'Derivative Free Optimisation: surface response methods

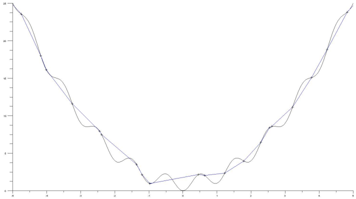
Laurent Dumas

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

Master AMS, V04 course, 2017-18

Modèles approchés

- De manière générale, l'objectif consiste à construire une fonction approchée $\tilde{J}$ (*surrogate ou metamodèle*) de la fonction exacte J à partir d'un certain nombre de points $(X_i, J(X_i))_{1 \leq i \leq N}$ où la fonction exacte est supposée connue.



- <u>Références</u> : *Giannakoglou* (2001), *Jin* (2005), etc...

Modèles approchés : méthode de krigeage

- La première méthode présentée ici, appelée méthode de krigeage, est une méthode probabiliste basée sur la minimisation de la variance de l'estimation en un point X donné.
- L'approximation de la fonction coût en un point $X \in \mathbb{R}^n$ s'écrit :

$$\hat{j}(X) = \sum_{i=1}^{N} \omega(X_i) j(X_i)$$

  où on suppose que $\hat{j}(X)$ et $j(X)$ sont des réalisation des variables aléatoires respectives $\hat{J}(X)$ et $J(X)$ (*dans cette approche $J(X)$ n'est plus un réel mais une variable aléatoire*).
- Afin de déterminer $\hat{j}(X)$, on suppose que la covariance de J est connue :

$$cov(J(X), J(Y)) = c(X, Y)$$

Modèles approchés : méthode de krigeage

- En cherchant à minimiser $var(\hat{J}(X) - J(X))$ tout en imposant $E(\hat{J}(X) - J(X)) = 0$, on aboutit à une relation permettant de déterminer $\hat{j}(X)$:

$$\hat{j}(X) = K^T C^{-1} z$$

  où K est le vecteur colonne de terme général $c(X_i, X)$, C est la matrice de terme général $c(X_i, X_j)$, et z le vecteur colonne de terme général $j(X_i)$.

- Une estimation de la variance au point X est également disponible :

$$var(\hat{J}(X) - J(X)) = c(X, X) - K^T C^{-1} K$$

Modèles approchés : méthode de krigeage

- La fonction de corrélation est en général choisie comme étant de type exponentielle :

$$c(X, Y) = \theta_1 \exp\left(-\frac{1}{2}\sum_{i=1}^{n}\frac{(x_i - y_i)^2}{r_i^2}\right) + \theta_2$$

- Les paramétres $\Theta = (\theta_1, \theta_2, r_1, ..., r_n)$ sont alors déterminés par le principe du maximum de vraisemblance, c'est à dire en maximisant la fonction :

$$\mathcal{L}(\Theta) = p(j(X_1), ..., j(X_N)) = \frac{1}{\sqrt{(2\pi)^N \det C}} \exp\left(-\frac{1}{2} z^T C^{-1} z\right)$$

Modèles approchés : méthode RBF

- Une autre méthode possible, appelée méthode RBF, est construite comme une combinaison linéaire de fonctions radiales centrées en chacun des points X_i.
- L'approximation de la fonction coût en un point $X \in \mathbb{R}^n$ s'écrit alors :

$$\tilde{J}(X) = \sum_{i=1}^{N} w_i h(||X - X_i||)$$

où h désigne une fonction $r \mapsto h(r)$ dite fonction de base radiale.

Modèles approchés : méthode RBF

- Les poids $(w_i)_{1 \leq i \leq N}$ sont calculés par résolution de l'équation matricielle $Aw = z$ traduisant l'exactitude du réseau sur les points $(X_i)_{1 \leq i \leq N}$, où la matrice $A \in \mathcal{M}_N(\mathbb{R})$ a pour terme général $a_{i,j} = h(||X_i - X_j||)$ et le second membre a pour terme général $z_i = J(X_i)$.
- On a donc ici :

$$\tilde{J}(X) = R^T A^{-1} z$$

  où R est le vecteur colonne de terme général $h(||X - X_i||)$.
- Pour des fonctions h bien choisies, on peut montrer que la matrice A est toujours inversible voire définie positive.

