

Optimisation pour les fonctions différentiables

Cornel Marius MUREA

Laboratoire de Mathématiques, Informatique et Applications,

Université de Haute-Alsace

e-mail: cornel.murea@uha.fr

<http://www.edp.lmia.uha.fr/murea/>

Mulhouse, le 19/11/2009

Optimisation sans ou avec contraintes

Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$.

Optimisation sans contraintes

minimiser $f(x)$, pour $x \in \mathbb{R}^n$.

Optimisation avec contraintes

minimiser	$f(x)$
sous les contraintes	$h_1(x) = 0$
	$\vdots$
	$h_m(x) = 0$
	$g_1(x) \leq 0$
	$\vdots$
	$g_p(x) \leq 0$

où $h_1, \dots, h_m : \mathbb{R}^n \rightarrow \mathbb{R}$ et $g_1, \dots, g_p : \mathbb{R}^n \rightarrow \mathbb{R}$.

Notations

Le gradient

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(x) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x) \end{pmatrix} \in \mathbb{R}^n$$

Le Hessien

$$H_f(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(x) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(x) \end{pmatrix}$$

Pour l'optimisation sans contraintes:

x^* est **minimum local** s'il existe un voisinage U de x^* tel que $f(x^*) \leq f(x)$, $\forall x \in U$;

x^* est **minimum global** si $f(x^*) \leq f(x)$, $\forall x \in \mathbb{R}^n$.

Conditions nécessaires ou suffisantes d'optimalité locale

Définition Une matrice A est **semi-définie positive** si et seulement si $x^T A x \geq 0, \forall x \in \mathbb{R}^n$.

Définition Une matrice A est **définie positive** si et seulement si $x^T A x > 0, \forall x \in \mathbb{R}^n, x \neq 0$.

Conditions nécessaires d'ordre 1

Si x^* est **minimum local** et f est différentiable, alors $\nabla f(x^*) = 0$.

Conditions nécessaires d'ordre 2

Si x^* est **minimum local** et f est deux fois différentiable, alors $\nabla f(x^*) = 0$ et en plus $H_f(x^*)$ est une matrice semi-définie positive.

Conditions suffisantes

Si f est deux fois différentiable, $\nabla f(x^*) = 0$ et $H_f(x^*)$ est une matrice définie positive, alors x^* est **minimum local**.

Conditions suffisantes d'optimalité globale

Définition $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est **convexe** si et seulement si $\forall x, y \in \mathbb{R}^n$ et $\forall \lambda \in [0, 1]$, $f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$.

Définition $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est **strictement convexe** ssi $\forall x, y \in \mathbb{R}^n$, $x \neq y$ et $\forall \lambda \in (0, 1)$, $f(\lambda x + (1 - \lambda)y) < \lambda f(x) + (1 - \lambda)f(y)$.

Si f est différentiable, convexe et $\nabla f(x^*) = 0$ alors x^* est **minimum global**.

Méthodes numériques de type descente

On veut construire la suite $x^k \in \mathbb{R}^n$, de la forme :

$$x^{k+1} = x^k + \rho_k d^k$$

où $\rho_k \in \mathbb{R}$ est le pas et $d^k \in \mathbb{R}^n$ est la direction.

On dit que $d^k \in \mathbb{R}^n$ est une **direction de descente** ssi $(d^k)^T \nabla f(x^k) < 0$.

$$f(x^k + h) \approx f(x^k) + h^T \nabla f(x^k)$$

$$\min_{h \in \mathbb{R}^n, \|h\|=1} h^T \nabla f(x^k) = -\frac{(\nabla f(x^k))^T \nabla f(x^k)}{\|\nabla f(x^k)\|}$$

On peut choisir $d^k = -\nabla f(x^k)$ (méthode de type gradient)

La méthodes du gradient à pas optimal (de la plus forte pente)

Pas 0: Choisir un point de départ $x^0 \in \mathbb{R}^n$, $k = 0$

Pas 1: Si $\nabla f(x^k) = 0$ alors x^k est un minimum local. Stop

Pas 2: Trouve $\rho_k \in \mathbb{R}$, la solution du problème

$$\min_{\rho \geq 0} f(x^k - \rho \nabla f(x^k)) = f(x^k - \rho_k \nabla f(x^k))$$

Pas 3: $x^{k+1} = x^k - \rho_k \nabla f(x^k)$, $k = k + 1$, retourner au Pas 1.

Définition $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est **elliptique** ssi $\lim_{\|x\| \rightarrow +\infty} f(x) = +\infty$.

Théorème Si f est elliptique et différentiable, alors la méthodes du gradient à pas optimal est convergente.

