2\theta'(0)t^i + \theta(0)$ , alors

$$t_-^{i+1} = t^i, t_+^{i+1} = t_+^i, \text{ on va à l'étape 2}$$

**Etape 2 :** si  $t_+^{i+1} = 0$  déterminer  $t^{i+1} \in ]t_-^{i+1}, +\infty[$

si  $t_+^{i+1} \neq 0$  déterminer  $t^{i+1} \in ]t_-^{i+1}, t_+^{i+1}[$

remplacer  $i$  par  $i + 1$  et aller à l'étape 1.

**Théorème 3.3.** Supposons que  $\theta$  est de classe  $C^1, \theta'(0) < 0$  et  $\exists t^* \in \mathbb{R}_+$  tel que,  $\theta(t) \geq \theta(t^*), \forall t \in \mathbb{R}_+$  alors la méthode de Goldstein est finie.

#### Démonstration.

Puisque  $\exists t^* \in \mathbb{R}_+$  tel que  $\theta(t) \geq \theta(t^*), \forall t \in \mathbb{R}_+$  alors  $\exists N \in \mathbb{N}, t_+^N \neq 0$  et de la remarque 3.1,  $\{t_-^i\}$  et  $\{t_+^i\}$  sont des suites adjacentes avec  $\lim t_+^i = \lim t_-^i = t^*, \theta$  est continue et  $\theta(t_+^i) < \theta(0)$  donc, de b) on a

$$\theta(t_+^i) > m_1 t_+^i \theta'(0) + \theta(0) \Rightarrow \theta(t^*) \geq m_1 t^* \theta'(0) + \theta(0),$$

$$\theta(t_-^i) < m_2 t_-^i \theta'(0) + \theta(0) \Rightarrow \theta(t^*) \leq m_2 t^* \theta'(0) + \theta(0).$$

Ce qui donne  $m_2 \leq m_1$  et puisque  $\theta'(0) < 0 \Rightarrow$  une contradiction avec les hypothèses.

