Minimisations stochastiques et gradient conjugué

Dans $\mathbb{R}^d$, on note $B(m,r)$ la boule ouverte pour la norme infinie centrée en $m \in \mathbb{R}^d$ et de rayon $r > 0$: la boule $B(O,r)$, avec $O = (0,\ldots,0)$ coïncide donc avec le pavé $(-r,r)^d$.

I

La fonction de Goldstein-Price U_{GP} est la fonction polynomiale de deux variables de degré 8

$$U_{\mathrm{GP}}(x,y) = \left(1 + (x+y+1)^2\left(19 - 14x + 3x^2 - 14y + 6xy + 3y^2\right)\right)\\ \left(30 + (2x-3y)^2\left(18 - 32x + 12x^2 + 48y - 36xy + 27y^2\right)\right).$$

dont on cherche le minimum sur la boule $B(0,2)$. On montre que c'est un minimum global et que 3 minima locaux non globaux existent, outre 1 maximum local et 4 points selle.

(1) Tracer les lignes de niveau de U_{GP}.

(2) Effectuer des recherches aléatoires dites *à l'aveugle* (ou de Monte-Carlo simples) de la valeur minimum de U_{GP}.

(3) On considère l'algorithme de recherche accélérée [1] stochastique [RAS] [2].

  (a) Utiliser cet algorithme pour chercher le minimum de la fonction U_{GP}.

  (b) Pour une approximation $\overline{v_*}$ issue d'une recherche à la Monte-Carlo avec $K = 10^6$ itérations de la valeur minimum v_*, déterminer le nombre L d'itérations nécessaires suivant une application de RAS pour passer en dessous de la valeur $\overline{v_*}$. Faire ces recherches combinées $N = 100$ fois et décrire statistiquement la série des valeurs $(L_n)_{1\leq n\leq N}$ obtenues.

(4) Utiliser l'algorithme de gradient conjugué après avoir opéré une recherche à l'aveugle. Comparer avec la méthode couplant recherche à l'aveugle et recherche accélérée.

1. Les boules considérées sont implicitement les traces des boules $B(x,r)$ dans $\mathbb{R}^d$ sur l'espace E des variables admissibles, *i. e.* $B_E(x,r) = \{y \in E, \|y-x\|_\infty < r\}$. Pour $A, B \subset A$ toutes deux de mesure non nulle finie, la méthode du rejet permet de simuler une loi uniforme sur B à partir d'une variable X_A de loi uniforme sur A : si $(X_k)_{k\geq 0}$ est une suite de variables indépendantes de loi uniforme sur A, alors la variable X_K avec $K = \min\{k \geq 0, X_k \in B\}$ est une variable de loi uniforme sur B.

2. *cf.* en page 3 la présentation de cet algorithme et de celui de gradient conjugué non quadratique.

II

Reprendre I.1-4 pour une (ou deux) des fonctions test [3] du tableau suivant où le domaine de la fonction considérée est $[-\ell, \ell]^d$.

Nom	*Fonction*	ℓ	d
Rosenbrock	$100(y-x^2)^2+(1-x)^2$	5	2
Himmelblau	$(x^2+y-11)^2+(x+y^2-7)^2$	5	2
Freudenstein-Roth	$(x-13+((5-y)y-2)y)^2+(x-29+((y+1)y-14)y)^2$	10	2
Jennrich-Sampson	$\sum_{k=1}^{10}(2+2k-(\mathrm{e}^{kx}+\mathrm{e}^{ky}))^2$	1	2
Griewank	$1+\sum_{k=1}^{n}x_k^2/10-\prod_{k=1}^{n}\cos(x_k/\sqrt{k})$	100	n
Rastrigin	$x^2+y^2-\cos(18x)-\cos(18y)$	1	2

III

On introduit des polynômes aléatoires à d variables et de degré $d(k+2)$:

$$P_{d,k}(x_1,\dots,x_d)=\prod_{i=1}^{d}\left[x_i(x_i-1)\prod_{j=1}^{k}(x_i-\zeta_{ij})\right].$$

où les $\zeta_{ij}, 1\le i\le d, 1\le j\le k$ sont i. i. d. avec loi uniforme sur $[0,1]$: le polynôme $P_{d,k}$, produit de polynômes d'une variable à zéros aléatoires dans $[0,1]$, est donc une fonction de plus en plus oscillante lorsque l'entier k augmente.