La méthodes du gradient à pas fixe

Pas 0: Choisir un point de départ $x^0 \in \mathbb{R}^n$, $k = 0$

Pas 1: Si $\nabla f(x^k) = 0$ alors x^k est un minimum local. Stop

Pas 2: $x^{k+1} = x^k - \rho \nabla f(x^k)$, $k = k + 1$, retourner au Pas 1.

Théorème On suppose que f est différentiable et qu'il existe deux constantes $\alpha > 0$, $M > 0$ telles que :

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \alpha \|x - y\|^2, \quad \forall x, y \in \mathbb{R}^n$$

$$\|\nabla f(x) - \nabla f(y)\| \leq M \|x - y\|^2, \quad \forall x, y \in \mathbb{R}^n$$

Si $0 < a \leq \rho \leq b \leq \frac{M^2}{\alpha}$, alors la méthodes du gradient à pas fixe est convergente.

La méthode du gradient conjugué (I)

Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ quadratique, c'est à dire $f(x) = \frac{1}{2}x^T Qx + c^T x$ où Q est une matrice symétrique, définie positive et $c \in \mathbb{R}^n$.

Le problème d'optimisation sans contraintes $\inf_{x \in \mathbb{R}^n} f(x)$ a une solution unique x^* caractérisée par $Qx^* + c = 0$.

Définition Deux directions $d^1, d^2 \in \mathbb{R}^n$ sont Q **conjuguées** ssi $(d^1)^T Q d^2 = 0$.

Théorème Soient $\{d^i\}_{i=0}^{k-1}$, k directions mutuellement conjuguées. Soit $x^0 \in \mathbb{R}^n$ et la suite générée par

$$\begin{cases} x^{i+1} = x^i + \rho_i d^i \\ \rho_i = -\frac{\nabla f(x^i)^T d^i}{(d^i)^T Q d^i} \end{cases}$$

Alors x^k est l'unique minimum de f restreint à $x^0 + B_k$ où $B_k = [d^0, \dots, d^{k-1}]$ est le sous espace vectoriel engendré par les vecteurs $d^0, \dots, d^{k-1}$.

La méthode du gradient conjugué (II)

Comment on peut construire les directions mutuellement conjuguées d^i ?

Pas 0: Choisir $x^0 \in \mathbb{R}^n$ et poser $d^0 = -\nabla f(x^0)$, $k = 0$.

Pas 1: Si $\nabla f(x^k) = 0$ alors STOP. $x^* = x^k$

Pas 2: Calculer $x^{k+1} = x^k + \rho_k d^k$ où

$$\rho_k = -\frac{\nabla f(x^k)^T d^k}{(d^k)^T Q d^k}$$

$$\beta_k = \frac{\nabla f(x^{k+1})^T Q d^k}{(d^k)^T Q d^k}$$

$$d^{k+1} = -\nabla f(x^{k+1}) + \beta_k d^k$$

Pas 3: $k = k + 1$, retourner au Pas 1.

Théorème La méthode du gradient conjugué converge en au plus n itérations.

Algorithmes de type Newton

On veut résoudre $\nabla f(x^*) = 0$.

Algorithme de Newton

Pas 0: Choisir $x^0 \in \mathbb{R}^n$, et poser $k = 0$

Pas 1: Si $\nabla f(x^k) = 0$ alors STOP. $x^* = x^k$

Pas 2: $x^{k+1} = x^k - (H_f(x^k))^{-1} \nabla f(x^k)$ $k = k + 1$, retourner au Pas 1.

Algorithmes de type Newton

$$x^{k+1} = x^k - \left(A_k(x^{k'}) \right)^{-1} \nabla f(x^k) \text{ où } 0 \leq k' \leq k$$

$$A_k(x^{k'}) = H_f(x^k), k' = k, \text{ (algorithme de Newton)}$$

$$A_k(x^{k'}) = H_f(x^{k'}), k' = p \left[\frac{k}{p} \right], p \geq 2$$

$$A_k(x^{k'}) = H_f(x^0)$$

$$A_k(x^{k'}) = A_0 \text{ où } A_0 \text{ est une matrice inversible.}$$

Algorithmes de type quasi Newton

Dans les algorithmes de type Newton, il faut calculer l'inverse d'une matrice $\left(A_k(x^{k'})\right)^{-1}$.

