Plan

Notes

(Rappels) Direction de descente : définition

Soit $f : \mathbb{R}^n \to \mathbb{R}$. Une direction d est une direction de descente en un point x si :

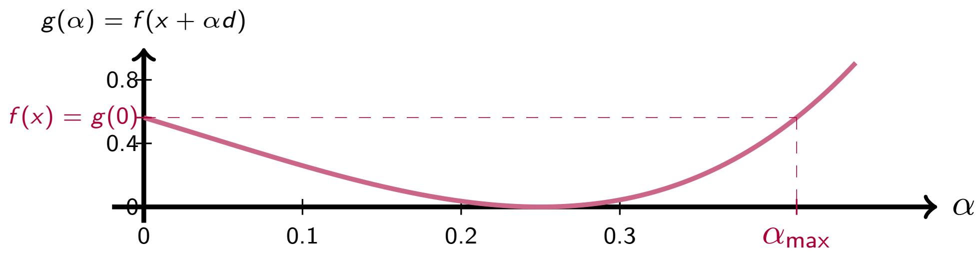
$$d^T \nabla f(x) < 0,$$

autrement dit, lorsque la dérivée directionnelle de f au point x dans la direction d est négative. Cela signifie que la fonction est décroissante dans la direction d.

Théorème.
Soit $f : \mathbb{R}^n \to \mathbb{R}$. Si d est une direction de descente en un point x, alors $\exists \alpha_{\max} > 0$:

$$f(x + \alpha d) < f(x), \quad \forall \alpha \in \left]0, \alpha_{\max}\right].$$

Illustration.
Pour la fonction $f(x_1, x_2) = (x_1 + x_2)^4 - 2(x_1 + x_2)^2 + 1$ au point $x = (\frac{1}{4}, \frac{1}{4})$ dans la direction $d = (1, 1)^T$, on a $f(x + \alpha d) = ((0.5 + 2\alpha)^2 - 1)^2$.



Dans cet exemple, $\alpha_{\max} = 0.411438$.

Notes

Recherche en ligne (unidimensionnelle) (*line search*)

Un grand nombre de méthodes pour l'optimisation sans contraintes,

$$\min_{x \in \mathbb{R}^n} f(x)$$

où f est suffisamment diffférentiable mais pas nécessairement convexe, sont basées sur une stratégie de recherche en ligne :

