3.2 Méthodes numériques de minimisation

On désire calculer de manière numérique le minimum (ou les minima) d'une fonction $f : U \to \mathbb{R}$ définie sur un ouvert de $\mathbb{R}^n$

$$f(x) = \min_{y \in U} f(y)$$

Tout algorithme de minimisation devrait avoir les propriétés

1) $$f(x_{k+1}) \leq f(x_k)$$

2) $$x_k \xrightarrow[k \to \infty]{} x = \text{le pt. minimisant}$$

Les méthodes sont regroupées en deux familles

a) les méthodes de descente

$$\begin{cases} x_0 \in \mathbb{R}^n \text{ donné} \\ x_{k+1} = x_k + \rho_k d_k \end{cases}$$

où d_k est une direction de descente et $\rho_k > 0$ est choisi de sorte que $f(x_{k+1}) \leq f(x_k)$

b) les méthodes de type Newton

On résout le système non linéaire

$$\nabla f(x) = 0.$$

3.3 Méthodes de gradient

Considérons un algorithme de descente et utilisons Taylor pour développer $f(x_{k+1})$ autour du pt. x_k.

$$f(x_{k+1}) = f(x_k + \rho_k d_k)$$

$$= f(x_k) + \rho_k (\nabla f(x_k), d_k) + \frac{1}{2}\rho_k^2 (H d_k, d_k)$$

$$\cong f(x_k) + \rho_k (\nabla f(x_k), d_k)$$

(si on néglige le terme d'ordre 2)

Une façon d'assurer l'inégalité $f(x_{k+1}) \le f(x_k)$ est de choisir

$$d_k = -\nabla f(x_k)$$

car

$$f(x_{k+1}) \cong f(x_k) - \rho_k \|\nabla f(x_k)\|^2 \le f(x_k)$$

Ceci motive l'algorithme du gradient

$$\begin{cases} x_0 \text{ donné et } \rho_k > 0 \\ x_{k+1} = x_k - \rho_k \nabla f(x_k) \end{cases}$$

On dit que l'algo. est à pas constant si $\rho_k = \rho$.

Quelques notes sur le critère d'arrêt.

On peut utiliser un (ou une combinaison) des critères suivants pour arrêter les calculs.

1) Etant donné que $\nabla f(x_k) \to \nabla f(x) = 0$

$$\left| \frac{\partial f}{\partial x_i}(x_k) \right| < \epsilon \qquad \forall i = 1, \dots, n$$

2) On a que $x_k \xrightarrow[k \to \infty]{} x$

$$\| x_{k+1} - x_k \| < \epsilon$$

3) La suite $f(x_k) \to f(x)$

$$| f(x_{k+1}) - f(x_k) | < \epsilon$$

Exemple:

$$f(x) = \frac{1}{2}(Ax, x) - (b, x)$$

$$\nabla f(x) = Ax - b$$

Alg. du gradient:

$$\begin{cases} x_0 \text{ donné} \\ x_{k+1} = x_k + \rho_k (b - Ax_k) \end{cases}$$

Voici un résultat de convergence de l'algorithme du gradient.

Théorème: Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction de classe C^2, coercive, convexe, et admettant un seul minimum.

De plus, on suppose qu'il existe $M > 0$ t.q.

$$(H_f(x)\, v, v) \leq M \|v\|^2 \qquad \forall v \in \mathbb{R}^n$$

Si on choisit les ρ_k dans l'intervalle

$$0 < \beta_1 \leq \rho_k \leq \beta_2 < 2/M$$

alors la méthode du gradient converge.

Preuve: On utilise Taylor autour de x_K

$$x_{K+1} = x_K - \rho_K \nabla f(x_K)$$

$$f(x_{K+1}) = f(x_K) + (\nabla f(x_K), x_{K+1} - x_K)$$
$$+ \frac{1}{2} (H (x_{K+1} - x_K), x_{K+1} - x_K)$$

$\Rightarrow$

$$f(x_{K+1}) - f(x_K) = -\frac{1}{\rho_K} \| x_{K+1} - x_K \|^2 + \frac{1}{2} (H(x_{K+1} - x_K), x_{K+1} - x_K)$$

$$\leq -\frac{1}{\rho_K} \| x_{K+1} - x_K \|^2 + \frac{M}{2} \| x_{K+1} - x_K \|^2$$

(*) $$\leq \left(\frac{M}{2} - \frac{1}{\rho_K} \right) \| x_{K+1} - x_K \|^2 < 0$$

car $$\rho_K < 2/M \Rightarrow \frac{M}{2} - \frac{1}{\rho_K} < 0$$

Par conséquent, la suite $\{ f(x_K) \}$ est strictement décroissante et minorée par le minimum

$$f(x) = \min_{y \in V} f(y) \leq f(x_K) \quad \forall K$$

Ceci signifie que la suite $f(x_K)$ est convergente.

Ceci indique que la suite $\{ x_K \}$ est bornée car sinon il existerait une sous-suite $\| x_{p_K} \| \to \infty$ et par coercivité de f

$$\Rightarrow \quad f(x_{p_K}) \to \infty$$

ce qui contredit le fait que la suite $f(x_K)$ est convergente.