(1) Étant donné un polynôme P, appliquer une recherche aléatoire avec $K=100d^2k^2$ itérations pour la recherche d'une valeur maximum V_P sur $B(1/2,1/2)$, puis estimer le nombre d'itérations K_{RAS} d'une recherche RAS pour atteindre au moins cette valeur V_P.

(2) Donner dans un tableau les résultats (K, K_{RAS}) obtenus par moyenne d'un échantillon de polynômes $P_{d,k}$ de taille $L=100$ pour différentes valeurs de (d,k) ((1,2-14), (2-5,1-5) par exemple).

Références :
M. Appel, R. Labarre D. Radulović, *On accelerated random search*, SIAM J. Optim. **14**#3 (2003) 708-731.
A. A. Goldstein, I. F. Price, *On descent from local minima*, Math. Comput., **25**#115 (569-574), 1971.

3. Certaines de ces fonctions apparaissent classiquement dans les tests d'algorithmes de recherche de minima : la première dite de Rosenbrock a un minimum à bassin d'attraction très effilé, la seconde dite de Rastringin a de multiples extrema locaux, la troisième dite de Himmelblau a quatre minima globaux.

Algorithme 1.1 Recherche accélérée stochastique [RAS] pour le programme $\min_E U$

Choisir $c > 1,\ \rho > 0,\ k_{\max}$
$r_0 = 2$; $k = 0$
Choisir m_0 suivant la loi uniforme sur $B(0, r_0)$
tant que $k < k_{\max}$ **faire**
  Choisir y suivant la loi uniforme sur $B(m_k, r_k)$
  si $U(y) < U(m_k)$ **alors**
    $m_{k+1} = y$ and $r_{k+1} = r_k$
  sinon
    $m_{k+1} = m_k$ and $r_{k+1} = r_k/c$
  fin si
  si $r_{k+1} < \rho$ **alors**
    $r_{k+1} = r_0$
  fin si
  $k = k + 1$
fin tant que

Algorithme 1.2 Gradient conjugué pour minimiser $\frac{1}{2}\langle Ax, x\rangle + \langle b, x\rangle$

$\varepsilon = 10$; choisir x_0 ; $r_0 = Ax_0 + b$; $d_0 = -r_0$; $h_0 = \dfrac{\langle r_0, d_0\rangle}{\langle Ad_0, d_0\rangle}$
tant que $k < K$ & $\varepsilon > tol$ **faire**
  $x_{k+1} = x_k - h_k d_k$
  $r_{k+1} = Ax_{k+1} + b$
  $d_{k+1} = -r_{k+1} + \dfrac{\|r_{k+1}\|^2}{\|r_k\|^2} d_k$
  $h_{k+1} = \dfrac{\langle r_{k+1}, d_{k+1}\rangle}{\langle Ad_{k+1}, d_{k+1}\rangle}$
  $\varepsilon = \|r_{k+1}\|$
  $k = k + 1$
fin tant que

Algorithme 1.3 : Algorithme du gradient conjugué (J non quadratique)

Choisir x_0 ; r_0 ; $d_0 = -r_0$; $k = 0$
tant que $\|\nabla J(x_k)\| \geq \varepsilon$ **faire**
  $\alpha_{k+1} = RDD(J, x_k + \mathbb{R}_+ d_k)$
  $x_{k+1} = x_k + \alpha_{k+1} d_k$
  $r_{k+1} = \nabla J(x_{k+1})$
  $\beta_{k+1} = \dfrac{\|r_{k+1}\|^2}{\|r_k\|^2}$ (Fletcher-Reeves) ou $\beta_{k+1} = \dfrac{\langle r_{k+1}, r_{k+1} - r_k\rangle}{\|r_k\|^2}$ (Polak-Ribière)
  $d_{k+1} = -r_{k+1} + \beta_{k+1} d_k$
  $k = k + 1$
fin tant que

L'instruction $RDD(J, x + \mathbb{R}_+ d)$ indique une Recherche de minima (exact ou approché) sur la Demi-Droite issue de x et de direction d.