Optimisation numérique 3
Méthodes de recherche par ligne I
(chapitre 3.1 et 3.2 du livre*)

JULIEN Cyril

Fribourg, le 8 octobre 2009

Table des matières

*NOCEDAL Jorge et WRIGHT Stephen J., *Numerical Optimization*, Springer-Verlag, 1999.

1 Introduction

1.1 Cadre général de l'optimisation sans contrainte de fonctions continues

Soient $x \in \mathbb{R}^n$ un vecteur de variables et $f : \mathbb{R}^n \to \mathbb{R}$ la fonction lisse de coût à optimiser. Alors, sans restriction de la généralité, le problème d'optimisation sans contrainte revient à calculer $\min_{x \in \mathbb{R}^n} f(x)$.

1.2 Algorithme itératif en tant que démarche numérique

Numériquement, on procédera comme suit : on choisit un x_0 (que l'on estime comme bon, assez proche de la solution si possible), et on considère une suite convenable $(x_k)_{k \in \mathbb{N}}$ (dépendante de f et déterminée récursivement), de telle sorte qu'elle converge vers la solution x^* t.q. $f(x^*) = \min_{x \in \mathbb{R}^n} f(x)$. Évidemment, on se contente enfin d'un résultat numérique x_m (pour un $m \in \mathbb{N}$) que l'on estime suffisamment proche de x^*. Traditionellement, il existe deux stratégies pour définir itérativement la suite $(x_k)_{k \in \mathbb{N}}$: celle de recherche par ligne et celle de la région de confiance. La première sera étudiée ici.

1.3 Méthodes de recherche par ligne

L'idée des méthodes de recherche par ligne est de déterminer à chaque pas la direction et la longueur du pas suivant. Mathématiquement, $\forall k \in \mathbb{N}$, on choisit une direction $p_k \in \mathbb{R}^n$ et une longueur de pas $\alpha_k \in \mathbb{R}_{>0}$, ensuite, on pose $x_{k+1} := x_k + \alpha_k p_k$. Pour que la suite soit appropriée, on demande que p_k et α_k soient choisis pour qu'approximativement, $f(x_{k+1}) \approx \min_{\alpha_k, p_k} f(x_k + \alpha_k p_k)$. On connait déjà certains résultats sur p_k, mais α_k et p_k seront respectivement plus étudiés dans les chapitres 2 et 3.

2 Longueur du pas

Soient un $k \in \mathbb{N}$ fixe et $p_k \in \mathbb{R}^n$ donné. On veut donc maintenant trouver une bonne longueur $\alpha > 0$ qui minimise $\phi(\alpha) := f(x_k + \alpha p_k)$. Idéalement, on chercherait à minimiser exactement ϕ globalement. Cependant, cette opération coûterait démesurément trop chère à cause des évaluations trop nombreuses de f et de ∇f[1]. Donc, on utilisera des stratégies

[1] $\nabla f = \text{grad} f$

plus pratiques qui, en un coût minimal, diminuent f de manière adéquate ; ces procédés sont des recherches inexactes par ligne. Typiquement, un tel algorithme essaye une suite de candidats-valeurs α et l'arrête quand un de ces candidats satisfait certaines propriétés préétablies. Plus précisément, une première étape consiste à trouver un intervalle contenant les différentes longeurs du pas désirées ; puis une seconde étape consiste à extraire la meilleure longueur contenue dans l'intervalle. Notons qu'il existe des méthodes plus compliquées.

2.1 Condition de Wolfe

Une condition populaire de recherche inexacte par ligne est la diminution suffisante, ou condition d'Armijo :

$$f(x_k + \alpha_k p_k) \leq f(x_k) + c_1 \alpha_k \nabla f(x_k)^t p_k \tag{1}$$

pour $0 < c_1 < 1$.

