Projet n°3 : Les méthodes de quasi-Newton

L'idée de ces méthodes est de proposer une alternative à la méthode (coûteuse) de Newton pour la recherche du minimum d'une fonction J. Rappelons que l'algorithme de Newton s'écrit :

$$u^0 \in \mathbb{R}^n, \quad u^{k+1} = u^k - [J''(u^k)]^{-1} \nabla J(u^k).$$

Une des difficultés de cette méthode est le coût du calcul de la hessienne $J''(u^k)$ et de son inverse. On introduit donc des algorithmes du type

$$u^0 \in \mathbb{R}^n, \quad u^{k+1} = u^k - \alpha^k S^k \nabla J(u^k),$$

où

- α^k est le pas optimal associé à la direction $S^k \nabla J(u^k)$, i.e.

$$J(u^{k+1}) = \inf_{\alpha \in \mathbb{R}} J(u^k - \alpha S^k \nabla J(u^k)),$$

- S^k est une matrice symétrique définie positive, raisonnablement proche de $[J''(u^k)]^{-1}$, qu'on peut calculer à l'aide d'une formule simple du type

$$S^{k+1} = S^k + C^k.$$

Dans cette formule, C^k est une matrice de correction qui sera choisie de sorte que S^{k+1} satisfasse la condition ci-dessous, dite "condition de quasi-Newton" :

$$S^{k+1} \left(\nabla J(u^{k+1}) - \nabla J(u^k) \right) = u^{k+1} - u^k. \tag{1}$$

On pourra remarquer qu'en dimension $n = 1$, cette condition définit entièrement le réel S^{k+1}, et qu'on obtient la méthode de la sécante. Cependant, en dimension supérieure ou égale à 2, il y a plusieurs choix possibles pour la matrice S^{k+1}. On s'intéressera particulièrement au choix de Davidon, Fletcher et Powell. D'autres méthodes de quasi-Newton sont présentées dans [1]

Dans tout le devoir, on note, pour tout $k \geq 1$,

$$\delta^k = u^{k+1} - u^k, \quad \gamma^k = \nabla J(u^{k+1}) - \nabla J(u^k),$$

et Δ, respectivement Γ, les matrices de colonnes $(\delta_0, \cdots \delta^{n-1})$, $(\gamma^0, \cdots \gamma^{n-1})$.

Suggestions pour l'étude du problème

Les suggestions ci-dessous ne sont que des pistes de réflexion, il n'est pas obligatoire de traiter toutes les questions, et d'autres aspects pourront être abordés. Cependant, le travail comportera obligatoirement une partie de programmation, dans un langage laissé au libre choix du binôme.

La méthode Davidon-Fletcher-Powell (DFP)

On définit l'algorithme D.F.P. de la façon suivante :
– on choisit un point u^0 et une matrice S^0 symétrique définie positive ;
– à l'étape $k \geq 0$, on définit α^k par

$$J(u^k - \alpha^k S^k \nabla J(u^k)) = \inf_{\alpha \in \mathbb{R}} J(u^k - \alpha S^k \nabla J(u^k)).$$

On calcule alors

$$u^{k+1} = u^k - \alpha^k S^k \nabla J(u^k), \quad S^{k+1} = S^k + \frac{(\delta^k)(\delta^k)^T}{(\delta^k)^T \gamma^k} - \frac{(S^k \gamma^k)(S^k \gamma^k)^T}{(\gamma^k)^T S^k \gamma^k}.$$

1. Montrer qu'à chaque étape, la matrice S^k est symétrique définie positive (en supposant que le α^k soit calculé exactement...) et vérifie la relation (1).
2. On suppose dans cette question que J est quadratique, de hessienne A, symétrique, définie positive.
   (a) Montrer par récurrence que
   $$\forall i, j,\ 0 \leq i < j \leq k,\ \langle \delta^i, A\delta^j \rangle = 0,$$
   i.e. les directions de descente sont A-conjuguées deux à deux.
   (b) On fixe $i \geq 0$; montrer par récurrence sur k que
   $$\forall k \geq i,\ S^{k+1}\gamma^i = \delta^i.$$
   En déduire que la méthode converge toujours en n itérations au plus, i.e. on a toujours $S^n = A^{-1}$.
   (c) Vérifier que si $S^0 = I$, on retrouve la méthode du gradient conjugué.
3. Tester cette méthode pour le problème de Rosenbrock, i.e.
   $$J(v_1, v_2) = (v_1 - 1)^2 + 10(v_1^2 - v_2)^2.$$
   On prendra $u^0 = (-1, -1)$ et $S^0 = J''(u^0)^{-1}$. Comparer avec les méthodes vues en cours.

En pratique, on ne peut bien-sûr pas calculer les α^k de façon exacte ; or, cette méthode est assez sensible à la précision du calcul de α^k. Pour cette raison, on lui préfère souvent celle de Broyden-Fletcher-Goldfarb-Shanno (BFGS), voir par exemple [1].

Références

[1] Jean-Baptiste Culioli, Introduction à l'optimisation. Paris : Ellipses, 1994.