Annexe B

Méthode du gradient à pas optimal

La méthode la plus classique pour minimiser une fonction de plusieurs variables $f(p)$ (où p est un vecteur) est la méthode de la plus grande pente (ou méthode du gradient à pas optimal). Cette méthode est basée sur la constatation élémentaire suivante : soit p_0 tel que $\nabla f(p_0) \neq 0$. On peut remplacer f par son approximation au premier ordre au voisinage de p_0. L'approximation de f est la fonction L définie par :

$$L(p) = f(p_0) + \nabla f(p_0)^T (p - p_0).$$

Considérons la droite passant par p_0 le long du gradient de f en p_0 :

$$p(\alpha) = p_0 - \alpha \cdot \nabla f(p_0).$$

Si α est choisi positif, alors

$$L(p(\alpha)) = f(p_0) - \alpha \parallel \nabla f(p_0) \parallel^2 < f(p_0).$$

Comme

$$f(p(\alpha)) = L(p(\alpha)) + o(\alpha \parallel \nabla f(p_0) \parallel),$$

pour α suffisamment petit, on aura

$$f(p(\alpha)) < f(p_0).$$

En d'autres termes, la direction opposée à celle du gradient est une direction de descente. C'est même la meilleure localement (i.e. pour α petit).

L'algorithme de la plus grande pente (voir l'algorithme 3, page 105), basé sur ce principe, a les propriétés suivantes :

- Deux directions de recherche successives sont orthogonales. En effet, dans le calcul de α_k, si la minimisation est exacte (i.e. si l'on trouve exactement le meilleur α), la condition nécessaire pour l'optimum de la recherche linéaire s'écrit
$$\frac{d}{d\alpha} f(p_k - \alpha \nabla f(p_k))|_{\alpha=\alpha_k} = -\nabla f(p_{k+1})^T \nabla f(p_k) = 0. \tag{B.1}$$
- Dans le cas où f est quadratique, de la forme $f(p) = \frac{1}{2} p^T M p - g^T p$, avec M symétrique définie positive, on peut calculer facilement le paramètre α_k. Il est donné par l'équation
$$\nabla f(p_{k+1})^T \nabla f(p_k) = 0$$
ou encore
$$(Mp_{k+1} - g)^T (Mp_k - g) = \Big(M\big(p_k - \alpha_k (Mp_k - g)\big)\Big)^T (Mp_k - g) = 0.$$
Soit, à l'aide des notations
$$\| p \|^2 = p^T p$$
et
$$\| p \|_M^2 := p^T M p, \tag{B.2}$$
(si M est définie positive, il s'agit bien d'une norme) :
$$\alpha_k = \frac{\| Mp_k - g \|^2}{\| Mp_k - g \|_M^2}. \tag{B.3}$$

Algorithme 3 Algorithme de la plus grande pente (ou à pas optimal)

1: choix du vecteur des poids p_0 et de la tolérance de convergence ϵ.

2: en partant de p_o, on peut au prix d'une recherche linéaire, déterminer

$$p_1 = p_0 - \alpha_0 \nabla f(p_0)$$

avec

$$\alpha_0 = min_{\alpha>0} f(p_0 - \alpha \nabla f(p_0)).$$

Le point p_1 obtenu est le meilleur possible dans la direction $-\nabla f(p_0)$. C'est ce que l'on résume par "plus grande pente" ou "gradient à pas optimal", α_0 étant le pas optimal correspondant.

3: **répéter**

4: pour $k = 1, 2, 3, ...$, calcul de p_{k+1} connaissant p_k

$$p_{k+1} = p_k - \alpha_k \nabla f(p_k)$$

avec

$$\alpha_k = min_{\alpha>0} f(p_k - \alpha \nabla f(p_k)).$$

5: **jusqu'à ce que** $\| \nabla f(p_k) \| < \epsilon$.

6: **FIN**