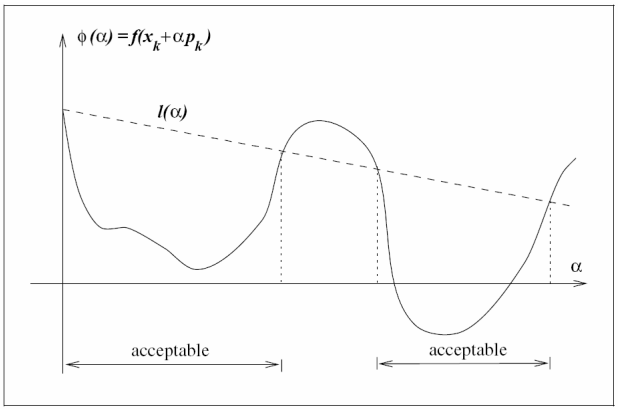

FIG. 1 – Condition de diminution suffisante

En d'autres termes, la réduction de f doit être proportionelle à la longueur du pas α_k et à la dérivée directionnelle $\nabla f(x_k)^t p_k$. En pratique, remarquons que la valeur de $c_1 = 10^{-4}$ s'utilise fréquemment. Si l'on pose $l(\alpha) := f(x_k) + c_1 \alpha \nabla f(x_k)^t p_k$ (à pente négative), on peut reformuler cette condition (1) comme suit : $\phi(\alpha_k) \leq l(\alpha_k)$.

L'ajout d'une seconde condition paraît indispensable pour éviter les pas trop courts, car la condition (1) peut renvoyer des valeurs trop petites pour qu'elles soient intéressantes. Cette seconde condition, appelée condition de courbure, est introduite comme suit :

$$\nabla f(x_k + \alpha_k p_k)^t p_k \geq c_2 \nabla f(x_k)^t p_k \tag{2}$$

pour $c_1 < c_2 < 0$, i.e. $\phi'(\alpha_k) \geq c_2\phi'(0)$.

Ce qui est justifié car d'une part, si $\phi'(\alpha)$ est trop petite, alors on peut aller chercher plus loin dans la même direction, i.e. prendre un α plus grand ; et d'autre part, si la pente $\phi'(\alpha)$ est à peine négative ou même positive, alors c'est un signe que l'on ne peut trouver d'autres α dans la même direction (f serait alors trop mal estimée), donc la recherche se terminera.

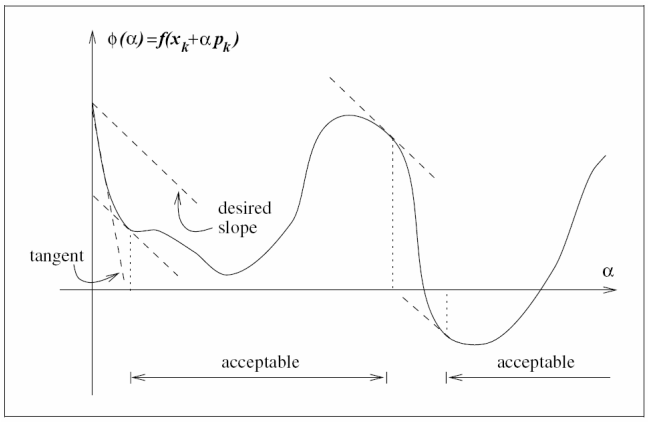

FIG. 2 – Condition de courbure

La condition de Wolfe, (W), regroupe les conditions de diminution suffisante (1) et de courbure (2) :

$$\begin{aligned} \phi(\alpha_k) &\leq l(\alpha_k) \\ \phi'(\alpha_k) &\geq c_2\phi'(0). \end{aligned}$$

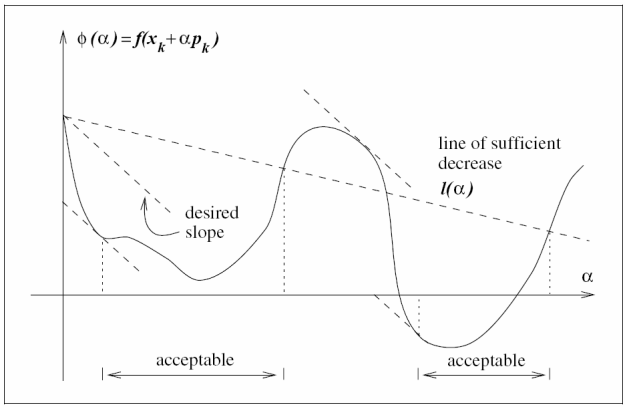

FIG. 3 – Condition de Wolfe

Il est possible de forcer (W) à devenir plus restrictive. Pour cela, on remplace (2) par $|\nabla f(x_k + \alpha_k p_k)^t p_k| \leq c_2 |\nabla f(x_k)^t p_k|$. On l'appelle la condition

forte de Wolfe, (W'), et la seule différence avec (W) est que les trop grandes pentes $\phi'(\alpha)$ ne sont pas acceptées.