Maintenant, on utilise le critère de Bolzano-Weierstrass pour conclure qu'il existe une sous-suite x_{p_K} qui converge (de la suite originale x_K)

$$x_{p_K} \xrightarrow[K \to \infty]{} x^*$$

f est de classe C^1

$$\Rightarrow \quad f(x_{p_k}) \to f(x^*)$$
$$\nabla f(x_{p_k}) \to \nabla f(x^*)$$

Mais $\nabla f(x_k) = \dfrac{x_{k+1} - x_k}{-\rho_k} \underset{k\to\infty}{\longrightarrow} 0$

à cause de (*)

Et ~~$\nabla f(x_{p_k}) \to x_p$~~

$\therefore \quad \nabla f(x^*) = 0$

Par le fait que f soit convexe et admet un seul minimum

$\Rightarrow \quad x^* = x$ = le pt. minimisant de f

Il reste à montrer que toute la suite $x_k \underset{k\to\infty}{\longrightarrow} x$

Par le même argument ci-dessus, on montre que toutes les sous-suites ~~qui convergent~~ convergeantes de x_k convergent vers la même limite $= x$.

En utilisant le fait que la suite originale x_k est bornée, ceci est suffisant pour conclure que l'algo. converge

$$x_k \underset{k\to\infty}{\longrightarrow} x$$

Voici une situation intéressante où le théorème précédent s'applique.

$$f(x) = \frac{1}{2}(Ax, x) - (b, x) \qquad \text{où } A > 0 \text{ et } b \in \mathbb{R}^n$$

En effet, f est coercive, convexe, de classe C^2. De plus, on a que

$$H_f(x) = A.$$

La fonction

$$R_A(v) = \frac{(Av, v)}{(v, v)}$$

est continue dans $\mathbb{R}^n \setminus \{0\}$ et atteint son maximum car

$$\max_{v \neq 0} R_A(v) = \max_{\|v\|=1} R_A(v)$$

et que $\{v \in \mathbb{R}^n / \|v\| = 1\}$ est un ensemble fermé et borné dans $\mathbb{R}^n$.

Par conséquent, $\exists M > 0$ t.q.

$$\frac{(Av, v)}{(v, v)} \leq M \qquad \forall v \neq 0$$

$$\text{ou} \qquad (Av, v) \leq M \|v\|^2 \quad \forall v \in \mathbb{R}^n$$

Par conséquent, l'algo. du gradient converge pour le calcul du minimum de la fonction quadratique $f(x)$.

Dans la pratique, l'algo. du gradient à pas constant ($\rho_k = \rho$) est lent au voisinage du pt. minimisant. En effet

$$\nabla f(x_k) \simeq 0$$

$$\Rightarrow \rho \nabla f(x_k) \simeq 0$$

$$\Rightarrow x_{k+1} = x_k - \rho \nabla f(x_k) \simeq x_k$$

Pour remédier à cette situation, on fait appel à la méthode du gradient à pas optimaux

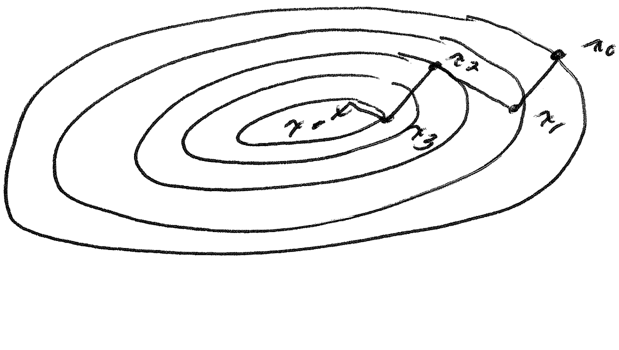
$$\begin{cases} x_0 \text{ donné} \\ x_{k+1} = x_k - \rho_k \nabla f(x_k) \end{cases}$$

où $\rho_k > 0$ est choisi de sorte que

$$\min_{\rho} f(x_k - \rho \nabla f(x_k))$$

Par le choix de ρ_k, l'algo. du gradient à pas optimaux possède la propriété particulière

$$\nabla f(x_{k+1}) \perp \nabla f(x_k)$$



En général, il n'est pas facile de calculer le ρ_k optimal. Par contre, dans le cas quadratique, la situation est facile.

Posons $$f(x) = \frac{1}{2}(Ax, x) - (b, x) \qquad A > 0,\ b \in \mathbb{R}^N$$

On définit le résidu au pt x par

$$r = -Df(x) = b - Ax$$

L'algo. du gradient s'écrit :

$$x_{k+1} = x_k - \rho_k\, Df(x_k) = x_k + \rho_k r_k$$

$$\Rightarrow \quad Ax_{k+1} - b = Ax_k - b + \rho_k A r_k$$

$$\Rightarrow \quad r_{k+1} = r_k - \rho_k A r_k$$

La condition d'optimalité pour ρ_k est :

$$Df(x_{k+1}) \perp Df(x_k)$$

$$\Leftrightarrow \quad (r_{k+1}, r_k) = 0$$

$$\Leftrightarrow \quad (r_k - \rho_k A r_k, r_k) = 0$$

$$\Rightarrow \quad \rho_k = \frac{\|r_k\|^2}{(A r_k, r_k)}$$

Remarque :

$$(A r_k, r_k) = 0 \quad \Leftrightarrow \quad r_k = 0 \quad \text{car } A > 0$$

et $$r_k = b - Ax_k = 0 \quad \Leftrightarrow \quad Ax_k = b$$

$$x_k = x$$

" Nous avons atteint le minimum ".