* $\beta_k = \dfrac{g_k^T g_k}{g_{k-1}^T g_{k-1}} = \dfrac{\|\nabla f(x_k)\|^2}{\|\nabla f(x_{k-1})\|^2}$ méthode de Fletcher-Reeves

* $\beta_k = \dfrac{(g_k - g_{k-1})^T g_k}{g_{k-1}^T g_{k-1}} = \dfrac{(\nabla f(x_k) - \nabla f(x_{k-1}))^T \nabla f(x_k)}{\|\nabla f(x_{k-1})\|^2}$ méthode de Polack-Ribière

Dans le cas quadratique, les deux méthodes sont identiques, mais dans le cas général, la méthode de Polack-Ribière est en général plus performante

7) La méthode de Newton

Méthode de Newton-Raphson

La méthode de Newton-Raphson est à l'origine une méthode itérative permettant de déterminer les zéros d'une fonction $f : \mathbb{R} \to \mathbb{R}$ (donc les solutions de $f(x) = 0$) supposée dérivable

A partir de $x_0 \in \mathbb{R}$ (idéalement assez proche de la solution)

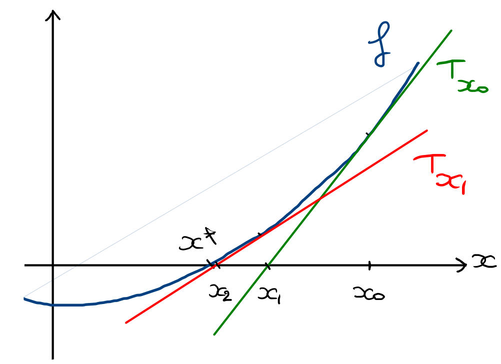
→ linéarisation de f en x_0 via sa tangente $T_{x_0} : x \mapsto f(x_0) + (x - x_0) f'(x_0)$

→ recherche du point x_1 tel que $T_{x_0}(x_1) = 0 = f(x_0) + (x_1 - x_0) f'(x_0)$

$$\iff x_1 - x_0 = -\frac{f(x_0)}{f'(x_0)}$$

$$\iff x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$$

→ on recommence en x_1



La méthode de Newton Raphson construit donc la suite d'itérés suivante : $x_0 \in D_f$; $x_{k+1} = x_k - \dfrac{f(x_k)}{f'(x_k)}$

La convergence échoue si $x_{k+1} \notin D_f$ ou si $f'(x_k) = 0$

Sinon, il existe un voisinage de x^* (solution de $f(x) = 0$) tel que la convergence est quadratique (modulo quelques hypothèses supplémentaires ...)

Méthode de Newton en optimisation

Pour $f : \mathbb{R} \to \mathbb{R}$ dans un premier temps, en supposant f convexe et 2-fois dérivable : $x^* = \underset{x}{\operatorname{argmin}} f(x) \iff f'(x^*) = 0$

→ on peut donc appliquer Newton Raphson sur f' pour en trouver un zéro

→ la méthode de Newton s'écrit donc : $x_0 \in D_f$; $\boxed{x_{k+1} = x_k - \dfrac{f'(x_k)}{f''(x_k)}}$

Cette règle peut s'interpréter comme la minimisation de l'approximation quadratique de f en x_k :

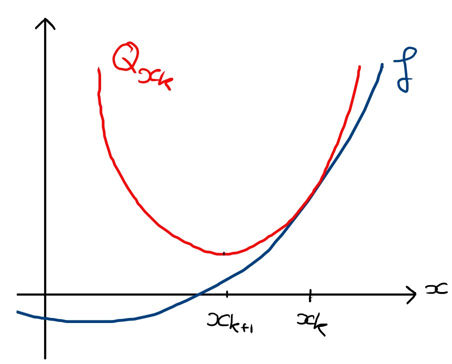
DL_2 de f en x_k : $f(x_k + h) = f(x_k) + h f'(x_k) + \dfrac{h^2}{2} f''(x_k) + o(h^2)$

$$\iff f(x) = f(x_k) + (x - x_k) f'(x_k) + \frac{(x - x_k)^2}{2} f''(x_k) + o((x - x_k)^2)$$

Soit $Q_{x_k} : x \mapsto f(x_k) + (x - x_k) f'(x_k) + \dfrac{(x - x_k)^2}{2} f''(x_k)$ l'approximation quadratique de f en x_k (parabole de même pente $f'(x_k)$ et de même courbure $f''(x_k)$)