□

Dans ce test on demande à  $t^*$  soit de façon à ce que

$$\theta_1(t^*) \leq \theta(t^*) \leq \theta_2(t^*), t > 0$$

où

$$\theta_1(t) = m_2\theta'(0)t + \theta(0)$$

$$\theta_2(t) = m_1\theta'(0)t + \theta(0)$$

Ce qui est schématisé par la figure 3.2

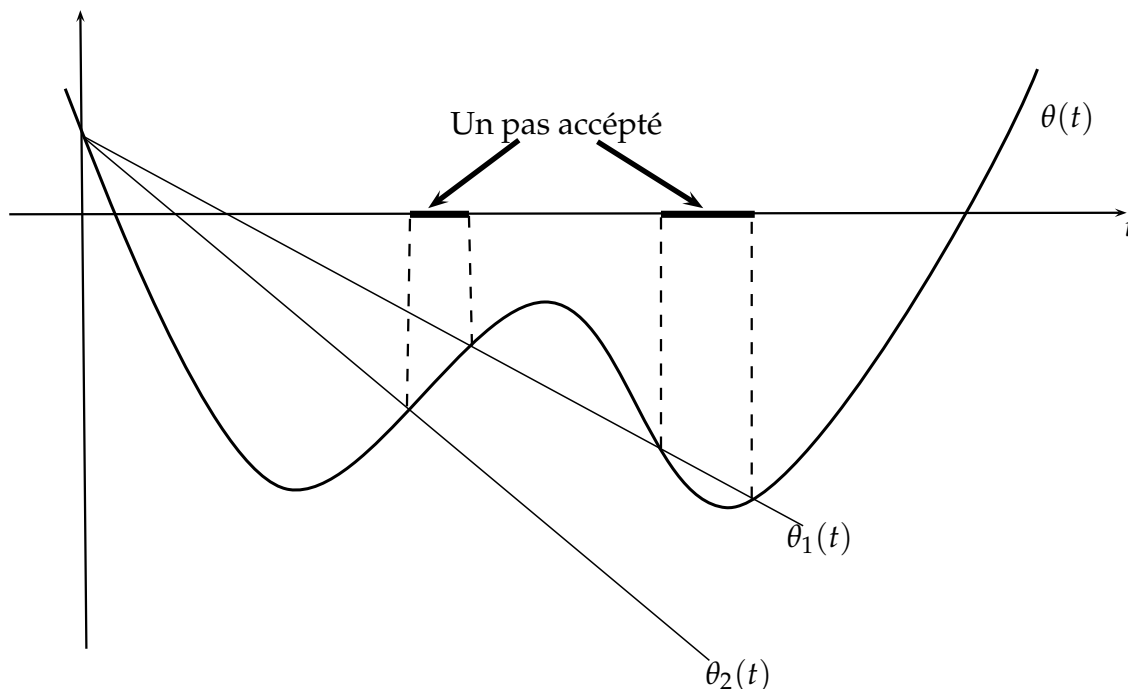


Figure 3.2: La sélection du pas par la règle de Goldstein.

### 3.3.2 La recherche linéaire d'Armijo

Cette méthode n'exige pas le calcul de  $\theta'$ . On choisit  $m \in ]0, 1[$ , et on cherche à réaliser une condition dite d'Armijo (ou condition de décroissance linéaire [1]), qui est

$$\theta(t^*) \leq m\theta'(0)t^* + \theta(0). \quad (3.5)$$



**Algorithme 3.4.****Etape 0 (Initialisation) :**  $i = 0, t^0 > 0, 0 < m < 1$ **Etape 1 :** si  $\theta(t^i) \leq m\theta'(0)t^i + \theta(0)$ , stop  $t^* = t^i$ si  $\theta(t^i) > m\theta'(0)t^i + \theta(0)$ , alors $t_+^{i+1} = t^i$ , on va à l'étape 2**Etape 2 :** Déterminer  $t^{i+1} \in ]0, t_+^{i+1}[$ remplacer  $i$  par  $i + 1$  et aller à l'étape 1.**Théorème 3.4.** Supposons que  $\theta$  est de classe  $C^1, \theta'(0) < 0$  et  $\exists t^* \in \mathbb{R}_+$ , tel que  $\theta(t) \geq \theta(t^*), \forall t \in \mathbb{R}_+$  alors la méthode d'Armijo est finie.**Démonstration.**

Pareille à celle du théorème 3.4

□

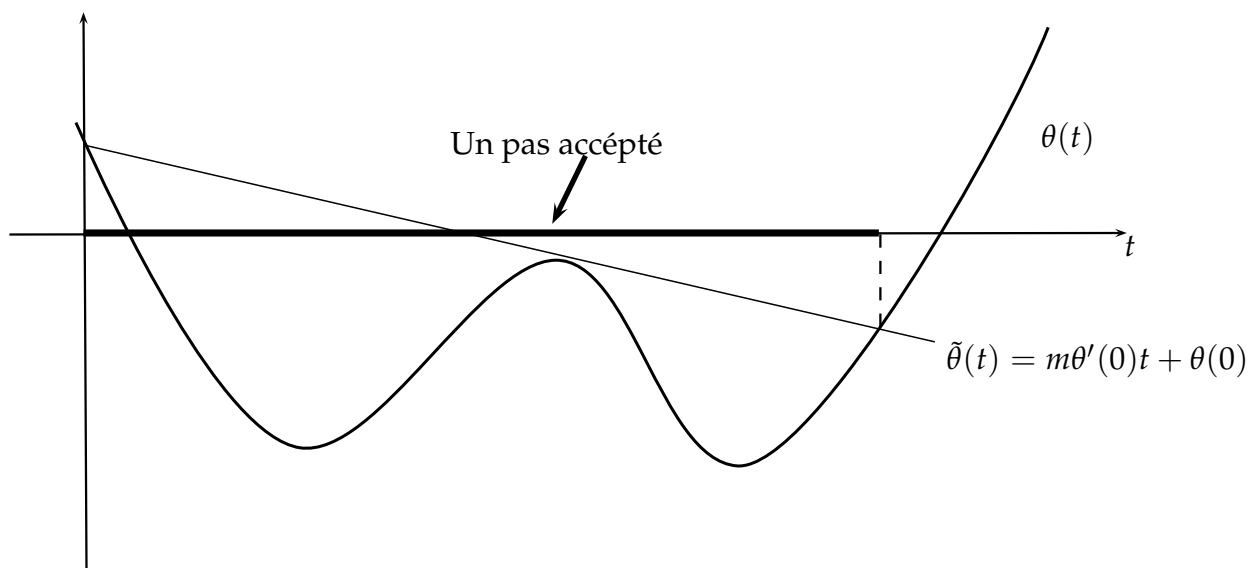
**Remarque 3.3.** L'astuce de cette méthode est de prendre un  $t^0$  raisonnablement grand, par exemple  $t^0 = 1$ , pour le nouveau pas  $t^{i+1}$  on utilise souvent la technique de rebroussement "backtraking" qui consiste à prendre  $t^i = \alpha^i$ , où  $\alpha \in ]0, 1[$  dans ce cas le pas choisit est :  $t^* = \alpha^i$  et  $i$  est le plus petit entier tel que l'on ait (3.5). Ce qui nous donne une suite  $\{1, \alpha^1, \alpha^2, \dots\}$  décroissante d'où le nom de rebroussement.

