Derivative Free Optimization: stochastic methods

Laurent Dumas

Laboratoire de Mathématiques de Versailles
Université de Versailles Saint Quentin en Yvelines

DFO course, Mauritius, 11 janvier 2018

Problème d'optimisation à résoudre

- On s'intéresse ici au problème de l'optimisation globale d'une fonction $J : \mathcal{O} \rightarrow \mathbb{R}$ dont le calcul peut éventuellement être complexe et coûteux.
- Dans le cas où $\mathcal{O} \subset \mathbb{R}^n$, le choix doit se porter vers une méthode d'optimisation sans gradient.

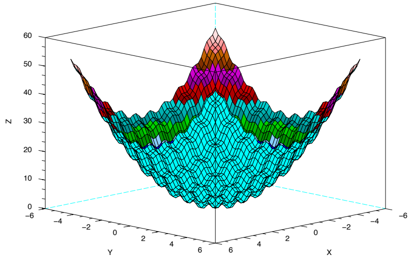
Algorithmes stochastiques : historique

- Les méthodes de type Monte Carlo et leurs variantes (recuit simulé) sont les premiers algorithmes stochastiques ayant été introduits en optimisation.
- Développés plus récemment, les algorithmes évolutionnaires (algorithmes génétiques, stratégies d'évolution, PSO, etc...) sont des algorithmes stochastiques d'optimisation qui tirent leur nom d'une analogie avec la théorie de l'évolution des espèces de Darwin.
- Références : *Holland* (1976), *Goldberg* (1989), *Cerf* (1994), *Schoenaueur* (1996), *Hansen* (2001), etc...

Algorithmes stochastiques

- On présente ici 4 méthodes (SA, GA, ES, PSO) qui seront testées sur la fonction de Rastrigin :

$$Rast(x_1, ..., x_n) = \sum_{i=1}^{n} \left(x_i^2 - \cos(2\pi x_i)\right) + n$$



- Celle-ci possède un grand nombre de minimas locaux (10^n sur $[-5, 5]$) et un seul minimum global.

Principe général du recuit simulé (SA)

- Choix d'un élément initial $X^1 \in \mathcal{O}$
- for n from 1 to $Ngen$
- Mutation : remplacer X^n par Y^n, choisi aléatoirement dans un voisinage.
- Evaluation de $J(Y^n)$.
  - si $J(Y^n) < J(X^n)$, alors $X^{n+1} = Y^n$.
  - si $J(Y^n) \geq J(X^n)$, alors $X^{n+1} = Y^n$ avec une probabilité $exp(-\frac{J(Y^n)-J(X^n)}{T})$ et $X^{n+1} = X^n$ sinon.
- Mise à jour du paramètre T ($T \to 0$ lentement)
- end for

Principe général d'un algorithme génétique (GA)

- Choix d'une population initiale $P_1 = \{X_i^1 \in \mathcal{O}, 1 \leq i \leq N_p\}$
- for n from 1 to $Ngen$
- Evaluation de $\{J(X_i^n), 1 \leq i \leq N_p\}$.
- Creation d'une population de N_p individus par :
  - Selection de (X_α^n, X_β^n) en fonction de leur facteur de santé.
  - Croisement : remplacer (X_α^n, X_β^n) par (Y_α^n, Y_β^n).
  - Mutation : remplacer (Y_α^n, Y_β^n) par (Z_α^n, Z_β^n).
- end for
- Generation de la nouvelle population P_{n+1}.
- end for

Principe général d'une stratégie d'évolution (ES)

- Choix d'une population initiale de μ parents : $P_1 = \{X_i^1 \in \mathcal{O}, \quad 1 \leq i \leq \mu\}$
- for n from 1 to $Ngen$
- Creation d'une population de $\lambda \geq \mu$ enfants O_n par :
  - Croisement à ω parents $Y_i^n = \frac{1}{\omega}(\sum_{j=1}^{\omega} X_j^n)$
  - Mutation : remplacer Y_i^n par Z_i^n
- Evaluation de $\{J(Z_i^n), 1 \leq i \leq \lambda\}$
- Selection des meilleurs μ parents dans la population $P_n \cup O_n$.
- end for

Principe général d'un essaim de particules (PSO)

- Choix d'une population initiale $P_1 = \{(X_i^1, v_i^1, p_i^1), 1 \leq i \leq N_p\}$ de particules ayant la position actuelle $X_i \in \mathcal{O}$, la vitesse v_i et une meilleure position p_i.
- for n from 1 to $Ngen$
- Evaluation de $\{J(X_i^n), 1 \leq i \leq N_p\}$.
- Actualisation de la meilleure position individuelle et globale(p_g^n)
  - Calcul des nouvelles vitesses de chaque particule :

$$v_i^{n+1} = \omega v_i^n + c_1\rho_1(p_i^n - x_i^n) + c_2\rho_2(p_g^n - x_i^n)$$

  - Calcul des nouvelles positions de chaque particule :

$$X_i^{n+1} = X_i^n + v_i^{n+1}$$

- end for

Principaux avantages

- Grâce à leur caractère stochastique, il s'agit de méthodes d'optimisation globale.
- Grâce à l'utilisation d'une population, il s'agit de méthodes facilement parallélisables.
- Toutes ces méthodes permettent de gérer les contraintes de manière relativement simple et efficace par un principe de pénalisation statique ou dynamique ou de stochastic ranking.
- La plupart de ces méthodes possèdent une version multi-objectif permettant de déterminer un front de Pareto.
- Ces méthodes sont également très stables par rapport aux erreurs numériques commises sur la fonction J.

Principaux inconvénients

- Le principal inconvénient de ces méthodes est leur coût de calcul important lié au grand nombre d'évaluations de la fonction J à effectuer.
- La vitesse de convergence est également très lente comparée à une méthode déterministe.
- Le choix des paramètres est très influent sur la qualité des résultats obtenus.
- Il existe très peu de résultats de convergence théorique de ces méthodes.

Un résultat de convergence pour les ES

- La plupart des résultats de convergence concernent des fonctions sphériques : $J(x) = g(||x||^2)$ avec g croissante.
- Soit par exemple une stratégie d'évolution $(1, 1)$ représentée par la suite de vecteurs aléatoires (X_n) et ayant une mutation gaussienne de variance $\sigma||X_n||$. Sous certaines hypothèses techniques et pour une fonction sphérique, X_n converge presque surement vers le minimum de J et :

$$\lim_{n \to +\infty} \frac{1}{n} \ln \left(\frac{||X_n||}{||X_0||} \right) = c < 0$$