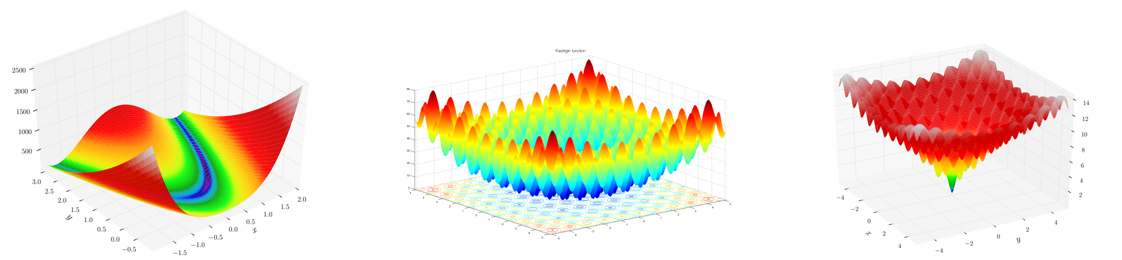
BE1 - Méthodes d'optimisation sans dérivée

Youssef Diouane youssef.diouane@isae-supaero.fr
ISAE - SUPAERO
Toulouse, France.

Dans ce bureau d'étude, nous nous intéressons à la résolution de problèmes d'optimisation sans contrainte en utilisant des méthodes d'optimisation sans dérivée. Le travail est à réaliser en binôme. Le compte-rendu est à déposer sur le LMS avant la date limite : 27/04/2020.

L'objectif de ce BE est d'implémenter puis évaluer les performances d'une méthode déterministe d'optimisation sans dérivée dite méthode de recherche directe directionnelle. Les fonctions à minimiser sont données par :

— Rosenbrock : $f_1(x, y) = 100(y - x^2)^2 + (1 - x)^2$.
— Rastrigin : $f_2(x, y) = 20 + (x^2 - 10\cos(2\pi x)) + (y^2 - 10\cos(2\pi y))$.
— Ackley : $f_3(x, y) = -20\exp\left(-0.2\sqrt{0.5(x^2 + y^2)}\right) - \exp\left(\frac{\cos(2\pi x)+\cos(2\pi y)}{2}\right) + 20 + \exp(1)$.



(a) Rosenbrock. (b) Rastrigin. (c) Ackley.

Pour la réalisation de ce BE, on vous fournit les fichiers Matlab, disponibles via le LMS, qui implémentent les fonctions f_1, f_2 et f_3. L'ensemble de ces fichiers est regroupé dans le répertoire `Problems` :

— `f_rosenbrock.m` : réalise le codage de la fonction f_1 ;
— `f_rastrigin.m` : réalise le codage de la fonction f_2 ;
— `f_ackley.m` : réalise le codage de la fonction f_3 ;

On vous fournit aussi les fichiers Matlab des algorithmes traités dans ce BE (**à compléter**) et un programme principal qui gére l'ensemble des affichages et prépare les entrées et sorties pour chaque algorithme :

— `algo_recherche_directe.m` : la méthode de recherche directe directionnelle ;
— `algo_stratégie_evolution.m` : une stratégie d'évolution ;
— `algo_stratégie_evolution_convergente.m` : une stratégie d'évolution globalement convergente ;

— `BE1_main.m` : un programme principal relatif à l'implémentation de ce BE;

La méthode de recherche directe directionnelle

L'algorithme de recherche directe directionnelle est un algorithme classique à directions de descente : le point courant est déplacé itérativement le long d'une direction qui fait décroître la valeur de la fonction objectif. Plus précisément, nous appellerons direction de descente de f en un point x de $\mathbb{R}^n$ tout vecteur $d \in \mathbb{R}^n$ tel que $f(x + \alpha d) < f(x)$ pour tout pas $\alpha \in \mathbb{R}_+^*$ suffisamment petit. On peut montrer qu'en chaque point x tel que $\nabla f(x) \neq 0$, tout vecteur $d \in \mathbb{R}^n$ tel que $\nabla f(x)^\top d < 0$ est une direction de descente en x (Voir plus de détails en BE2). Autrement dit, tout vecteur non nul formant un angle aigu avec l'opposé du gradient (non nul) de f en x est une direction de descente.