Figure 3.3: La sélection du pas selon la règle d'Armijo.

### 3.3.3 Règle de Wolfe

C'est la méthode qui semble actuellement la plus efficace, on choisit  $0 < m_1 < m_2 < 1$  et le test est une double inégalité, appelées conditions de Wolfe [40]

$$\theta(t^*) \leq m_1 t^* \theta'(0) + \theta(0), \quad (3.6)$$

$$\theta'(t^*) \geq m_2 \theta'(0). \quad (3.7)$$

#### Algorithme 3.5.

**Etape 0 (Initialisation) :**  $i = 0, t^0 > 0, 0 < m_1 < m_2 < 1$

**Etape 1 :** si  $\theta(t^i) \leq m_1 \theta'(0)t^i + \theta(0)$  et  $\theta'(t^i) \geq m_2 \theta'(0)$ , stop  $t^* = t^i$

si  $\theta(t^i) > m_1 \theta'(0)t^i + \theta(0)$ , alors

$$t_+^{i+1} = t^i, t_-^{i+1} = t^i \text{ on va à l'étape 2}$$

si  $\theta(t^i) \leq m_1 \theta'(0)t^i + \theta(0)$  et  $\theta'(t^i) < m_2 \theta'(0)$ , alors

$$t_-^{i+1} = t^i, t_+^{i+1} = t_+^i \text{ on va à l'étape 2}$$

**Etape 2 :** si  $t_+^{i+1} = 0$  déterminer  $t^{i+1} \in ]t_-^{i+1}, +\infty[$

si  $t_+^{i+1} \neq 0$  déterminer  $t^{i+1} \in ]t_-^{i+1}, t_+^{i+1}[$

remplacer  $i$  par  $i + 1$  et aller à l'étape 1.

**Remarque 3.4.** Cette méthode respecte bien le principe d'une recherche linéaire inexacte. D'une part, on assure la décroissance de  $\theta$  puisque

$$\frac{\theta(t^*) - \theta(0)}{t^*} \leq m_1 \theta'(0) < 0 \Rightarrow \theta(t^*) < \theta(0).$$

Et d'autre part du fait que  $\theta'(t^*) \geq m_2 \theta'(0)$ , la dérivée a suffisamment augmenté (donc  $x_{k+1}$  ne sera pas trop proche de  $x_k$ ).

**Théorème 3.5.** Supposant que :  $\theta$  est de classe  $C^1$ , bornée inférieurement et  $\theta'(0) < 0$ . Alors la recherche linéaire de Wolfe est finie.

