Chapitre 2

Théorèmes de convergence associés aux méthodes à directions de descente et recherches linéaires inexactes

2.1 Introduction

Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$.On concidère le probléme de minimisation sans contraintes (P) suivant :

$$\min \{f(x) \; : x \in \mathbb{R}^n\} \tag{P}$$

Rappelons que pour résoudre le probléme (P) , les méthodes à directions de descente et rechreches linéaires inexactes génèrent une suite $\{x_k\}_{k\in\mathbb{N}}$de la manière suivante :

Supposons qu'à l'itération k on ait un point $x_k \in \mathbb{R}^n$ et une direction de descente $d_k \in \mathbb{R}^n$,c'est à dire d_k vérifie :

$$\nabla f(x_k)^\top d_k < 0 \tag{2.1}$$

Le sucusseur x_{k+1} de x_k est obtenu comme suit :

$$x_{k+1} = x_k + \alpha_k d_k \,;\, \alpha_k \in \mathbb{R}_+^* \tag{2.2}$$

α_k est obtenu par une recherche linéaire inexacte du type Armijo ou de Goldstein ou de Wolfe (wolfe-forte) .

2.1.1 Algorithme général des directions de descente

L'algorithme général associé aux rechreches linéaires inexactes et aux directions de descente est le suivant :

Algorithme des directions de descente

Etape1.

Prendre $x_0 \in \mathbb{R}^n$ quelconque, $\varepsilon = 10^{-6}$(par exemple),$k = 0$.

Etape2.

Si $\|\nabla f(x_k)\| < \varepsilon$,stop ; sinon, trouvez une direction de descente $d_k \in \mathbb{R}^n$ vérifiant $\nabla f(x_k)^\top d_k < 0$.

Etape3.

Calculez le pas α_k en utilisant l'une des règles a) ou b) ou c) suivantes :

a) Armijo

b) (Goldstein1) et (Goldstein2)

c) (Wolfe1) et (Wolfe2)

Etape4.

Posez

$$x_{k+1} = x_k + \alpha_k d_k$$

Posez $k = k + 1$ et allez à Etape2

2.1.2 Convergence globale

Les méthodes à directions de descente et rechreches linéaires inexactes génèrent une suite $\{x_k\}_{k\in\mathbb{N}}$. L'déal est que cette suite $\{x_k\}_{k\in\mathbb{N}}$ converge vers la solution minimale globale du problème (P). En général on n'obtient pas ce résultat et se contente vérifie seulement

$$\lim_{k\longrightarrow\infty} \|\nabla f(x_k)\| = 0, \qquad \text{(conv-glob)}$$

ou un résultat encore plus faible

$$\lim_{k\longrightarrow\infty} \inf \|\nabla f(x_k)\| = 0. \qquad \text{(conv-glob(bis))}$$

Définition : On dit que la suite $\{x_k\}_{k\in\mathbb{N}}$ est globalement convergente, si pour tout point de départ x_0, l'une des deux conditions (conv-glob) ou (conv-glob(bis)) est réalisée.

Remarque : La condition (conv-glob(bis)) est plus faible que la condition (conv-glob).Elle implique qu'il existe seulement une sous suite $\{\|\nabla f(x_k)\|\}_{k\in\mathbb{N}_1}$ $(\mathbb{N}_1 \subset \mathbb{N})$ qui converge vers 0

2.1.3 Théorème de Zoudentijk

Introduction

Dans cette section, on va étudier la *contribution* de la recherche linéaire inexacte dans la convergence des algorithmes à directions de descente. Ce n'est qu'une contribution, parce que la recherche linéaire ne peut à elle seule assurer la convergence des itérés. On comprend bien que le choix de la directionde descente joue aussi un rôle. Cela se traduit par une condition, dite de *Zoutendijk*, dont on peut tirer quelques informations qualitatives intéressantes.

Définition : On dit qu'une règle de recherhe linéaire satisfait la *condition de Zoutendijk* s'il existe une constante $c > 0$ telle que pour tout indice $k \geq 1$ on ait

$$f(x_{k+1}) \leq f(x_k) - c\|\nabla f(x_k)\|^2 \cos^2\theta_k, \tag{2.3}$$

où θ_k est l'angle que fait d_k avec $-\nabla f(x_k)$, défini par

$$\cos\theta_k = \frac{-\nabla^{\intercal} f(x_k)\, d_k}{\|\nabla f(x_k)\|\, \|d_k\|}. \tag{2.4}$$

Dans la proposition suivante , on va servir de la condition de Zoutendijk.