Lemme 1 *Soient p_k une direction convenable de x_k, et $f \in C^1(\mathbb{R}^n)$ bornée inférieurement sur $\{x_k + \alpha p_k | \alpha > 0\}$. Si $0 < c_1 < c_2 < 1$, alors il existe des intervalles de longueurs du pas satisfaisant (W), respectivement (W').*
Preuve :

Comme $\phi(\alpha) = f(x_k + \alpha p_k)$ est borné inférieurement et $0 < c_1 < 1$, alors la ligne $l(\alpha) = f(x_k) + c_1 \alpha \nabla f(x_k)^t p_k$ intersecte ϕ au moins une fois. Soit $\alpha' > 0$ la plus petite valeur t.q. $f(x_k + \alpha' p_k) = f(x_k) + c_1 \alpha' \nabla f(x_k)^t p_k$. Alors (1) est vérifiée pour $\alpha < \alpha'$. De plus :

$$f(x_k + \alpha' p_k) - f(x_k) = \alpha' \nabla f(x_k + \alpha'' p_k)^t p_k$$

pour un $\alpha'' \in]0, \alpha'[$ par le théorème des valeurs intermédiaires. Donc :

$$\nabla f(x_k + \alpha'' p_k)^t p_k = c_1 \nabla f(x_k)^t p_k > c_2 \nabla f(x_k)^t p_k$$

comme $c_1 < c_2$ et $\nabla f(x_k)^t p_k < 0$. Enfin, la continuité de f implique qu'il existe un intervalle autour de α'' t.q. (W) et (W') soient vérifiées.

□

2.2 Condition de Goldstein

Comme la condition de Wolf, celle de Goldstein, (G), se constitue de (1) et d'une autre partie qui sert à ne pas s'arrêter pour des α trop petits :

$$f(x_k) + (1 - c)\alpha \nabla f(x_k)^t p_k \leq f(x_k + \alpha p_k) \tag{3}$$

pour $0 < c < 1/2$, le même que dans (2), ($c = c_1$).

Elle a l'énorme désavantage d'exclure parfois tous les minima de ϕ. Néanmoins, elle a en commun avec (W) la convergence de même type.

2.3 Diminution suffisante

En fait, si l'on choisit à l'avance certains α (dans un intervalle), alors la condition (1) suffit. Un algorithme exprimant plus précisément le concept multiplierait une valeur de départ par un scalaire tant que la condition (1) n'est pas vérifiée et terminerait avec une valeur retournée juste avant que (1) soit vérifiée.

3 Convergence des méthodes de recherche par ligne

Dans cette section, on discutera de la direction p_k qui est cruciale, en plus de la longueur α_k, pour la convergence. Une notion importante pour ce but est la suivante : il s'agit de l'angle entre p_k et $-\nabla f(x_k)$ (la direction la plus pentue), noté θ_k et défini donc par

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

On peut montrer que la méthode qui a la plus petite pente (la plus profonde descente) converge globalement. Pour d'autres algorithmes, on décrit de combien p_k peut différer de la plus grande descente pour toujours avoir une convergence globale. De manière générale, on prouve le théorème suivant, dû à Zoutendijk.

Théorème 2 *Soient l'itération $x_{k+1} = x_k + \alpha_k p_k$ t.q. α_k satisfait la condition de Wolf, et $f \in C^1(N)$ et bornée dans $\mathbb{R}^n$ pour $N \subseteq \mathbb{R}^n$ ouvert avec $\{x \in \mathbb{R}^n | f(x) \leq f(x_0)\} \subseteq N$ où x_0 est le point de départ. Soit ∇f lipschitzienne. Alors :*

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

Preuve :

D'une part, par (2),

$$\begin{aligned} \nabla f(\underbrace{x_k + \alpha p_k}_{=x_{k+1}})^t p_k &\geq c_2 \nabla f(x_k)^t p_k \qquad | - \nabla f(x_k)^t p_k \\ \Rightarrow \quad (\nabla f(x_{k+1}) - \nabla f(x_k))^t p_k &\geq (c_2 - 1)\nabla f(x_k)^t p_k. \end{aligned}$$

D'autre part, comme ∇f est lipschitzienne, avec constante L,

$$(\nabla f(x_{k+1}) - \nabla f(x_k))^t p_k \leq |(\nabla f(x_{k+1}) - \nabla f(x_k))^t p_k| \leq \alpha_k L \|p_k\|^2.$$

Donc,

$$(c_2 - 1)\nabla f(x_k)^t p_k \leq \alpha_k L \|p_k\|^2 \Rightarrow \alpha_k \geq \frac{(c_2 - 1)\nabla f(x_k)^t p_k}{L\|p_k\|^2}.$$