*Démonstration.*

$\exists N \in \mathbb{N}$  tel que  $t_+^i \neq 0$  sinon  $t^i \rightarrow +\infty$ , avec

$$\theta(t^i) \leq \theta(0) + m_1 t^i \theta'(0).$$

Donc  $\theta(t^i)$  tendrait vers  $-\infty$  d'où une contradiction avec les hypothèses

Les suites  $\{t_-^i\}$  et  $\{t_+^i\}$  sont des suites adjacentes avec

$$\lim t_+^i = \lim t_-^i = t^*$$

$$\theta(t_+^i) > m_1 t_+^i \theta'(0) + \theta(0) \Rightarrow \theta(t^*) \geq m_1 t^* \theta'(0) + \theta(0)$$

$$\theta(t_-^i) \leq m_1 t_+^i \theta'(0) + \theta(0) \Rightarrow \theta(t^*) \leq m_1 t^* \theta'(0) + \theta(0)$$

Ce qui donne

$$\theta(t^*) = m_1 t^* \theta'(0) + \theta(0)$$

Et puisque  $\theta(t_-^i) < m_2 \theta'(0)$  alors

$$\theta(t^*) \leq m_2 \theta'(0)(t)$$

D'autre part pour tout  $t_+^i$

$$\begin{aligned} \theta(t_+^i) &> \theta(0) + m_1 \theta'(0)(t^* + t_+^i - t^*) = \\ &= \theta(0) + m_1 \theta'(0)(t_+^i - t^*) + m_1 \theta'(0)t^*. \end{aligned}$$

Donc

$$\theta(t_+^i) > \theta(t^*) + m_1 \theta'(0)(t_+^i - t^*),$$

ce qui entraîne  $[\theta(t_+^i) - \theta(t^*)]/(t_+^i - t^*) \geq m_1 \theta'(0)$ . Passant à la limite, on obtient

$$\theta'(t^*) \geq m_1 \theta'(0) > m_2 \theta'(0).$$

Ce qui contredit les hypothèses sur  $m_1$  et  $m_2$ .  $\square$

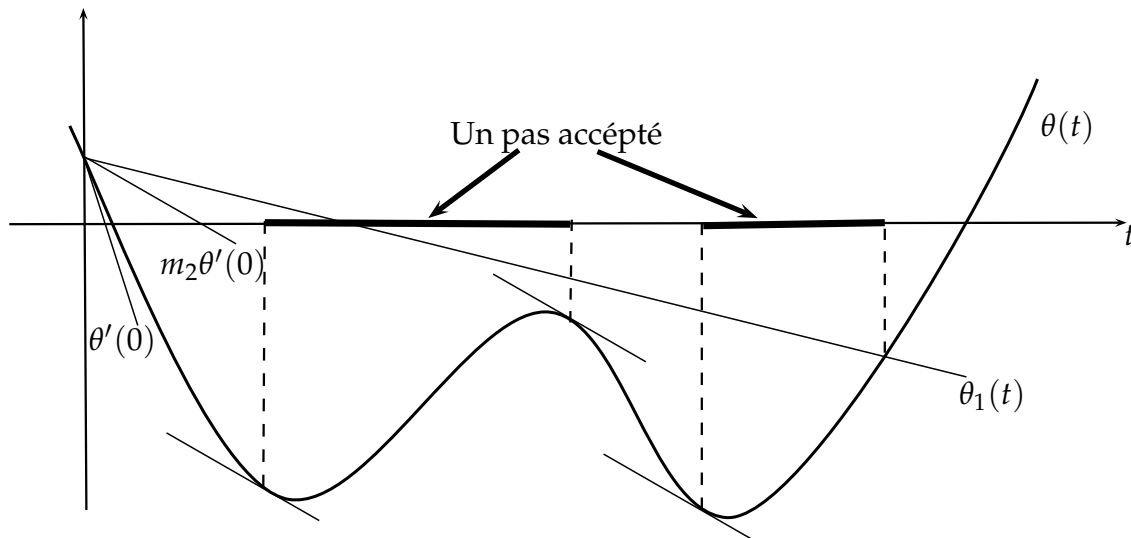


Figure 3.4: La sélection du pas selon la règle de Wolfe.

**Remarque 3.5.** Pour démontrer la convergence de certains algorithmes il est parfois nécessaire d'exiger des conditions plus strictes que (3.7), dans la règle de Wolfe forte [5, 27], (3.7) est remplacé par

$$|\theta'(t^*)| \leq m_2 |\theta'(0)|. \quad (3.8)$$

Le théorème suivant, montre qu'il convient de choisir le coefficient de descente ( $m$  pour Armijo,  $m_1$  pour Goldstein et Wolfe) inférieur à  $\frac{1}{2}$ , et dans la règle de Goldstein il faudrait mieux choisir  $m_2 > \frac{1}{2}$ .

