BE1 - Optimisation quadratique successive

Le travail est à réaliser **en binôme**.
Le compte-rendu et le code sont à déposer sur le LMS avant le **vendredi 18/10/2024 à 12h**.

Dans ce bureau d'étude, nous nous intéressons à une classe de méthodes de type newtonienne dites par *Optimisation Quadratique Successive* ou *Sequential Quadratic Programming* (SQP) en anglais. Ces méthodes sont conçues pour la résolution des problèmes d'optimisation avec des contraintes d'égalité de la forme:

$$\min_{\mathbf{x}\in C} J(\mathbf{x}) \tag{$\mathcal{P}$}$$

où $C = \{\mathbf{x} \in \mathbb{R}^n \,|\, c_i(\mathbf{x}) = 0,\, i = 1, \dots, p\}$.

Les méthodes d'optimisation proposées dans ce BE seront testées sur les problèmes suivants définis, pour $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$ comme suit :

- Problème (`P1`) : la fonction objectif est $J_1(\mathbf{x}) = x_1^2 + x_2^2$ avec une seule contrainte $x_1 + x_2 = 1$.
- Problème (`P2`) : la fonction objectif est $J_2(\mathbf{x}) = x_1^2 + x_2^2$ avec une seule contrainte $x_1^2 x_2 = 16$.
- Problème (`P3`) : la fonction objectif est $J_3(\mathbf{x}) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$ avec une seule contrainte $x_1 + x_2 = 1$.

On rappelle que le Lagrangien L associé au problème $(\mathcal{P})$ est donné par:

$$L(\mathbf{x}, \lambda) = J(\mathbf{x}) + \lambda^T c(\mathbf{x}).$$

Nous avons vu en cours que sous certaines conditions une solution $\mathbf{x}_*$ de $(\mathcal{P})$ est un point critique du Lagrangien L. Comme pour l'algorithme de Newton, l'idée essentielle de la méthode SQP consiste à résoudre une succession de problèmes quadratiques qui sont des approximations du problème de départ, mais qui, dans ce cas, comporteront des contraintes linéaires. En effet, étant donné $(\mathbf{x}_k, \lambda_k) \in \mathbb{R}^n \times \mathbb{R}^p$, on cherche $\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{d}_k$, où $\mathbf{d}_k$ est une direction de descente calculée comme la solution du problème suivant :

$$\min_{\substack{\mathbf{d} \in \mathbb{R}^n \\ c'(\mathbf{x}_k)\mathbf{d} + c(\mathbf{x}_k) = 0}} \langle \nabla_{\mathbf{x}} J(\mathbf{x}_k), \mathbf{d}\rangle + \frac{1}{2}\langle \nabla^2_{\mathbf{xx}} L(\mathbf{x}_k, \lambda_k)\mathbf{d}, \mathbf{d}\rangle \tag{$\mathcal{QP}_k$}$$

où $c'(\mathbf{x}_k)$ représente la matrice Jacobienne des contraintes c en $\mathbf{x}_k$. Le vecteur $\lambda_{k+1} \in \mathbb{R}^p$ sera choisi comme un multiplicateur de Lagrange optimal associé au problème $(\mathcal{QP}_k)$ si un tel vecteur existe.

Dans un premier temps, on souhaite justifier théoriquement la faisabilité de l'algorithme SQP.

Question 1 : *Montrer que si $\mathbf{d}_k$ est solution de $(\mathcal{QP}_k)$ alors il existe un multiplicateur optimal noté λ_{k+1} tel que $(\mathbf{d}_k, \lambda_{k+1})$ est une solution du système linéaire:*

$$\begin{bmatrix} \nabla^2_{\mathbf{xx}} L(\mathbf{x}_k, \lambda_k) & c'(\mathbf{x}_k)^T \\ c'(\mathbf{x}_k) & 0 \end{bmatrix} \begin{bmatrix} \mathbf{d}_k \\ \lambda_{k+1} \end{bmatrix} = - \begin{bmatrix} \nabla_{\mathbf{x}} J(\mathbf{x}_k) \\ c(\mathbf{x}_k) \end{bmatrix} \tag{$\mathcal{LS}_k$}$$

