Propriété : indépendance linéaire

Si $Q \in \mathbb{R}^{n\times n}$ est symétrique définie positive et $d_1, \dots, d_k \in \mathbb{R}^n$ sont Q-conjugués, alors ils sont linéairement indépendants

Preuve : par contraposition, supposons $d_k = \alpha_1 d_1 + \dots + \alpha_{k-1} d_{k-1}$ avec $(\alpha_1, \dots, \alpha_{k-1}) \neq 0$

Alors $0 < d_k^T Q d_k$ (Q définie positive)

$< d_k^T Q (\alpha_1 d_1 + \dots + \alpha_{k-1} d_{k-1}) = \alpha_1 \underbrace{d_k^T Q d_1}_{=0} + \dots + \alpha_{k-1} \underbrace{d_1^T Q d_{k-1}}_{=0}$

$\to$ contradiction

On en revient à la minimisation de $\frac{1}{2} x^T A x - b^T x$:

$x^* = \underset{x \in \mathbb{R}^n}{\text{argmin}} \frac{1}{2} x^T A x - b^T x \iff \nabla f(x^*) = 0$

$\iff Ax^* - b = 0$

$\iff x^*$ est solution du système linéaire $Ax = b$: $Ax^* - b = 0 \iff Ax^* = b$

La propriété d'indépendance linéaire nous dit que si $d_0, \dots, d_{n-1}$ sont des vecteurs A-conjugués, ils forment une base de $\mathbb{R}^n$ (car famille libre de n vecteurs)

On peut donc écrire $x^* = \alpha_0 d_0 + \dots + \alpha_{n-1} d_{n-1} = \sum_{i=0}^{n-1} \alpha_i d_i$ pour des poids α_i

$\to d_i^T A x^* = d_i^T A (\alpha_0 d_0 + \dots + \alpha_{n-1} d_{n-1}) = \alpha_0 d_i^T A d_0 + \dots + \alpha_{n-1} d_i^T A d_{n-1}$

$= \alpha_i d_i^T A d_i$ puisque $d_i^T A d_j = 0$ (d_i et d_j sont A-conjugués)

$\to \alpha_i = \dfrac{d_i^T A x^*}{d_i^T A d_i} = \dfrac{d_i^T b}{d_i^T A d_i}$ puisque $Ax^* = b$

$\to$ si on connait les directions d_i, on peut trouver les poids $\alpha_i = \dfrac{d_i^T b}{d_i^T A d_i}$ associés sans avoir besoin de connaitre x^* ni d'inverser A

Et $x^* = \sum_{i=0}^{n-1} \alpha_i d_i = \sum_{i=0}^{n-1} \left(\dfrac{d_i^T b}{d_i^T A d_i}\right) d_i = \dfrac{d_0^T b}{d_0^T A d_0} d_0 + \dots + \dfrac{d_{n-1}^T b}{d_{n-1}^T A d_{n-1}} d_{n-1}$

$\to x^*$ peut se construire itérativement en n étapes, en rajoutant à chaque fois $\alpha_i d_i$ à l'itération précédente si on connait d_i (auquel cas $\alpha_i = \dfrac{d_i^T b}{d_i^T A d_i}$)

Théorème des directions conjuguées

Soient $d_0, \dots, d_{n-1} \in \mathbb{R}^n$ des vecteurs A-conjugués et $x_0 \in \mathbb{R}^n$. Alors les itérations suivantes :

$\to g_k = \nabla f(x_k) = Ax_k - b$

$\to \alpha_k = -\dfrac{g_k^T d_k}{d_k^T A d_k}$ (pas optimal de la descente de gradient)

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

convergent en n itérations : $x_n = x^*$ solution de $Ax^* = b$ ($\iff \nabla f(x^*) = 0$)

$\Rightarrow$ On ne sait toujours pas comment trouver d_k en pratique (pour le moment), mais la convergence est garantie en n itérations peu importe $x_0 \in \mathbb{R}^n$!

Choix de la direction de descente : la direction d_k est construite à partir du gradient $g_k = \nabla f(x_k)$ et de la direction précédente d_{k-1}

$$d_k = \underbrace{-g_k}_{\text{descente de plus forte pente}} + \underbrace{\beta_k d_{k-1}}_{\text{"momentum"}}$$ avec β_k choisi de manière à ce que d_k et d_{k-1} soient A-conjugués

→ β_k est donné par $d_{k-1}^T A d_k = 0 = d_{k-1}^T A(-g_k + \beta_k d_{k-1})$

$$\Leftrightarrow \ -d_{k-1}^T A g_k + \beta_k d_{k-1}^T A d_{k-1} = 0$$

$$\Leftrightarrow \ \beta_k = \frac{d_{k-1}^T A g_k}{d_{k-1}^T A d_{k-1}}$$

Une fois la nouvelle direction d_k obtenue, on construit le prochain itéré avec $x_{k+1} = x_k + \alpha_k d_k$ (la première direction d_0 est donnée par $d_0 = -g_0 = -\nabla f(x_0)$)

On obtient ainsi la version finale de l'algorithme du gradient conjugué

Algorithme du gradient conjugué (cas quadratique)

Initialisation : $x_0 \in \mathbb{R}^n$; $g_0 = \nabla f(x_0) = Ax_0 - b$; $d_0 = -g_0$; $\alpha_0 = -\dfrac{g_0^T d_0}{d_0^T A d_0}$

$x_1 = x_0 + \alpha_0 d_0$

Boucle : pour $k = 1, \dots, n-1$

$g_k = \nabla f(x_k) = Ax_k - b$ — gradient au point actuel

$\beta_k = \dfrac{d_{k-1}^T A g_k}{d_{k-1}^T A d_{k-1}}$ — coefficient pour que la nouvelle direction de descente soit A-conjuguée à la précédente

$d_k = -g_k + \beta_k d_{k-1}$ — nouvelle direction de descente (avec momentum)

$\alpha_k = -\dfrac{d_k^T g_k}{d_k^T A d_k}$ — pas de descente optimal

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

Avec quelques manipulations supplémentaires, on peut réécrire les poids α_k et β_k par les formules suivantes (utilisées en pratique)

$$\alpha_k = \frac{g_k^T g_k}{d_k^T A d_k} \quad \text{et} \quad \beta_k = \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}}$$

→ l'algorithme a convergé lorsque $x_k = x_n$, c'est à dire $\boxed{x_n = x^*}$

Si f n'est plus quadratique : on peut toujours utiliser le gradient conjugué, mais :

→ il n'y a plus la garantie que ça converge en n itérations. Dans ce cas, on relance n itérations supplémentaires en prenant comme x_0 le point terminal x_n des n itérations précédentes

→ même si la mise à jour $x_{k+1} = x_k + \alpha_k d_k$ reste la même, il n'est plus possible d'utiliser $\alpha_k = \dfrac{g_k^T g_k}{d_k^T A d_k}$ (formule du pas optimal dans le cas quadratique)

→ α_k est déterminé par la règle d'Armijo

→ la direction conjuguée reste la même : $d_k = -g_k + \beta_k d_{k-1}$, avec deux formules possibles pour β_k