**Théorème 3.6.** Supposons  $\theta$  quadratique avec un minimum  $t^*$ . Alors

$$\theta(t^*) = \theta(0) + \frac{1}{2}\theta'(0)t^*.$$

C'est-à-dire si  $\theta$  est quadratique, le pas optimale  $t^* = 1$  sera refusé par le test de descente si  $m_1 > \frac{1}{2}$ . On doit donc prendre

$$0 < m_1 < \frac{1}{2} < m_2 < 1. \quad (3.9)$$

**Remarque 3.6.**

- Nous avons définie plusieurs méthodes de recherche linéaire, ce qui permet de calculer  $x_{k+1}$  à partir de  $x_k$  suivant le schéma itératif déjà défini, donc

$$x_{k+1} = x_k + t_k d_k,$$

où  $d_k$  est une direction et  $t_k$  un pas choisit par une recherche linéaire. On peut se demander si la suite ainsi générée converge vers une solution du problème (2.1). Bien entendu que cette propriété ne peut pas avoir indépendamment du choix de  $d_k$ , la recherche linéaire sera impuissante si  $d_k$  est "trop orthogonale" à  $g(x_k)$ . Ainsi on défini l'angle  $\varphi_k$  entre  $-g(x_k)$  et  $d_k$  par

$$c_k = \cos(\varphi_k) = -\frac{g(x_k)^\top d_k}{\|g(x_k)\| \|d_k\|} \quad (3.10)$$

et d'après le théorème 2.1  $d_k$  serait une direction de descente si et seulement si  $c_k > 0$ .

- Dans le chapitre 1, on a vu un théorème essentiel de convergence (théorème 1.1), ce théorème couvre un domaine très large d'algorithmes. Sachant que notre but est l'étude des méthodes utilisant essentiellement des directions de descentes, avec des recherches linéaires qui sont dans la plus part du temps inexactes, on utilise souvent le théorème suivant.

**Théorème 3.7** (Zoutendijk). Supposons que  $g$  (le gradient de  $f$  est Lipchitzien sur la tranche  $\mathcal{L} = \{x \in \mathbb{R}^n : f(x) \leq f(x_0)\}$ ). Soit un algorithme dans lequel la direction est telle que  $\sum c_k^2$  diverge, et en utilisant la règle de Wolfe. Alors

- Soit  $\theta(t_k) \rightarrow -\infty$ .
- Soit  $\liminf \|g(x_k)\| = 0$ .

**Démonstration.**

Le premier cas est vrai si  $f$  n'est pas bornée inférieurement donc  $t_k \rightarrow +\infty$  et  $\theta(t_k) \rightarrow -\infty$ . Sinon la règle de Wolfe se termine après un nombre finie d'itération (voir théorème 3.5). Donc on a  $f(x_k) < f(x_0)$  et  $g(x_k)$  est lipschitzien dans  $\mathcal{L}$ ,  $\forall i \in \mathbb{N}^*$  et soit  $t_k$  le pas optimal donné par Wolfe à l'itération  $k$ . De (3.5)

$$t_k \theta'(0) = t_k g(x_k)^\top d_k = -t_k c_k \|d_k\| \|g(x_k)\| = -c_k \|x_{k+1} - x_k\| \|g(x_k)\|.$$

Et le test de descente entraîne directement

$$\theta(t_k) \leq \theta(0) + m_1 t_k \theta'(0) \Leftrightarrow f(x_{k+1}) \leq f(x_k) - m_1 c_k \|x_{k+1} - x_k\| \|g(x_k)\|.$$

Donc

$$f(x_k) - f(x_{k+1}) \geq m_1 c_k \|x_{k+1} - x_k\| \|g(x_k)\|. \quad (3.9)$$

La règle de Wolfe, donne

$$\theta'(t_k) \geq m_2 \theta'(0),$$

donc

$$\begin{aligned} g(x_{k+1})^\top d_k &\geq m_2 g_k^\top d_k \Rightarrow \\ (g(x_{k+1}) - g(x_k))^\top d_k &\geq (1 - m_2) g(x_k)^\top d_k = (1 - m_2) c_k \|g(x_k)\| \|d_k\|. \end{aligned}$$

