Méthode de la plus grande pente / du gradient

On décrit quelques versions de la méthode du gradient pour résoudre le problème

$$(P) \qquad \min\left\{J(x)\ :\ x\in\mathbb{R}^d\right\}.$$

Méthode de la plus grande pente / du gradient : pas constant

1. $x^0\in\mathbb{R}^d$ fixé, pas $\rho>0$ fixé, précision $\varepsilon>0$ fixée,
2. $x\leftarrow x^0$,
3. Tant que $\|\nabla J(x)\|>\varepsilon$ faire
   — $\rho\leftarrow\rho$
   — $x\leftarrow x-\rho\nabla J(x)$
4. afficher x.

Dans la méthode ci-dessus, on peut remplacer l'erreur $\|\nabla J(x)\|$ (qui permet d'évaluer la précision de la solution courante x) par d'autres erreurs vues en cours ...

Méthode de la plus grande pente / du gradient : pas optimal

Dans l'algorithme précédent, on remplace l'étape $\rho\leftarrow\rho$ par $\rho\leftarrow\rho_{opt}$ où ρ_{opt} est la solution du problème

$$\min\left\{q(t)=J\left(x-t\nabla J(x)\right)\ :\ t>0\right\}$$

Méthode de la plus grande pente : recherche linéaire de Wolfe

On reprend la méthode à pas constant, dans laquelle on remplace l'étape $\rho\leftarrow\rho$ par l'algorithme suivant (recherche linéaire de Wolfe), qui se compose de deux étapes : une boucle principale qui cherche un intervalle $[t_g,t_d]$ dans lequel on peut trouver une valeur ρ qui satisfait les conditions de Wolfe fortes, et une fonction "zoom" qui fait la recherche dans cet intervalle.

1. $x\in\mathbb{R}^d$ et $\rho>0$ donnés,
2. paramètres $0<c_2<c_1<1$ fixés,
3. $t_g=0$, t_d fixé grand (par exemple $t_d=2\rho$), $t\leftarrow\rho$, test $\leftarrow 0$,
4. Tant que $test=0$ faire

   SI $[q(t)\geq q(0)+c_2q'(0)t]$ ou $[q'(t)\geq 0]$ ALORS $test\leftarrow 1$ et $t\leftarrow zoom(tg,td)$
   SINON <u>si</u> $|q'(t)|\leq -c_1q'(0)$ <u>alors</u> $test\leftarrow 1$
   <u>sinon</u> $t_g\leftarrow t$, $t_d\leftarrow 10t_d$ et $t\leftarrow (t_g+t_d)/2$
5. renvoyer t.

où $q(t):=J(x-t\nabla J(x))$, donc $q(0)=J(x)$ et $q'(0)=-\|\nabla J(x)\|^2$.

La fonction zoom recherche dans l'intervalle $[t_g,t_d]$ une valeur ρ qui satisfait les conditions de Wolfe, par exemple par dichotomie.

suite au dos

Méthode proximale

1. $x^0 \in \mathbb{R}^d$ fixé, pas $\rho > 0$ fixé, paramètre $\lambda > 0$ fixé, précision $\varepsilon > 0$ fixée,
2. $x \leftarrow x^0$,
3. Tant que $\|\nabla J(x)\| > \varepsilon$ faire
   — Résoudre le problème
   $$\inf\left\{\varphi(x) = J(z) + \frac{1}{2\lambda}\|z - x\|^2 \,:\, z \in \mathbb{R}^d\right\}$$
   par exemple via la méthode de la plus grande pente à pas constant ρ. On note y la solution trouvée.
   — $x \leftarrow y$
4. afficher x.

On remarque que le gradient de φ est donné par

$$\nabla\varphi(z) = \nabla J(z) + \frac{1}{\lambda}(z - x)$$

et est donc facile à calculer si on sait calculer celui de J.

Méthodes de Newton

La méthode de Newton pour l'optimisation consiste en l'algorithme suivant

1. $x^0 \in \mathbb{R}^d$ fixé, pas $\rho > 0$ fixé, précision $\varepsilon > 0$ fixée,
2. $x \leftarrow x^0$,
3. $W \leftarrow D^2 J(x^0)^{-1}$,
4. Tant que $\|\nabla J(x)\| > \varepsilon$ faire
   — $\rho \leftarrow 1$
   — $x_{new} \leftarrow x - \rho W \nabla J(x)$
   — $W \leftarrow D^2 J(x_{new})^{-1}$
   — $x \leftarrow x_{new}$
5. afficher x.

et pour la méthode de quasi-Newton BFGS on remplace $\rho \leftarrow 1$ par une recherche linéaire de type Wolfe pour $q(t) = J(x - t\,W\nabla J(x))$ et on remplace l'étape $W \leftarrow D^2 J(x_{new})^{-1}$ par la formule

$$W = W - \frac{sy^tW + Wys^t}{y^ts} + \left(1 + \frac{y^tWy}{y^ts}\right)\frac{ss^t}{y^ts}$$

où $s = x_{new} - x$ et $y = \nabla J(x_{new}) - \nabla J(x)$.

Méthode des différences finies

Dans les algorithmes précédents, on peut remplacer le calcul de $\nabla J(x)$ par l'algorithme qui calcul un vecteur g qui est une approximation par différences finies de $\nabla J(x)$:

1. $x \in \mathbb{R}^d$ fixé, paramètre $0 < \delta << 1$ fixé,
2. Pour $i = 1$ à d faire
   $$g_i \;:=\; \frac{J(x + \delta\, e_i) - J(x - \delta\, e_i)}{2\,\delta}$$
   où e_i est le i-ième vecteur de la base canonique de $\mathbb{R}^d$,
3. renvoyer $g = (g_1, \ldots, g_d)$.