Proposition : si la suite $\{x_k\}$ générée par un algorithme d'optimisation vérifie la condition de Zoutendijk (2.3) et si la suite $\{f(x_k)\}$ est minorée, alors

$$\sum_{k\geq 1} \|\nabla f(x_k)\|^2 \cos^2\theta_k < \infty \tag{2.5}$$

Preuve : En sommant les inégalités (2.3) ,on a

$$\sum_{k=1}^{l} \|\nabla f(x_k)\|^2 \cos^2\theta_k \leq \frac{1}{c}\left(f(x_1) - f(x_{l+1})\right), \tag{2.6}$$

Preuve. et puisque la suite $\{f(x_k)\}$ est minorée, c'est-à-dire $\exists \acute{c} > 0$ telle que pour tout k, $f(x_k) \geq \acute{c}$ alors

$$\sum_{k\geq 1} \|\nabla f(x_k)\|^2 \cos^2\theta_k < \infty \text{ quand } l \longrightarrow \infty,$$

■

Théorème de Zoudentijk

Le Théorème de Zoudentijk est le principal Théorème qu'on utilise pour démontrer la convergence des méthodes du gradient conjugué non quadratique et les méthodes quasi-Newtonniennes.Ce théorème est surtout efficace avec les rechreches linéaires inexactes de Wolfe et de Goldstein et nécessite quelques hypothèses sur f (f différentiable ∇f Lipschitzien) .

On dispose d'un résultat de convergence des algorithmes de descente avec recherche linéaire de Wolfe 1 et 2.

Théorème : (Théorème de Zoutendijk)

Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ supposée différentiable, de gradient Lipschitz et bornée inférieurement. Soit A un algorithme générant des itérés définis par :

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

où d_k est une direction de descente de f en x_k et $s_k > 0$ un pas vérifiant les conditions de Wolfe 1 et 2. Alors :

$$\sum \cos(\theta_k)^2 \|\nabla f(x_k)\|^2 < +\infty.$$

Preuve : D'après la seconde condition de Wolfe, on sait que : $\forall k$, $\nabla f(x_{k+1})^\top d_k \geq \epsilon_2 \left(\nabla f(x_k)^\top d_k\right)$, d'où :

$$\forall k, \ \ (\nabla f(x_{k+1}) - \nabla f(x_k))^\top d_k \geq (\sigma - 1)\left(\nabla f(x_k)^\top d_k\right).$$

De plus :

$$\begin{aligned}
(\nabla f(x_{k+1}) - \nabla f(x_k))^\top d_k &\leq \|\nabla f(x_{k+1}) - \nabla f(x_k)\| \, \|d_k\| && \text{(Cauchy-Schwartz)} \\
&\leq L \|x_{k+1} - x_k\| \, \|d_k\| && (\nabla f \ \ L\text{-Lipschitz}) \\
&\leq L\alpha_k \|d_k\|^2.
\end{aligned}$$

D'où en combinant les deux inégalités précédentes, on obtient :

$$\forall k, 0 \ \leq (\sigma - 1)\left(\nabla f(x_k)^\top d_k\right) \leq L\alpha_k \|d_k\|^2,$$

soit un pas α_k vérifiant :

$$\alpha_k \geq \frac{\sigma - 1}{L} \frac{\nabla f(x_k)^\top d_k}{\|d_k\|^2} > 0. \tag{2.7}$$

Une technique classique en optimisation pour obtenir des résultats de convergence consiste à se servir de la série de terme général $f(x_k) - f(x_{k+1})$: c'est une série à termes positifs qui converge si sa somme partielle de rang n (égale à $f(x_0) - f(x_{n+1})$)est majorée. C'est ici le cas puisque f est est supposée bornée inférieurement :

$$\begin{aligned} f(x_k) - f(x_{k+1}) &\geq -\beta_1 \alpha_k \nabla f(x_k)^\top d_k && \text{(condition de Wolfe 1)} \\ &\geq \beta_1 \frac{1-\sigma}{L} \frac{\left(\nabla f(x_k)^\top d_k\right)^2}{\|d_k\|^2} && \text{(d'après (2.7))} \\ &\geq \beta_1 \frac{1-\sigma}{L} \cos(\theta_k)^2 \|\nabla f(x_k)\|^2 . \end{aligned}$$

D'après le Théorème de comparaison de deux séries numériques à termes positifs, sachant que la série de terme général $f(x_k) - f(x_{k+1})$ converge, on en déduit la convergence de la série $\sum \cos(\theta_k)^2 \|\nabla f(x_k)\|^2$.