Algorithmes de type quasi Newton

Pas 0: Choisir $x^0 \in \mathbb{R}^n$, et poser $k = 0$

Pas 1: Si $\nabla f(x^k) = 0$ alors STOP. $x^* = x^k$

Pas 2: $x^{k+1} = x^k - \rho_k H_k \nabla f(x^k)$ où H_k est une matrice qui approche $(H_f(x^k))^{-1}$ et le pas ρ_k est déterminé comme la solution d'un problème d'optimisation mono-dimensionnelle

$$\inf_{\rho \geq 0} f(x^k - \rho H_k \nabla f(x^k))$$

Pas 3: $k = k + 1$, retourner au Pas 1.

Formules de correction pour la matrice H_k

Soit $f(x) = \frac{1}{2}x^T Qx + c^T x$ quadratique.

On a $\nabla f(x) = Qx$ et $H_f(x) = Q$ donc $H_f^{-1}(x) = Q^{-1}$.

$$Q(x - y) = \nabla f(x) - \nabla f(y), \quad x - y = Q^{-1}(\nabla f(x) - \nabla f(y))$$

$$x^{k+1} - x^k = Q^{-1}(\nabla f(x^{k+1}) - \nabla f(x^k))$$

On cherche H_{k+1} qui vérifie

$$x^{k+1} - x^k = H_{k+1}(\nabla f(x^{k+1}) - \nabla f(x^k))$$

Différentes formules de correction de type $H_{k+1} = H_k + \Delta_k$ ont été proposées.

Algorithme de Davidon-Fletcher-Powell

Pas 0: Choisir une matrice H_0 symétrique, définie positive, $x^0 \in \mathbb{R}^n$, et poser $k = 0$

Pas 1: Si $\nabla f(x^k) = 0$ alors STOP, si non $d^k = -H_k \nabla f(x^k)$

Pas 2: Déterminer le pas ρ_k

$$\inf_{\rho \geq 0} f(x^k + \rho d^k) = f(x^k + \rho_k d^k)$$

Pas 3: Poser $x^{k+1} = x^k + \rho_k d^k$,
 $p_k = x^{k+1} - x^k$ et $q_k = \nabla f(x^{k+1}) - \nabla f(x^k)$

Pas 4: La formule de correction est

$$H_{k+1} = H_k + \frac{p_k p_k^T}{p_k^T q_k} - \frac{H_k q_k q_k^T H_k}{q_k^T H_k q_k}$$

Pas 5: $k = k + 1$, retourner au Pas 1.

Algorithme de Broyden, Fletcher, Goldforb, Shano

$$\delta_k = x^{k+1} - x^k$$

$$\gamma_k = \nabla f(x^{k+1}) - \nabla f(x^k)$$

La formule de correction

$$H_{k+1} = H_k + \left(1 + \frac{\gamma_k^T H_k \gamma_k}{\delta_k^T \gamma_k}\right) \frac{\delta_k \delta_k^T}{\delta_k^T \gamma_k} - \frac{\delta_k \gamma_k^T H_k + H_k \gamma_k \delta_k^T}{\delta_k^T \gamma_k}$$

Tests d'arrêt

Critère 1

$$\left\| \nabla f(x^k) \right\|_{\infty} = \max_{i=1, \dots, n} \left| \frac{\partial f}{\partial x_i}(x^k) \right| \leq \epsilon$$

Critère 2

$$\left\| \nabla f(x^k) \right\|_2 = \sqrt{\sum_{i=1}^n \left(\frac{\partial f}{\partial x_i}(x^k) \right)^2} \leq \epsilon$$

Critère 3

$$\left| f(x^{k+1}) - f(x^k) \right| \leq \epsilon$$

Recherches mono-dimensionnelles économiques

$$\inf_{\rho \geq 0} f(x^k + \rho d^k)$$

On pose $g : [0, +\infty) \rightarrow \mathbb{R}$, $g(\rho) = f(x^k + \rho d^k)$.

Critère de Goldstein et Armijo : trouver ρ_k tel que

$$g(0) + (1 - \lambda) \theta_k g'(0) \leq g(\theta_k) \leq g(0) + \lambda \theta_k g'(0)$$

où $\lambda \in (0, 1/2)$.

Bibliographie

Minoux, Michel Programmation mathématique: théorie et algorithmes. Tome 1. Collection Technique et Scientifique des Télécommunications. Dunod, Paris, 1983

Luenberger, David G. Linear and nonlinear programming. Second edition. Kluwer Academic Publishers, Boston, MA, 2003.

Dennis, J. E., Jr.; Schnabel, Robert B. Numerical methods for unconstrained optimization and nonlinear equations. Classics in Applied Mathematics, 16. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1996.

Gill, Philip E.; Murray, Walter; Wright, Margaret H. Practical optimization. Academic Press, Inc. London-New York, 1981.