Lorsque le gradient de la fonction f n'est pas accessible il n'est pas possible de trouver une direction de descente de manière aussi directe. Cependant nous pouvons assurer l'existence d'une direction de descente au sein de familles de vecteurs positivement génératrices. Une famille finie D de vecteurs $d_1, \ldots, d_r$ est dite positivement génératrice si tout vecteur v de $\mathbb{R}^n$ peut s'écrire comme une combinaison linéaire d'éléments de D dont les coefficients sont positifs (i.e., il existe des coefficients $\lambda_1, \ldots, \lambda_r$ dans $\mathbb{R}_+$ tels que $v = \sum_{i=1}^r \lambda_i d_i$). Un exemple élémentaire de famille positivement génératrice de $\mathbb{R}^n$ est la réunion des vecteurs de la base canonique de $\mathbb{R}^n$ et de leurs opposés. Nous appellerons directions les éléments de D. On peut montrer (voir le polycopié du cours) qu'une famille de vecteurs non nuls de $\mathbb{R}^n$ est positivement génératrice si et seulement si, pour tout vecteur non nul v de $\mathbb{R}^n$, il existe une direction d dans D telle que $v^\top d > 0$. Autrement dit, pour tout vecteur v non nul, il existe une direction d dans D formant un angle aigu avec v. Ainsi, en parcourant les directions de D – selon un ordre arbitraire – nous sommes assurés d'en trouver une qui soit une direction de descente pour f au point courant.

L'algorithme de recherche directe directionnelle découle naturellement des remarques précédentes. Donnons-nous une famille positivement génératrice D,un point initial x_0 de $\mathbb{R}^n$ et un pas initial α_0 strictement positif. Au début de chaque itération k, un point courant x_k et un pas α_k sont donnés. L'ensemble D est parcouru jusqu'à trouver une direction d telle que $f(x_k + \alpha_k d) < f(x_k)$ – nous savons qu'il en existe une si α est suffisamment petit. Si une telle direction est obtenue, le nouvel itéré est défini par $x_{k+1} = x_k + \alpha_k d$ et le pas est maintenu tel quel voire augmenté : $\alpha_{k+1} = \gamma \alpha_k$ où γ est un élément de $[1, +\infty[$. L'itération k est alors qualifiée de succès. Si D a été parcouru entièrement sans qu'aucune direction satisfaisante n'ait été trouvée, le nouvel itéré est identique au point courant et le pas est réduit : $x_{k+1} = x_k$ et $\alpha_{k+1} = \beta \alpha_k$ où β est un élément de $]0, 1[$. L'itération k est alors qualifiée d'échec.

Afin d'assurer la convergence globale de l'algorithme une diminution suffisante de la valeur de l'objectif au point courant est exigée : donnons-nous une fonction ρ de $\mathbb{R}_+^*$ dans $\mathbb{R}_+^*$ croissante tel que $\lim_{\alpha \to 0} \rho(\alpha)/\alpha = 0$. La condition de réduction suffisante est donnée par :

$$f(x_k + \alpha_k d) < f(x_k) - \rho(\alpha_k \|d\|)$$

Nous ferons le choix classique $\rho(\alpha\|d\|) = \frac{C}{2}\|d\|^2\alpha^2$ pour tout α dans $\mathbb{R}_+^*$. Dans le cadre du BE, on choisit $C = 10^{-2}$.

Question 1 *Écrire un pseudo-code pour l'algorithme de recherche directe directionnelle. Expliciter deux critères d'arrêt convenables (autres que le nombre maximum d'évaluation de la fonction objectif ou d'itérations).*

Question 2 *Implémenter l'algorithme de recherche directe directionnelle. Commenter les résultats obtenus et trouver un choix efficace pour les paramètres suivants :*
— *La valeur des paramètres α_0, γ et β.*
— *Le choix de la famille positivement génératrice D.*

Pour montrer la convergence globale de l'algorithme de recherche directe directionnelle, nous faisons les hypothèses suivantes :
— La fonction f est minorée et continûment différentiable sur $\mathbb{R}^n$. Son gradient ∇f est lipschitzien sur $\mathbb{R}^n$. Notons L une constante de Lipschitz du gradient de f, i.e.,

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\| \qquad \text{pour tout} x, y \in \mathbb{R}^n.$$

— La suite des itérés $\{x_k\}_{k\in\mathbb{N}}$ génerée par l'algorithme de recherche directe directionnelle est uniformément bornée.
— Les éléments d de la famille positivement génératrice D sont bornées en norme, i.e., $\max_{d\in D} \|d\| < \infty$.

La mesure cosinus d'une famille positivement génératrice D est définie par

$$\text{cm}(D) = \min_{v\in\mathbb{R}^n\setminus\{0\}} \max_{d\in D} \frac{v^\top d}{\|v\|\|d\|}.$$

Enfin, nous notons que, par construction, $\text{cm}(D) > 0$.

Question 3 *On considère l'algorithme de recherche directe directionnelle sans critère d'arrêt. Montrer que,*

$$\liminf_{k\to+\infty} \alpha_k = 0.$$

Indication : Par absurde, supposer qu'il existe un $\alpha > 0$ tel que $\alpha_k > \alpha$ pour tout k. Ensuite, utiliser la condition de réduction suffisante pour trouver une contradiction.