Ce Théorème implique que si la direction de descente d_k n'est pas orthogonale à la direction de $-\nabla f(x_k)$ à l'infini i.e. si

$$\exists c > 0, \forall k \in \mathbb{N}, \ \cos(\theta_k) \geq c,$$

alors la série $\sum \|\nabla f(x_k)\|^2$ converge et la suite $(\nabla f(x_k)_k)$ tend vers 0 quand $k \longrightarrow +\infty$. Ce Théorème nous donne donc la convergence globale de l'algorithme de descente considéré.

Conséquences tirées du théorème de Zoutendijk

$\sum_{k=0}^{\infty} \cos^2(\theta_k) \|\nabla f(x_k)\|^2 < +\infty$ implique

$$\lim_{k \longrightarrow \infty} \cos^2(\theta_k) \|\nabla f(x_k)\|^2 = 0 \tag{2.8}$$

La relation (2.8) n'implique pas en général que

$$\lim_{k \longrightarrow \infty} \|\nabla f(x_k)\|^2 = 0. \tag{2.9}$$

Si $\cos(\theta_k)$ tend vers zéro quand k tend vers l'infini, $\|\nabla f(x_k)\|$ peut ne pas tendre vers zéro. Il faut donc éviter que θ_k tendre vers $\frac{\pi}{2}$,pour s'assurer que : $\lim_{k \longrightarrow \infty} \|\nabla f(x_k)\|^2 = 0$. C'est le but de la définition suivante :

Définition : Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$, une fonction bornée inférieurement, différentiable.Soit un algorithme itératif qui génère une suite $\{x_k\}_{k \in \mathbb{N}}$ dans $\mathbb{R}^n$ définie par $x_0 \in \mathbb{R}^n$ quelconque et les itérations

$$x_{k+1} = x_k + \alpha_k d_k, k = 0, 1, 2, ...$$

La suite $\{d_k\}_{k \in \mathbb{N}}$ est dite en relation gradient avec la suite $\{x_k\}_{k \in \mathbb{N}}$ si on a

$$\exists \mu > 0 : \cos(\theta_k) \geq \mu : \forall k \in \mathbb{N} \tag{2.10}$$

Sens géométrique de la condition (2.10)

La condition (2.10) veut dire que la suite $\{\cos(\theta_k)\}_{k \in \mathbb{N}}$est minorée par un nombre strictement positif μ. La condition (2.9) n'est pas garantie, si la suite $\cos(\theta_k) \longrightarrow 0$ quand $k \longrightarrow \infty$.Géométriquent, cette condition veut dire que la suite angles

$$\left\{\theta_k = \widehat{-\nabla f(x_k), d_k}\right\}_{k \in \mathbb{N}}$$

ne doit pas tendre vers $\frac{\pi}{2}$ quand k tend vers l'infini ou en d'autre termes, il ne faut pas que les deux suites de directions

$$\{d_k\}_{k\in\mathbb{N}} \text{ et } \{-\nabla f(x_k)\}_{k\in\mathbb{N}} \ ,$$

finissent par devenir orthogonales quand $k \longrightarrow \infty$.

Si la suite $\{d_k\}_{k\in\mathbb{N}}$ est en relation gradient avec la suite $\{x_k\}_{k\in\mathbb{N}}$. On obtient le Théorème suivant qui assure la convergence globale de la suite $\{x_k\}_{k\in\mathbb{N}}$

Théorème : **(de Zoudentijk-bis)**

Soit f :$\mathbb{R}^n \longrightarrow \mathbb{R}$, une fonction bornée inférieurement, différentiable,dont le gradient est continu au sens de Lipshitz, c'est à dire qu'il existe $M > 0$ telle que

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

Soit un algorithme itératif qui génère une suite $\{x_k\}_{k\in\mathbb{N}}$ dans $\mathbb{R}^n$ définie par $x_0 \in \mathbb{R}^n$ quelconque et les itérations

$$x_{k+1} = x_k + \alpha_k d_k, \ \ k = 0, 1, 2, ...$$

avec d_k direction de descente (c'est à dire $\nabla f(x_k)^\top d_k < 0$) et α_k vérifiant les conditions de (Wolfe1) et (Wolfe2). Si on plus on a

$$\exists \mu > 0 : \cos(\theta_k) \geq \mu : \forall k \in \mathbb{N}. \tag{2.11}$$

où θ_k est l'angle défini par

$$\cos(\theta_k) = \frac{-\nabla f(x_k)^\top d_k}{\|\nabla f(x_k)\| \|d_k\|}.$$

Alors on a

$$\lim_{k\longrightarrow\infty} \|\nabla f(x_k)\|^2 = 0.$$

Preuve : D'après le Théorème de Zoutendijk, on a

$$\sum_{k=0}^{\infty} \cos^2(\theta_k) \|\nabla f(x_k)\|^2 < +\infty.$$

Ceci implique

$$\lim_{k \longrightarrow \infty} \cos^2(\theta_k) \|\nabla f(x_k)\|^2 = 0. \tag{2.12}$$