La formule de *Cauchy-Schwarz* et le fait que  $g(x_k)$  est lipchitzien sur  $\mathcal{L}$  entraîne l'existence d'une constante  $L$  (dépendant de  $x_0$ ) telle que

$$(g(x_{k+1}) - g(x_k))^\top d_k \leq \|g(x_{k+1}) - g(x_k)\| \|d_k\| \leq L \|d_k\|^2,$$

donc

$$(1 - m_2) c_k \|g(x_k)\| \leq L \|d_k\|. \quad (3.10)$$

On déduit des deux arguments (3.1) et (3.2) que (avec  $r = \frac{m_1(1 - m_2)}{L}$ )

$$r c_k^2 \|g(x_k)\|^2 \leq f(x_k) - f(x_{k+1})$$

donc

$$r \sum_{i=1}^{i=n} c_i^2 \|g_i\|^2 \leq f(x_0) - f(x_{k+1}).$$

Si  $f$  est bornée inférieurement (et puisque  $\{f(x_k)\} \searrow$ ), alors

$$\sum c_i^2 \|g_i\|^2 < \infty.$$

Et si de plus on suppose que  $\|g(x_k)\| \geq c > 0, \forall i \in \mathbb{N}$  implique

$$\sum c_i^2 < \infty,$$

contrairement aux hypothèses.  $\square$

**Remarque 3.7.** .

- On démontre dans [27] que le théorème 3.7 reste valable si on utilise une recherche linéaire vérifiant ce qu'on appelle condition de Zoutendijk : Il existe une constante  $M > 0$  telle que pour tout  $k \geq 1$  on ait

$$f(x_{k+1}) \leq f(x_k) - M \|g(x_k)\|^2 c_k. \quad (3.11)$$

- On note que les recherches linéaires exactes ainsi que la recherche linéaire du type Armijo et Goldstein vérifie la condition (3.11).

**Exemple 3.1.** Nous allons tester l'algorithme "steepest descent" avec différentes recherches linéaires, comme fonction test on prend : la vallée - banane - de Rosenbrock. Il s'agit d'une fonction à deux variables du quatrième degré qui se présente comme une vallée en U dont le fond assez plat est incurvé suivant une parabole.

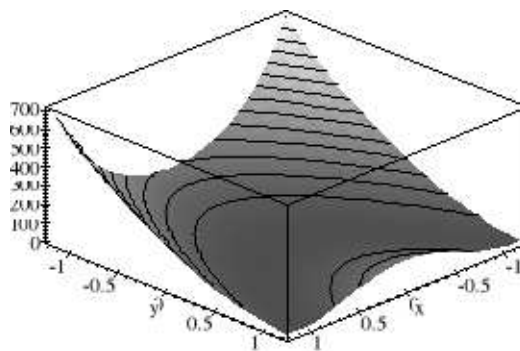


Figure 3.5: La fonction de Rosenbrock ( $f(x, y) = 100 \cdot (x^2 - y)^2 + (x - 1)^2$ )

La figure 3.6 situe le point solution  $(x, y) = (1, 1)$  du problème de minimisation de la fonction

de Rosenbrock, le point de départ qui sera systématiquement utilisé  $(x, y) = (-1.2, 1)$ , ainsi que le acheminement typique de l'algorithme Steepest descent avec une recherche linéaire.

Les résultats obtenus sont illustrés dans les tableaux 3.1, 3.2, 3.3 et 3.4.

$k$	$t_k$	$x_k$	$f(x_k)$	$\ g(x_k)\ $
0	—	(-1.20, 1.000)	24.200	232.867
10	$4.777 \times 10^{-03}$	(-1.009, 1.019)	4.037	2.928
100	$3.811 \times 10^{-03}$	(-0.771, 0.596)	3.138	3.092
200	$8.999 \times 10^{-03}$	(0.147, 0.018)	0.727	1.642
500	$1.634 \times 10^{-03}$	(0.807, 0.650)	$3.726 \times 10^{-02}$	0.220
1000	$1.243 \times 10^{-03}$	(0.944, 0.891)	$3.095 \times 10^{-03}$	$5.547 \times 10^{-02}$
1500	$1.161 \times 10^{-03}$	(0.981, 0.962)	$3.590 \times 10^{-04}$	$1.826 \times 10^{-02}$
2000	$1.136 \times 10^{-03}$	(0.993, 0.986)	$4.538 \times 10^{-05}$	$6.4229 \times 10^{-03}$

Tableau 3.1: Steepest descent avec recherche linéaire exacte (dichotomie).