Question 4 *En déduire qu'il existe un point $x_* \in \mathbb{R}^n$ et $\{k_i\}$ une sous-suite d'itérations déclarées échec tel que $\lim_{i\to+\infty} \alpha_{k_i} = 0$ et $\lim_{i\to+\infty} x_{k_i} = x_*$*

Question 5 *Soit k une itération déclarée échec (i.e., pour tout $d \in D$ on a $f(x_k + \alpha_k d) \geq f(x_k) - \frac{C}{2}\|d\|^2\alpha_k^2$). Montrer qu'il existe $d \in D$ tel que*

$$cm(D)\|\nabla f(x_k)\|\|d\|\alpha_k - \frac{C}{2}\|d\|^2\alpha_k^2 \leq \int_0^1 \left(\nabla f(x_k + t\alpha_k d) - \nabla f(x_k)\right)^\top (\alpha_k d)\delta t.$$

En déduire que

$$\|\nabla f(x_k)\| \leq \frac{L + C}{2} cm(D)^{-1} \max_{d\in D} \|d\|\alpha_k.$$

Indication : Verifier qu'il existe $d \in D$, tel que $\forall x \in \mathbb{R}^n$ si $\nabla f(x) \neq 0$ alors

$$cm(D)\|\nabla f(x)\|\|d\| \leq -\nabla f(x)^\top d.$$

Question 6 *Montrer que*

$$\liminf_{k\to+\infty} \|\nabla f(x_k)\| = 0 \quad \text{et} \quad \nabla f(x_*) = 0.$$

Un exemple de stratégie d'évolution

Les stratégies d'évolution utilisent un ensemble de μ points (i.e., parents) pour produire λ nouveaux points (i.e., enfants). Pour produire chacun des enfants, μ parents se recombinent. Une fois produits, les enfants sont mutés, généralement par ajout d'une variable aléatoire suivant une loi normale. L'étape de sélection peut s'appliquer, soit uniquement aux enfants, soit à l'ensemble enfants + parents. Dans le premier cas, l'algorithme est noté $(\mu/\mu_W, \lambda) - ES$, dans le second $(\mu/\mu_W + \lambda, \lambda) - ES$. L'algorithme suivant explicite un exemple de stratégie d'évolution de type $(\mu/\mu_W, \lambda) - ES$.

Initialisation : Soit λ et μ deux nombres positifs tels que $\lambda \geq \mu$. On choisit un point de départ x_0, un pas initial $\sigma_0^{\mathrm{ES}} > 0$, une distribution initiale $\mathcal{C}_0$, et des poids initiaux $(\omega_0^1, \ldots, \omega_0^\mu) \in S$. Posons $k = 0$.
Tant que le test de convergence est non satisfait :
1.Génération d'enfants : Calculer les nouveaux points $Y_{k+1} = \{y_{k+1}^1, \ldots, y_{k+1}^\lambda\}$ tel que

$$y_{k+1}^i \;=\; x_k + \sigma_k^{\mathrm{ES}} d_k^i,$$

où d_k^i suit la distribution $\mathcal{C}_k$, $i = 1, \ldots, \lambda$.
2. Sélection de parents : Evaluer $f(y_{k+1}^i)$, $i = 1, \ldots, \lambda$, et réordonner les points enfants dans $Y_{k+1} = \{\tilde{y}_{k+1}^1, \ldots, \tilde{y}_{k+1}^\lambda\}$ dans l'ordre croissant : $f(\tilde{y}_{k+1}^1) \leq \cdots \leq f(\tilde{y}_{k+1}^\lambda)$.
Choisir les nouveaux parents comme les meilleurs μ enfants points $\{\tilde{y}_{k+1}^1, \ldots, \tilde{y}_{k+1}^\mu\}$, et calculer leur moyenne pondérée

$$x_{k+1} \;=\; \sum_{i=1}^{\mu} \omega_k^i \tilde{y}_{k+1}^i.$$

3. Mise à jour : Mettre à jour le pas $\sigma_{k+1}^{\mathrm{ES}}$, la distribution $\mathcal{C}_{k+1}$, et les poids $(\omega_{k+1}^1, \ldots, \omega_{k+1}^\mu) \in S$. Incrementer k et retourner à l'étape 1.

Question 7 *Résoudre les problèmes d'optimisation f_1 f_2 et f_3 avec cette stratégie d'évolution. Commenter.*

Indication : La stratégie d'evolution est fournie dans le fichier `algo_stratégie_evolution.m`

Question 8 *Par rapport à la méthode de recherche directe directionnelle, citer les avantages et inconvénients de cette stratégie d'évolution.*

Question 9 ***(Bonus)*** *Les stratégies d'évolution ne possèdent pas de preuve de convergence globale. Proposer des modifications de cet algorithme pour assurer sa convergence globale.*

Indication : En s'inspirant de la méthode de recherche directe directionnelle, introduire une condition de réduction suffisante dans l'algorithme.

Question 10 ***(Bonus)*** *Évaluer les performances de ce nouvel algorithme. Vos résultats sont-ils les mêmes ?*

Indication : Implémenter votre solution dans `algo_stratégie_evolution_convergente.m`