$f(x) = Q_{x_k}(x) + o((x - x_k)^2)$ pour x proche de x_k

→ x_{k+1} est définit comme $\underline{x_{k+1} = \operatorname{argmin} Q_{x_k}(x)}$



$\rightarrow$ f convexe, donc $f''(x_k) > 0$

$\rightarrow$ $x_{k+1} = \underset{x}{\operatorname{argmin}}\ Q_{x_k}(x) \iff Q'_{x_k}(x_{k+1}) = 0$

Or $Q'_{x_k}(x) = f'(x_k) + (x - x_k) f''(x_k)$ (puisque $Q_{x_k}(x) = f(x_k) + (x-x_k)f'(x_k) + \frac{(x-x_k)^2}{2} f''(x_k)$)

Donc $Q'_{x_k}(x_{k+1}) = 0 = f'(x_k) + (x_{k+1} - x_k) f''(x_k)$

$$\iff \boxed{x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}}$$

Pour $f : \mathbb{R}^n \rightarrow \mathbb{R}$ maintenant, à supposer que f est convexe et 2-fois différentiable, de hessienne H_f

L'idée reste la même, à savoir pour un point $x_k \in \mathbb{R}^n$ donné, on modélise f par son approximation quadratique Q_{x_k} de même gradient $\nabla f(x_k)$ et même matrice hessienne $H_f(x_k)$, et l'itéré suivant est donné par le minimiseur de Q_{x_k} :

$x_{k+1} = \underset{x}{\operatorname{argmin}}\ Q_{x_k}(x)$

Avec $Q_{x_k}(x) = f(x_k) + \nabla f(x_k)^T (x - x_k) + \frac{1}{2}(x - x_k)^T H_f(x_k)(x - x_k)$ (DL_2 de f en x_k : $f(x) = Q_{x_k}(x) + o(\|x - x_k\|^2)$)

Donc x_{k+1} est donné par $\nabla Q_{x_k}(x_{k+1}) = 0$, avec $\nabla Q_{x_k}(x) = \nabla f(x_k) + H_f(x_k)(x - x_k)$

Donc $\nabla Q_{x_k}(x_{k+1}) = 0 = \nabla f(x_k) + H_f(x_k)(x_{k+1} - x_k)$

$$\iff \boxed{x_{k+1} = x_k - H_f(x_k)^{-1} \nabla f(x_k)}$$

On appelle pas de Newton la direction $d_k = -H_f(x_k)^{-1} \nabla f(x_k)$

Si $H_f(x_k)$ est définie positive (elle est à minima semi-définie positive puisque f est convexe), alors d_k est bien une direction de descente : $\nabla f(x_k)^T d_k = -\nabla f(x_k)^T H_f(x_k)^{-1} \nabla f(x_k) < 0$ (l'inverse d'une matrice définie positive est elle même définie positive)

L'itération de la descente de Newton peut donc s'écrire :

Initialisation : $x_0 \in D_f$

Boucle : Tant que critère d'arrêt non satisfait

$\rightarrow$ $d_k = -H_f(x_k)^{-1} \nabla f(x_k)$ pas de Newton

$\rightarrow$ $x_{k+1} = x_k + d_k$ le pas η est implicitement défini comme $\eta = 1$, mais on peut utiliser un pas $\eta \neq 1$ (critère d'Armijo)

On appelle décrément de Newton en x_k la quantité $\lambda(x_k) = \left(\nabla f(x_k)^T H_f(x_k)^{-1} \nabla f(x_k)\right)^{1/2}$, qui peut aussi s'écrire $\lambda(x_k) = \left(d_k^T H_f(x_k) d_k\right)^{1/2}$ (avec d_k le pas de Newton)

$\rightarrow$ $\lambda^2(x_k) < \varepsilon$ sert en général de condition d'arrêt pour la méthode de Newton

Avantages et inconvénients de la méthode de Newton

$\oplus$ Si f est quadratique, la méthode de Newton converge en 1 itération

$\oplus$ D'une manière générale, la convergence est nettement plus rapide qu'une descente de gradient : si x^* est critique et non dégénéré et H_f lipschitzien au voisinage de x^*, alors :
pour x_k "pas trop loin", $\exists c > 0$ tq $\|x_{k+1} - x^*\| \leq c\|x_k - x^*\|^2$ $\rightarrow$ la convergence est quadratique