BE1 - Optimisation quadratique successive (SQP: Sequential Quadratic Programming)

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

Le travail est à réaliser en binôme. Le compte-rendu est à déposer sur le LMS avant la date limite : le vendredi 08/10/2021.

Dans ce bureau d'étude, nous nous intéressons à une classe de méthodes applée SQP (Sequential Quadratic Programming) conçue pour la résolution des problèmes d'optimisation avec des contraintes en égalité de la forme :

$$\mathcal{P} \ : \ \min_{x \in \mathcal{C}} \ J(x),$$

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

Les méthodes d'optimisation, proposées dans ce BE, seront testées sur les problèmes suivants :

— Problem (`P1`) : $x \in \mathbb{R}^2$, la fonction objectif est $f(x) = x_1^2 + x_2^2$ avec une seule contrainte $x_1 + x_2 = 1$.

— Problem (`P2`) : $x \in \mathbb{R}^2$, la fonction objectif est $f(x) = x_1^2 + x_2^2$ avec une seule contrainte $x_1^2 x_2 = 16$.

— Problem (`P3`) : $x \in \mathbb{R}^2$, la fonction objectif est $100(x_2 - x_1^2)^2 + (1 - x_1)^2$ avec une seule contrainte $x_1 + x_2 = 1$.

On rappelle que le Lagrangien $\mathcal{L}$ associé au problem $\mathcal{P}$ est donnée par :

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

Nous avons vu en cours qu'une solution x^* de ce problème est un point critique du Lagrangien $\mathcal{L}$ mais ce n'est pas en général un minimum local du Lagrangien. Comme pour l'algorithme de Newton, l'idée essentielle de la méthode SQP consiste à résoudre une succession de problèmes quadratiques avec contraintes linéaires (ces problèmes sont relativement simples à résoudre) qui sont des approximations du problème de départ. En effet, étant donné x_k, on cherche $x_{k+1} = x_k + d_k$ où d_k est une direction de descente calculée de la façon suivante : on cherche x_{k+1} qui diminue la valeur du Lagrangien (car c'est le Lagrangien qui joue le role de la fonction objectif en présence des contraintes). Pour un vecteur $\lambda_k \in \mathbb{R}^p$, on construit une approximation du Lagrangien au voisinage de x_k, i.e.,

$$\mathcal{L}(x_k + d, \lambda_k) = \mathcal{L}(x_k, \lambda_k) + \langle \nabla_x \mathcal{L}(x_k, \lambda_k), d \rangle + \frac{1}{2} \langle \nabla_x^2 \mathcal{L}(x_k, \lambda_k) d, d \rangle + \mathcal{O}(\|d\|^3).$$

Si on néglige les termes d'ordre supérieur ou égal à 3, on voit qu'il suffit de trouver un d_k qui minimise

$$\langle \nabla_x \mathcal{L}(x_k, \lambda_k), d \rangle + \frac{1}{2} \langle \nabla_x^2 \mathcal{L}(x_k, \lambda_k) d, d \rangle$$

pour espérer minimiser le Lagrangien.

Question 1 *Donner l'expression de $\nabla_x \mathcal{L}(x,\lambda)$ et $\nabla_x^2 \mathcal{L}(x,\lambda)$, puis montrer que d_k est solution du problème quadratique :*

$$\mathcal{QP}_k \quad : \quad \min_{D_c(x_k)d + c(x_k) = 0} \quad \langle \nabla_x J(x_k), d\rangle + \frac{1}{2}\langle \nabla_x^2 \mathcal{L}(x_k,\lambda_k)d, d\rangle.$$

où $D_c(x_k)$ représente la matrice Jacobienne des contraintes.

Le vecteur λ_{k+1} sera choisi comme étant le multiplicateur optimal de Lagrange associé au problème $\mathcal{QP}_k$.

Question 2 *Montrer que $(d_k, \lambda_{k+1})^\top$ est solution du système linéaire :*

$$\mathcal{LS}_k \quad : \quad \begin{pmatrix} \nabla_x^2 \mathcal{L}(x_k,\lambda_k) & D_c(x_k)^\top \\ D_c(x_k) & 0 \end{pmatrix} \begin{pmatrix} d \\ \lambda \end{pmatrix} = - \begin{pmatrix} \nabla J(x_k) \\ c(x_k) \end{pmatrix}.$$

Question 3 *On suppose que la matrice $\nabla_x^2 \mathcal{L}(x_k,\lambda_k) \in \mathbb{R}^{n\times n}$ est symétrique définie positive, et que $D_c(x_k) \in \mathbb{R}^{p\times n}$ est surjective (i.e. de rang maximum $\mathrm{rang}(B) = p < n$). Montrer que le système linéaire $\mathcal{LS}_k$ admet une solution unique.*

Question 4 *Écrire un pseudo-code pour l'algorithme SQP. Expliciter deux critères d'arrêt convenables (autres que le nombre maximum d'évaluation de la fonction objectif ou d'itérations).*

Question 5 *Implémenter l'algorithme SQP et commenter les résultats obtenus.*

Le hessien du Lagrangien $Q = \nabla_x^2 \mathcal{L}(x_k,\lambda_k)$ peut s'avérer très coûteux à calculer (voire à estimer). Pour cela, en pratique dans ce genre d'algorithme, on remplace Q par une approximation H_k construite à partir du dernier déplacement en utilisant un algorithm quasi-Newton de type BFGS. En effet, à chaque itération k, on pose $s_{k-1} = x_k - x_{k-1}$ et $y_{k-1} = \nabla_x \mathcal{L}(x_k,\lambda_k) - \nabla_x \mathcal{L}(x_{k-1},\lambda_{k-1})$: si $y_{k-1}^\top s_{k-1} > 0$, on choisit

$$H_k = H_{k-1} + \frac{y_{k-1}y_{k-1}^\top}{y_{k-1}^\top s_{k-1}} - \frac{H_{k-1}s_{k-1}s_{k-1}^\top H_{k-1}}{s_{k-1}^\top H_{k-1}s_{k-1}},$$

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

Question 6 *En se basant sur la formule BFGS, donner un pseudo-code de la variante de l'algorithme SQP qui n'utilise pas la dérivée seconde. Implémenter l'algorithme et commenter les résultats obtenus par rapport aux précédents.*

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

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

où h est une tolérance de perturbation (suffisamment petite, e.g., $h = 10^{-5}$) et e_i est le ième vecteur de la base canonique.

Question 7 *Implémenter l'algorithme SQP sans utiliser les dérivées. Commenter les résultats obtenus par rapport aux précédents.*

Question 8 *Citer les avantages et les inconvenients des méthodes SQP (telles qu'elles sont présentées dans ce BE). Que suggérez-vous pour améliorer ces méthodes ?*