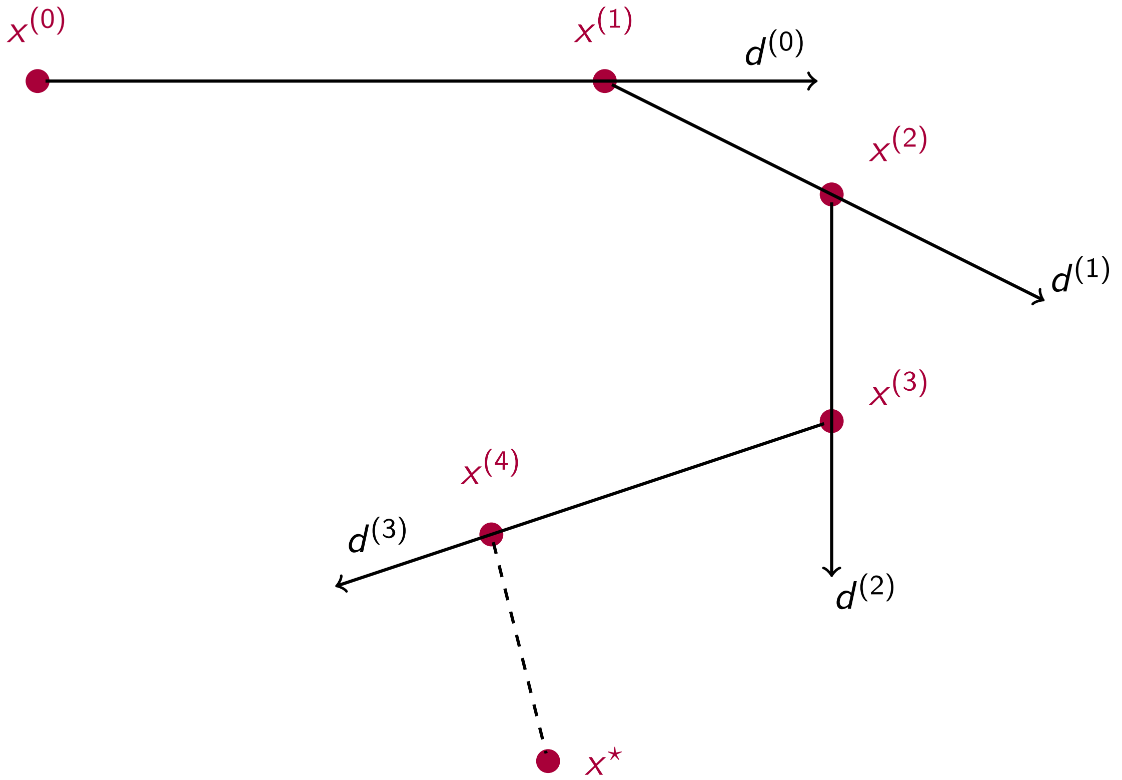
1. Initialisation : un point de départ $x^{(0)} \in \mathbb{R}^n$ et un indice $k := 0$
2. Tant qu'on n'est pas suffisamment proche d'un minimum
   1. Trouver une direction de recherche $d^{(k)} \in \mathbb{R}^n$
   2. Trouver une longueur de pas $\alpha^{(k)}$ qui minimise

      $$\min_{\alpha > 0} g(\alpha) = f(x^{(k)} + \alpha d^{(k)})$$

   Mettre à jour $\boldsymbol{x^{(k+1)} := x^{(k)} + \alpha^{(k)} d^{(k)}}$ puis $k := k + 1$

Ces méthodes n'ont pour vocation que d'identifier un minimum local (à moins d'avoir affaire à une problème d'optimisation convexe)

Notes



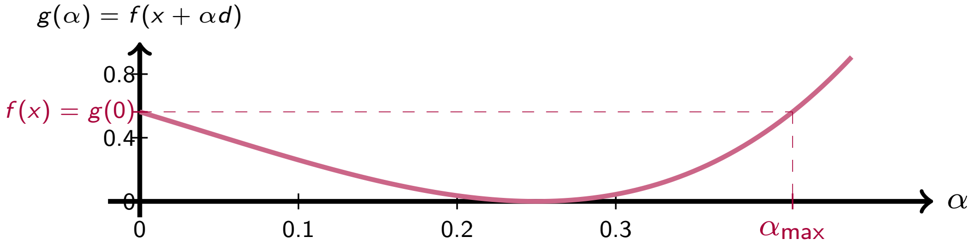
Lorsqu'on se trouve au point $x^{(k)}$,

- **quelle direction $d^{(k)}$ choisir ?** une direction de descente! (prochains cours)
- **quel pas $\alpha^{(k)}$ choisir le long de $d^{(k)}$?**

Notes

Choix du pas α

En l'itéré x et étant donné une direction de descente d, comment trouver le pas α ?



- Calculer la longueur de pas qui minimise exactement $g(\alpha)$ est possible dans certains cas, par exemple quand $f = \frac{1}{2}x^T Ax - b^T x + c$:

$$\alpha^\star = \frac{-{d^{(k)}}^T(Ax^{(k)} - b)}{{d^{(k)}}^T Ad^{(k)}}$$

- Dans la plupart des cas, calculer la longueur exacte du pas est couteux (problème d'optimisation en tant que tel) et inutile si la direction n'est pas bonne.
- Des petits pas devraient être utilisés afin que l'approximation linéaire de la fonction reste valable. Cependant, on désire avancer rapidement et faire des pas plus longs.
- Afin de concilier ces deux objectifs contradictoires, on introduit des conditions à respecter : les conditions de Wolfe. La longueur du pas n'est donc pas optimale mais est plus facile à calculer.