$k$	$t_k$	$x_k$	$f(x_k)$	$\ g(x_k)\ $
0	—	(-1.200, 1.000)	24.200	232.867
10	$2.499 \times 10^{-03}$	(-1.016, 1.035)	4.067	3.239
100	$4.999 \times 10^{-03}$	(0.360, 0.123)	0.413	1.399
200	$2.499 \times 10^{-03}$	(0.683, 0.466)	0.100	0.409
500	$2.499 \times 10^{-03}$	(0.930, 0.865)	$4.841 \times 10^{-03}$	0.124
1000	$1.250 \times 10^{-03}$	(0.976, 0.953)	$5.501 \times 10^{-04}$	$2.304 \times 10^{-02}$
1500	$2.499 \times 10^{-03}$	(0.988, 0.977)	$1.225 \times 10^{-04}$	$1.390 \times 10^{-02}$
2000	$4.999 \times 10^{-03}$	(0.994, 0.989)	$2.838 \times 10^{-05}$	$1.037 \times 10^{-02}$

Tableau 3.2: Steepest descent avec recherche linéaire d'Armijo.

$k$	$t_k$	$x_k$	$f(x_k)$	$\ g(x_k)\ $
0	—	(-1.200, 1.000)	24.200	232.867
10	$9.024 \times 10^{-03}$	(-0.939, 0.878)	3.765	5.702
100	$8.573 \times 10^{-04}$	(-0.483, 0.248)	2.221	2.956
200	$8.573 \times 10^{-04}$	(0.386, 0.149)	0.376	1.221
500	$8.573 \times 10^{-04}$	(0.737, 0.543)	$6.894 \times 10^{-02}$	0.428
1000	$9.024 \times 10^{-03}$	(0.906, 0.821)	$8.749 \times 10^{-03}$	0.142
1500	$8.573 \times 10^{-04}$	(0.989, 0.978)	$1.117 \times 10^{-04}$	$1.356 \times 10^{-02}$
2000	$8.573 \times 10^{-04}$	(0.998, 0.996)	$2.716 \times 10^{-06}$	$1.478 \times 10^{-03}$

Tableau 3.3: Steepest descent avec recherche linéaire de Goldstein.



$k$	$t_k$	$x_k$	$f(x_k)$	$\ g(x_k)\ $
0	—	(-1.200, 1.000)	24.200	232.867
10	$2.499 \times 10^{-03}$	(-1.016, 1.03)	4.067	3.239
100	$4.999 \times 10^{-03}$	(0.593, 0.347)	0.166	0.762
200	$2.499 \times 10^{-03}$	(0.716, 0.511)	$8.059 \times 10^{-02}$	0.455
500	$2.000 \times 10^{-02}$	(0.927, 0.859)	$5.380 \times 10^{-03}$	0.332
1000	$9.999 \times 10^{-03}$	(0.975, 0.951)	$5.919 \times 10^{-04}$	$8.178 \times 10^{-02}$
1500	$9.999 \times 10^{-03}$	(0.988, 0.977)	$1.279 \times 10^{-04}$	$1.259 \times 10^{-02}$
2000	$1.250 \times 10^{-03}$	(0.994, 0.989)	$2.887 \times 10^{-05}$	$6.108 \times 10^{-03}$

Tableau 3.4: Steepest descent avec recherche linéaire de Wolfe.