Modèles approchés : méthode RBF

- Une fonction continue f définie sur $\mathbb{R}_+^*$ est dite (définie) positive si pour toute famille de points distincts $X_1, ..., X_N$ de $\mathbb{R}^n$, la forme quadratique

$$q(c_1, ..., c_N) = \sum \sum_{1 \leq i,j \leq N} c_i c_j f(||X_i - X_j||)$$

  est (définie) positive.
- **Théorème** (Schoenberg) : Une fonction f est totalement monotone sur $\mathbb{R}_+^*$ si et seulement si la fonction $r \rightarrow f(r^2)$ est positive.
- Ainsi, les fonctions $r \mapsto e^{-r^2}$ et $r \mapsto (1+r^2)^{-\alpha}$ avec $\alpha > 0$ peuvent être utilisées comme fonctions de base dans les réseaux RBF.

Modèles approchés : méthode RBF

- **Théorème** (Micchelli) : soit h une fonction dérivable sur $\mathbb{R}_+$, strictement positive sur $\mathbb{R}_+^*$. Si la première dérivée de h est totalement monotone et non constante sur $\mathbb{R}_+^*$, alors pour toute famille de points distincts $X_1, ..., X_N$ de $\mathbb{R}^n$:

$$(-1)^{N-1} det\left(\left[h(||X_i - X_j||^2)\right]\right) > 0$$

- Ainsi, les fonctions $r \to (c^2 + r^2)^\alpha$ avec $c \in \mathbb{R}$ et $0 < \alpha < 1$ peuvent être utilisées comme fonction de base radiale dans les réseaux RBF.

Modèles approchés : méthode RBF

- Afin de déterminer les paramètres optimaux des fonctions de base du modèle RBF, une méthode de type 'leave-one out' peut être utilisée.
- Cette méthode consiste à entrainer le réseau sur tous les points sauf un et à tester l'erreur commise sur ce point. En répétant ce procédé sur tous les points, on aboutit à une erreur globale qu'il s'agit de rendre minimale.
- La méthode RBF rejoint alors la méthode de krigeage si une fonction gaussienne est choisie dans les deux cas. Seule la façon de chercher les paramètres optimaux de cette gaussienne diffèrent.

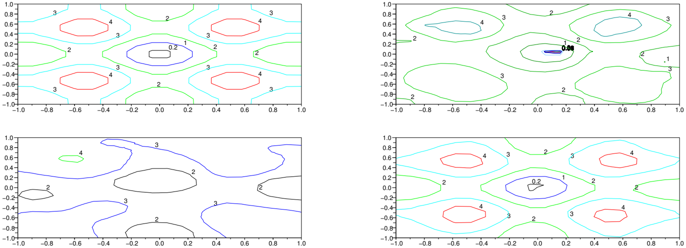
Modèles approchés : méthode RBF

- Dans le cas où le nombre de points N est trés grand, la matrice A peut être mal conditionnée. Afin d'éviter ce problème, deux choix sont possibles.
- Soit un procédé de régularisation de Tychonov est ajouté permettant de réduire le conditionnement de A. Dans ce cas, la méthode RBF cesse d'être une méthode d'interpolation.
- Soit le nombre de points N est réduit à m en ne considérant que les plus proches points du point X à calculer. Dans ce cas, le réseau construit devient local.

Modèles approchés : méthode RBF

- Exemple sur la fonction de Rastrigin :

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



- La figure ci-dessus compare les contours de la fonction de Rastrigin et trois modèles approchés, construit à partir d'un réseau RBF avec 40 ou 200 points d'exemples.