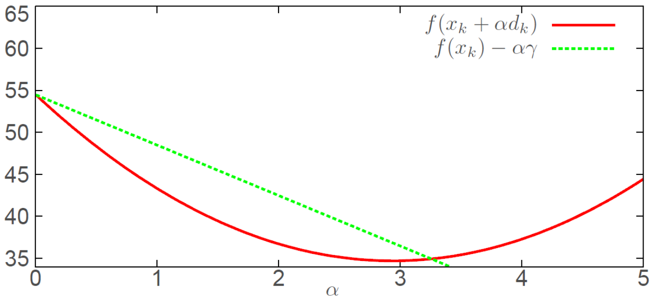
Notes

Condition de décroissance suffisante

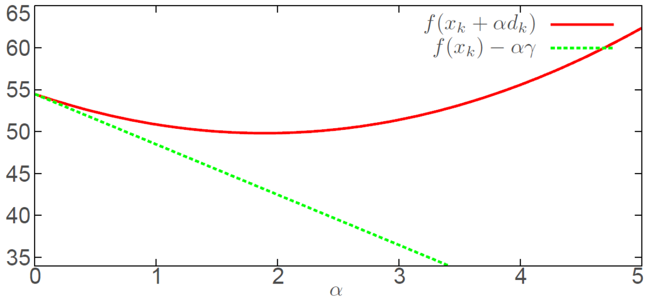
Etant donnés un point itéré x et une direction d, ou souhaite que le pas α vérifie une certaine décroissance (imposée à l'aide d'un paramètre $\gamma > 0$) :

$$f(x + \alpha d) \leq f(x) - \alpha\gamma.$$

Exemple : $f(x) = \frac{1}{2}x_1^2 + \frac{9}{2}x_2^2$ en l'itéré $(10, 1)$ et $\gamma = 6$



pour $d_1 = \begin{pmatrix} \frac{-10}{\sqrt{181}} & \frac{-9}{\sqrt{181}} \end{pmatrix}^T$



pour $d_2 = \begin{pmatrix} \frac{-2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix}^T$

Observations :

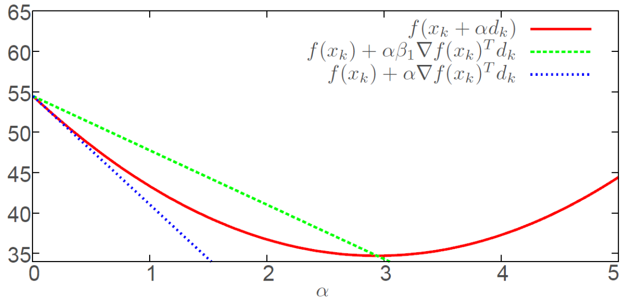
- γ ne peut pas être constant,
- il doit dépendre de la direction.

Notes

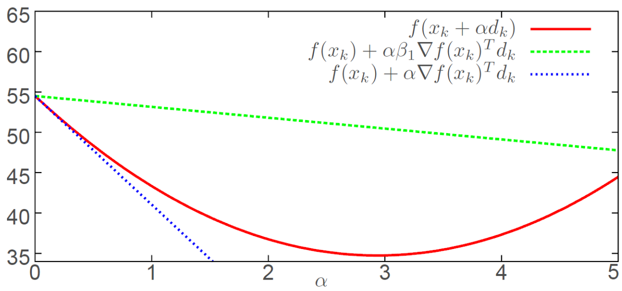
Condition de décroissance suffisante

Il est possible de montrer que pour tout $\beta_1 < 1$, il existe $\alpha \in]0, \alpha_{\max}]$ tel que :