Notons

$$u_k = \cos^2(\theta_k), \quad v_k = \|\nabla f(x_k)\|^2, \; w_k = u_k \; v_k \tag{2.13}$$

Alors (2.12) et (2.13) impliquent

$$u_k \geq \mu^2 > 0, \; v_k > 0, \; \lim_{k \longrightarrow \infty} \; w_k = 0 \;, \tag{2.14}$$

(2.13) implique

$$v_k = \frac{1}{u_k} w_k, \tag{2.15}$$

(2.14) et (2.15) donnent

$$0 \leq \; v_k \leq \frac{1}{\mu^2} \; w_k, \tag{2.16}$$

$\lim_{k \longrightarrow \infty} \; w_k = 0$ et (2.16) impliquent

$$\lim_{k \longrightarrow \infty} \|\nabla f(x_k)\| = 0.$$

2.2 Convergence de la méthode du gradient.

2.2.1 Cas de la recherche linéaire inexacte de Wolfe

Théorème : (convergence de la méthode de la plus forte pente avec la recherche linéaire de Wolfe).

Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$, une fonction bornée inférieurement, différentiable,dont le gradient est continu au sens de Lipshitz, c'est à dire qu'il existe $G > 0$ telle que

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

dans l'ensemble $\delta(x_0) = \{x \in \mathbb{R}^n : f(x) \leq f(x_0)\}$, $x_0 \in \mathbb{R}^n$ quelconque .

Considérons l'algorithme de la méthode de descente avec la recherche linéaire de Wolfe (voir Chapitre3_S1). Soit $\{x_k\}$ une suite génère par cet algorithme. Alors ou bien l'algorithme s'arrete après un nombre fini d'itérations eu n un poit x_{k0} tel que $\nabla f(x_{k0}) = 0$, ou bien il génère une suite infini $\{x_k\}_{k\in\mathbb{N}}$ telle que

$$\nabla f(x_k)_{k\longrightarrow\infty} \longrightarrow 0.$$

Preuve : C'est une conséquence directe du Théorème de Zoudentijk-bis . En effet les hypothèses sur f sont les mêmes hypothèses du théorème de Zoudentijk-bis.

Les donnés de notre Théorème sont :

$$d_k = -\nabla f(x_k) \Longrightarrow \theta_k = \widehat{-\nabla f(x_k), \ -} \nabla f(x_k) = 0$$

Preuve. donc

$$\cos(\theta_k) = 1 : \forall k.$$

par conséquent, en prenant par exemple $\mu = \frac{1}{2}$, on a

$$\forall k \in \mathbb{N} : \cos(\theta_k) > \frac{1}{2}.$$

D'après le Théorème de Zoudentijk, on a

$$\lim_{k\longrightarrow\infty} \|\nabla f(x_k)\| = 0.$$

Puisque $\|.\|$ est une fonction continue, donc

$$\nabla f\left(x_k\right)_{k \longrightarrow \infty} \longrightarrow 0.$$

■

2.2.2 Convergence globale de l'algorithme de Newton avec recherche linéaire inexacte de Wolfe

Théorème : Soit f :$\mathbb{R}^n \longrightarrow \mathbb{R}$, une fonction bornée inférieurement, différentiable,dont le gradient est est continu au sens de Lipshitz.On supose que le choix de la direction de descente est dans cet algorithme est le suivant : $d_k = -H_k^{-1}(\nabla f(x_k)$, où où H_k est une matrice symétrique définie positive. Couplée avec une recherche linéaire de Wolfe.On suppose aussi que le conditionnement de la matrice H_k reste borné au cours des itérations, alors l'algorithme de Newton avec recherche linéaire converge globalement.

Preuve : **(voir td)**

Bibliographie

[1] BENZINE RACHID. COURS OPTIMISATION SANS CONTRAINTES TOME1.sur internet.

[2] Aude RONDEPIERRE & Pierre WEISS.MÉTHODES STANDARDS EN OPTIMISATION NON LINÉAIRE DÉTERMINIS