Question 2 : *On suppose que la matrice $\nabla^2_{\mathbf{xx}} L(\mathbf{x}_k, \lambda_k) \in \mathbb{R}^{n\times n}$ est symétrique définie positive, et que $c'(\mathbf{x}_k) \in \mathbb{R}^{p\times n}$ est surjective (c'est-à-dire de rang maximum, rang$(c'(\mathbf{x}_k)) = p < n$). Montrer que le système linéaire $(\mathcal{LS}_k)$ admet une solution unique notée $(\mathbf{d}_k, \lambda_{k+1})$. En déduire que, dans ce cas, $\mathbf{d}_k$ est* *une solution de* $(\mathcal{QP}_k)$*.*

Question 3 : *Écrire un pseudo-code pour l'algorithme SQP. Expliciter deux critères d'arrêt convenables (autres que le nombre maximum d'évaluations de la fonction objectif ou d'itérations).*

Question 4 : *Implémenter l'algorithme SQP et commenter les résultats obtenus.*

La hessienne du Lagrangien ($Q_k = \nabla^2_{\mathbf{xx}} L(\mathbf{x}_k, \lambda_k)$) peut s'avérer très coûteuse à calculer (voire à estimer). Pour cela, en pratique dans ce type d'algorithme, on souhaite remplacer Q_k par une approximation H_k construite à partir du dernier déplacement en utilisant un algorithme de quasi-Newton de type BFGS. En effet, à chaque itération k, on pose $\mathbf{s}_{k-1} = \mathbf{x}_k - \mathbf{x}_{k-1}$ et $\mathbf{y}_{k-1} = \nabla_{\mathbf{x}} L(\mathbf{x}_k, \lambda_k) - \nabla_{\mathbf{x}} L(\mathbf{x}_{k-1}, \lambda_{k-1})$. Si $(\mathbf{y}_{k-1})^T \mathbf{s}_{k-1} > 0$, on choisit

$$H_k = H_{k-1} + \frac{\mathbf{y}_{k-1}(\mathbf{y}_{k-1})^T}{(\mathbf{y}_{k-1})^T \mathbf{s}_{k-1}} - \frac{H_{k-1}\mathbf{s}_{k-1}(\mathbf{s}_{k-1})^T H_{k-1}}{(\mathbf{s}_{k-1})^T H_{k-1} \mathbf{s}_{k-1}},$$

sinon $H_k = H_{k-1}$.

Question 5 : *En se basant sur la formule de BFGS, donner un pseudo-code de la variante de l'algorithme SQP qui n'utilise pas les dérivées secondes. Implémenter l'algorithme et commenter les résultats obtenus par rapport aux précédents.*

Les gradients de la fonction objectif et des contraintes peuvent également s'avérer coûteux voire impossibles à calculer pour certaines applications. On peut adapter l'algorithme SQP (à base de BFGS) à ce problème en utilisant une méthode de différences finies. Par exemple, dans le cas de la fonction objectif J, pour tout $i = 1, \ldots, n$ et $y \in \mathbb{R}^n$, on a :

$$\frac{\partial J}{\partial x_i}(y) \approx \frac{J(y + he_i) - J(y)}{h},$$

où $h > 0$ est un pas de discrétisation (suffisamment petit, par exemple $h = 10^{-5}$) et e_i est le i-ème vecteur de la base canonique.

Question 6 : *Implémenter l'algorithme SQP de BFGS à base de différences finies. Commenter les résultats obtenus par rapport aux précédents.*

Question 7 : *Citer les avantages et les inconvénients des méthodes SQP (telles qu'elles sont présentées dans ce BE). Que suggéreriez-vous pour améliorer ces méthodes?*