$$f(x + \alpha d) < f(x) + \alpha \beta_1 d^T \nabla f(x).$$



d_1 et $\beta_1 = 0.5$



d_1 et $\beta_1 = 0.1$

On choisit donc, avec $0 < \beta_1 < 1$ (typiquement $\beta_1 = 10^{-4}$) :

$$\gamma = -\beta_1 d^T \nabla f(x).$$

Première condition de Wolfe (W1). (également appelée condition d'Armijo-Wolfe)
La fonction f diminue suffisamment en $x^{(k)} + \alpha^{(k)} d^{(k)}$ par rapport à $x^{(k)}$ si, avec $0 < \beta_1 < 1$:

$$f(x^{(k)} + \alpha^{(k)} d^{(k)}) \leq f(x^{(k)}) + \alpha^{(k)} \beta_1 {d^{(k)}}^T \nabla f(x^{(k)}).$$

Inconvé

cette condition est toujours satisfaite lorsque $\alpha \to 0$, et n'empêche pas les pas de devenir trop petits $\to$ il est nécessaire d'introduire une condition supplémentaire.

Notes

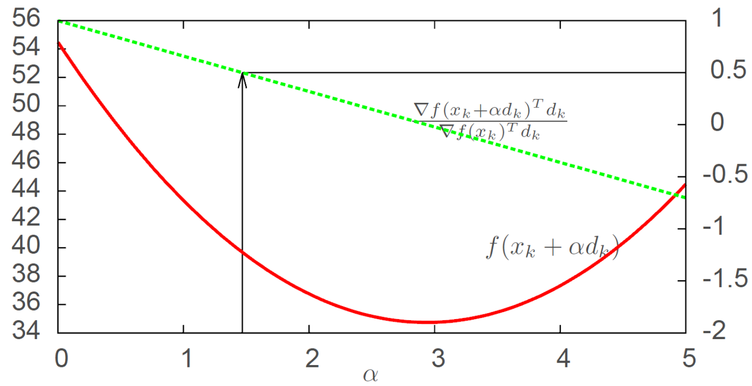
Condition de progrès suffisant (condition de courbure)

Au point $x^{(k)}$, puisque $d^{(k)}$ est une direction de descente, on a bien ${d^{(k)}}^T \nabla f(x^{(k)}) < 0$.
Si on applique le pas optimal dans la direction d_k, on a $g'(\alpha) = 0$, ce qui équivalent à :

$$ {d^{(k)}}^T \nabla f(x^{(k)} + \alpha^{(k)} d^{(k)}) = 0. $$

On assiste donc à une **augmentation** de la dérivée directionnelle.
L'idée est donc d'obtenir un pas α tel que cette dérivée augmente suffisamment.

Seconde condition de Wolfe (W2).
Avec $\beta_1 < \beta_2 < 1$ (typiquement $\beta_2 = 0.9$) : ${d^{(k)}}^T \nabla f(x^{(k)} + \alpha^{(k)} d^{(k)}) \geq \beta_2 {d^{(k)}}^T \nabla f(x^{(k)})$.



(en choisissant par exemple $\beta_2 = 0.5$, le pas α doit être tel que $\alpha \geq 1.4687$.)

Condition W2 forte : $\left| {d^{(k)}}^T \nabla f(x^{(k)} + \alpha d^{(k)}) \right| \leq \beta_2 \left| {d^{(k)}}^T \nabla f(x^{(k)}) \right|$.

Notes

Interprétation à l'aide de $g(\alpha)$

On sait que pour $g(\alpha) = f(x^{(k)} + \alpha d^{(k)})$, on a

$$g'(\alpha) = {d^{(k)}}^T \nabla f(x^{(k)} + \alpha d^{(k)}).$$

- La condition de décroissance suffisante $W1$ impose un taux de décroissance au moins égal à $\beta_1 |g'(0)|$:

$$g(\alpha) \leq g(0) - \alpha\beta_1 \big|g'(0)\big|.$$

- La condition de courbure $W2$ borne inférieurement $g'(\alpha)$ (c'est-à-dire l'empêche de rester trop négatif)

