TD 6: algorithmes d'optimisation sans gradient

Exercice 1

On considère la méthode 'pattern search' pour minimiser la fonction $f : \mathbb{R}^n \to \mathbb{R}$ avec un nombre fini de directions $\mathcal{D}$ tel que

$$\forall d \in \mathcal{D}, \quad ||d|| = 1$$

et

$$\kappa = \min_{||v||=1} \max_{d \in \mathcal{D}} v^T d > 0$$

On note $(x_k)_{k\in\mathbb{N}}$ la suite de points construite avec la méthode et $(\alpha_k)_{k\in\mathbb{N}}$ le pas associé. On rappelle que le critère de succès est le suivant :

$$f(x_k + \alpha_k d) < f(x_k) - c\frac{\alpha_k^2}{2}$$

où $c > 0$ est fixé et que $\alpha_{k+1} = \theta\alpha_k$ (respectivement $\alpha_{k+1} = \gamma\alpha_k$) dans le cas d'un échec (respectivement succès) avec $\theta \in]0, 1[$ et $\gamma \geq 1$.

Le lemme suivant peut être démontré :

Lemma On suppose que f est C^1, ∇f est ν-Lipschitz et que f est minorée par $m \in \mathbb{R}$. Alors, la suite de pas satisfait pour tout $N \in \mathbb{N}$:

$$\sum_{k=0}^{N} \alpha_k^2 \leq \frac{2\gamma^2}{c(1-\theta^2)} \left(\frac{c\alpha_0^2}{2\gamma^2} + f(x_0) - m \right)$$

1. Prouver que

$$\forall (x, y) \in \mathbb{R}^n \times \mathbb{R}^n, \quad |f(y) - f(x) - \nabla f(x)^T (y - x)| \leq \frac{\nu}{2} ||y - x||^2$$

2. Prouver que $\lim_{k\to\infty} \alpha_k = 0$ et que le nombre de pas d'échecs est infini.
3. Prouver que pour un pas d'échec, on a

$$||\nabla f(x_k)|| \leq \frac{c + \nu}{2\kappa} \alpha_k$$

4. Prouver que

$$\liminf_{k\to\infty} ||\nabla f(x_k)|| = 0$$

Exercice 2

On considère la méthode de Nelder Mead pour minimiser la fonction $\mathbb{R}^2$:

$$f(x,y) = x^2 + y^2$$

et un simplexe inital formé de $A = (4,5)$, $B = (5,3)$ et $C = (5,6)$.

Calculer la solution obtenue après une itération.

Exercice 3

Prouver qu'aucun pas de 'rétrécissement' n'est possible pour la minimisation d'une fonction convexe avec la méthode de Nelder Mead.

Exercice 4

On note $S_k = \{y_k^0, ..., y_k^n\}$ le simplexe obtenu par la méthode de Nelder Mead à l'itération k associé aux valeurs suivantes pour la fonction f : $f_k^0 \leq f_k^1 \leq ... \leq f_k^n$.

On suppose que la fonction$f$ est minorée.

1. Prouver que la suite $(f_k^0)_{k\in\mathbb{N}}$ est convergente.
2. Si un nombre fini de 'rétrécissements' survient, prouver que toutes les suites $(f_k^i)_{k\in\mathbb{N}}$ sont convergentes ($0 \leq i \leq n$).
3. On définit le volume du simplexe :

   $$V(S_k) = \frac{1}{n!}\det[y_k^0 - y_k^n, ...y_k^{n-1} - y_k^n]$$

   Vérifier que pour $n = 2$ la définition de $V(S_k)$ correspond à l'aire du triangle (S_k).
4. Donner la valeur de $V(S_{k+1})$ en fonction de $V(S_k)$ quand un pas d'expansion se produit (on pourra supposer pour simplifier quet $y_k^n = 0$).
5. Donner la valeur de $V(S_{k+1})$ en fonction de $V(S_k)$ quand un pas de rétrécissement se produit.

Exercice 5

Le modèle de krigeage est un modèle approché de $J : \mathbb{R}^n \to \mathbb{R}$ qui peut être écrit ainsi :

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

où les N points $X_i \in \mathbb{R}^n$ ont été évalués par J.

Dans cette expression, J et $\hat{J}$ sont des champs de vecteurs aléatoires et J est à moyenne nulle en tout point. De plus, la covariance entre les points X et Y a la forme suivante :

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

où la fonction $c : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ est supposée connue.

En chaque point $X \in \mathbb{R}^n$, le champs de vecteurs aléatoire $\hat{J}$ minimise la variance entre $J(X)$ et $\hat{J}(X)$, tout en assurant $E(\hat{J}(X)) = E(J(X))$.

1. Prouver que
$$\hat{J}(X) =^t KC^{-1}z$$
où $K =^t (c(X_1, X), ..., c(X_N, X))$, $z =^t (J(X_1), ...J(X_N))$ et C est une $N \times N$ matrice telle que $C_{i,j} = c(X_i, X_j)$.
2. Vérifier que la fonction de krigeage est une fonction d'interpolation, c'est à dire que $\hat{J}(X_i) = J(X_i)$ pour tout $i \in \{1, ..., N\}$.
3. Donner une expression de $Var(J(X) - \hat{J}(X))$ en utilisant $c(X, X)$, K et C.

Exercice 6

On considère les deux modèles approchés pour une fonction f définie sur $\mathbb{R}^n$:

— Un modèle RBF avec une fonction de base radiale :
$$h(r) = e^{-cr^2}$$
— Un modèle de krigeage avec une fonction de corrélation :
$$c(x, y) = \theta_1 + \theta_2 \exp(-\sum_{i=1}^{n} \frac{(x_i - y_i)^2}{2\sigma_i})$$

Prouver que pour un certain jeu de paramètres qui seront déterminés, les deux modèles approchés sont identiques.