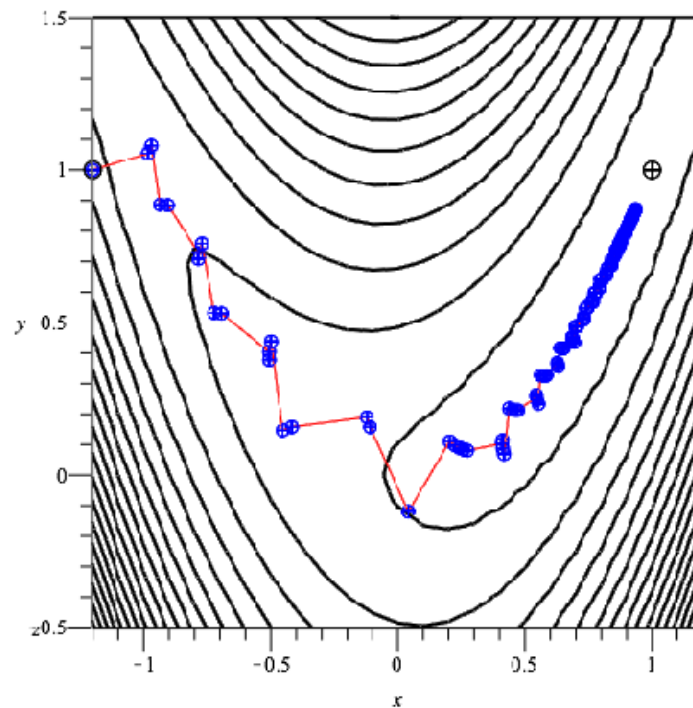


Figure 3.6: Acheminement typique de l'algorithme Steepest descent avec une recherche linéaire.

## Exercices

**Exercice 3.1.** Montrer que si  $\theta$  est quadratique avec un minimum  $t^*$ . Alors

$$\theta(t^*) = \theta(0) + \frac{1}{2}\theta'(0)t^*.$$

**Exercice 3.2.** Soit  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ , une fonction différentiable. On suppose que  $x^* \in \mathbb{R}^2$  est un minimum local le long de n'importe quelle ligne qui passe par  $x^*$ , autrement,  $\alpha = 0$  est un minimum local de la

fonction définie par  $\theta(\alpha) = f(x^* + \alpha d)$ ,  $\forall d \in \mathbb{R}^n$ .

1) Montrez que  $g(x^*) = 0$ .

2) Est-ce que  $x^*$  est un minimum local pour  $f$  ?

**Exercice 3.3.** Soit le problème de minimisation de la fonction  $f : \mathbb{R} \rightarrow \mathbb{R}$  donnée par

$$f(x) = x^2 - \frac{x^3}{3}.$$

Supposons qu'on utilise un l'algorithme

$$x_{k+1} = x_k - \alpha f'(x_k),$$

avec un pas  $\alpha = 1/2$  et une condition initiale  $x_1 = 1$ .

1) Montrer que cet algorithme converge vers un minimum local de  $f$ .

2) Trouver la vitesse de convergence de cet algorithme.

**Exercice 3.4.** Nous voulons minimiser la fonction définie par

$$f(x, y) = 10x^2 + 5xy + 10(y - 3)^2.$$

Trouver le minimum de cette fonction en utilisant la méthode de la plus profonde pente (steepest descent), avec une recherche linéaire de Cauchy et partant du point initial  $(x_0, y_0) = (10, 15)$  (prenez une tolérance de l'ordre de  $10^{-3}$ ).

**Exercice 3.5.** Pour la fonction définie par

$$f(x, y) = x^2 + 2y^2 + 4x + 4y.$$

1) Montrer que la méthode "steepest descent" appliquée avec le point initial  $x_0, y_0) = (0, 0)$  génère une suite  $\{x_k\}$  telle que

$$x_{k+1} = \left( \frac{2}{3^k} - 2, \left( -\frac{1}{3} \right)^k - 1 \right)^T.$$

2) Déduisez le minimum de  $f$ .

**Exercice 3.6.** Montrer que si  $f$  est une fonction quadratique convexe,  $f(x) = \frac{1}{2}x^\top Qx + b^\top x + c$ .

Alors la solution du problème de recherche linéaire

$$\min_{t>0} f(x_k + td_k),$$

peut être calculée de manière analytique et est donnée par

$$t_k = -\frac{g(x_k)^\top d_k}{d_k^\top Q d_k}.$$