$$g'(\alpha) \geq \beta_2 g'(0).$$

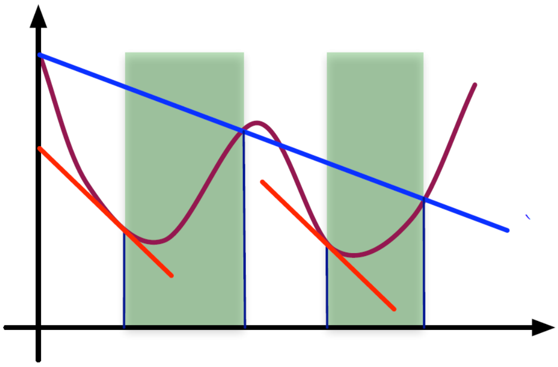
- La condition forte borne supérieurement $|g'(\alpha)|$, c'est-à-dire empêche $g'(\alpha)$ de trop s'éloigner de zéro :

$$|g'(\alpha)| \leq \beta_2 \, |g'(0)| \, .$$

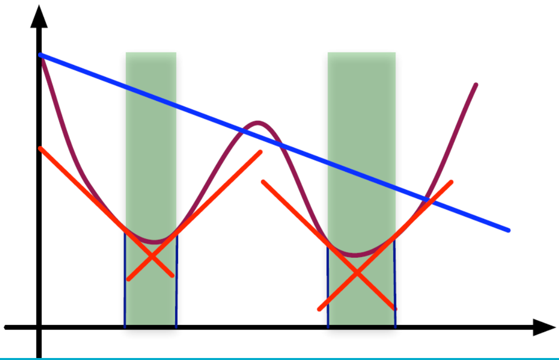
Notes

Conditions : illustration $W1{+}W2$ vs. $W1{+}W2$ forte

Pas α acceptables pour $W1$ (bleue) et $W2$ (rouge) :



Pas α acceptables pour $W1$ (bleue) et $W2$ forte (rouge) :



Notes

Existence

- Si on suppose $f \in C^1$ et
  - $d^{(k)}$ est une direction de descente,
  - $g(\alpha) = f(x^{(k)} + \alpha d^{(k)})$ est bornée inférieurement pour $\alpha > 0$, et
  - $0 < \beta_1 < \beta_2 < 1$,

  alors il existe des intervalles de longueurs de pas vérifiant les conditions de Wolfe (et les conditions fortes de Wolfe)
- Si $0 < \beta_2 < \beta_1 < 1$, il est possible qu'aucune valeur α ne vérifie les conditions de Wolfe.
- Il existe des algorithmes de calcul garantissant l'obtention d'un de ces points (voir slide suivant)
- Le minimum le long de la direction de recherche ne satisfait pas toujours aux conditions de Wolfe (d'où le choix de β_1 très petit)
- Ces conditions sont indépendantes de l'échelle

Notes

Algorithme de calcul de α

Méthode permettant de déterminer un pas α vérifiant les conditions $W1$ et $W2$.

On se donne : α_0, $0 < \beta_1 < \beta_2 < 1$ et $\lambda > 1$

1. Initialisation : $i = 0$, $\alpha_l = 0$ et $\alpha_r = +\infty$.
2. Si α_i vérifie $W1$ et $W2$, alors $\alpha^* = \alpha_i$, STOP.
3. Si α_i viole $W1$, alors le pas est trop long.

$$\alpha_r = \alpha_i$$

$$\alpha_{i+1} = \frac{\alpha_l + \alpha_r}{2}$$

4. Si α_i ne viole pas $W1$ mais viole $W2$, alors le pas est trop court.

$$\alpha_l = \alpha_i$$

$$\alpha_{i+1} = \begin{cases} \frac{\alpha_l+\alpha_r}{2} & \text{si } \alpha_r < +\infty \\ \lambda\alpha_i & \text{sinon.} \end{cases}$$

5. $i = i + 1$

Notes