En insérant cela dans (1) et en itérant,

$$
\begin{aligned}
f(x_{k+1}) \leq f(x_k) + c_1\alpha_k \overbrace{\nabla f(x_k)^t p_k}^{<0} &\leq f(x_k) + \frac{c_1(c_2-1)(\nabla f(x_k)^t p_k)^2}{L\|p_k\|^2} \\
&\leq f(x_k) + \underbrace{\frac{c_1(c_2-1)}{L}}_{:=-c} \cos^2\theta_k \|\nabla f(x_k)\|^2 \\
&\leq \ldots \\
&\leq f(x_0) - c\sum_{j=0}^{k} \cos^2\theta_j \|\nabla f(x_j)\|^2.
\end{aligned}
$$

D'où, en prenant la limite,

$$
\sum_{j\geq 0} \cos^2\theta_j \|\nabla f(x_j)\|^2 \leq \frac{f(x_0) - f(x_{k+1})}{c} < \infty,
$$

parce que f est bornée.

□

Remarquons les différents points suivants à propos du théorème : tout d'abord, il existe des résultats similaires avec (W') ou (G) à la place de (W) ; ensuite, les hypothèses de ce théorème ne sont pas trop restrictives en pratique ; enfin, une conséquence de ce théorème est que $\cos^2\theta_k\|\nabla f(x_k)\|^2 \to 0$. D'une part, cette dernière affirmation permet de montrer la convergence globale d'algorithmes de recherche par ligne, et d'autre part, elle entraîne que si $\cos\theta_k \neq 0$, (i.e. par exemple si $\theta_k \in]-\pi/2, \pi/2[$), alors $\lim_{k\to\infty} \|\nabla f(x_k)\| = 0$. En d'autres termes, le fait d'être certain que la direction de recherche n'est jamais suceptible à l'orthogonalité avec le gradient assure que la norme du gradient converge vers zéro. En particulier, la méthode à la plus profonde descente (pour laquelle $\theta_k = 0$) obtient une suite de gradient qui converge vers zéro car cette recherche par ligne satisfait la condition de Wolfe.

Clarifions encore la définition d'une convergence globale dans notre contexte. On dit qu'un algorithme converge globalement ssi $\lim_{k\to\infty} \|\nabla f(x_k)\| = 0$. Cette définition reste générale ; on ne peut garantir que la méthode converge vers un minimum, mais seulement vers des points stationnaires. Pour inclure la convergence vers un minimum local, on ajoutera des conditions sur p_k.

4 Conclusion

Pour conclure, traitons deux problèmes. D'abord, considérons la méthode de Newton, qui satisfait (W) et où $p_k = -B_k^{-1}\nabla f(x_k)$ avec B_k une matrice

symétrique et définie positive. Si B_k a un nombre de condition borné, i.e. il existe une constante $M \in \mathbb{R}$ t.q. $\|B_k\|\|B_k^{-1}\| \leq M \quad \forall k$, alors $\cos\theta_k \geq 1/M$. Donc $\lim_{k\to\infty} \|\nabla f(x_k)\| = 0$ et cette méthode converge globalement.

Enfin, pour d'autres méthodes satisfaisant (W), comme celle du gradient conjugué, on est seulement capable de montrer que $\liminf_{k\to\infty} \|\nabla f(x_k)\| = 0$, i.e. que seulement une sous-suite de $(\|\nabla f(x_k)\|)_{k\in\mathbb{N}}$ converge vers zéro. L'idée de la preuve utilise Zoutendijk et procède par contradiction : si l'affirmation s'avérerait fausse, alors il existerait $\gamma > 0$ t.q. $\|\nabla f(x_k)\| > \gamma \ \forall k$ assez grand ; donc, par Zoutendijk, $\cos\theta_k \to 0$ et il suffirait donc de voir qu'une sous-suite de $(\cos\theta_k)_{k\in\mathbb{N}}$ est bornée en dessus de zéro (par contraposition).

On applique cette technique de preuve pour démontrer la convergence globale d'ensembles généraux d'algorithmes. Par exemple, on considère n'importe quel algorithme pour lequel chaque itération produit une diminution de la fonction f avec la pente la plus raide et une longueur de pas qui vérifie la condition de Wolfe (ou celle de Goldstein). Ainsi, $\cos\theta_k = 1 \Rightarrow \lim_{k\to\infty} \|\nabla f(x_k)\| = 0$. Par conséquent, bien que le progrès du pas à pente la plus grande ne soit pas optimal, il converge tout de même globalement. Affaire à suivre...