Notes du cours MTH1101 – Calcul I
Partie III: Optimisation

Guy Desaulniers

Département de mathématiques et de génie industriel
École Polytechnique de Montréal

Automne 2022

Table des matières

Table des matières

Soit un domaine $D \subseteq \mathbb{R}^n$ et une fonction $f : D \to \mathbb{R}$ de n variables.
On cherche un point $\vec{x} \in D$ qui minimise la valeur de $f(\vec{x})$ sur D.
On écrit :

$$\min_{\vec{x} \in D} \; f(\vec{x}).$$

La fonction f est appelée la fonction objectif.

Définition : minimum absolu

$\vec{x}^* \in D$ est un minimum absolu (global) de $f(\vec{x})$ sur D si

$$f(\vec{x}^*) \leq f(\vec{x}), \qquad \forall\, \vec{x} \in D.$$

Définition : minimum local

$\vec{x}^* \in D$ est un minimum local de $f(\vec{x})$ sur D s'il existe $\epsilon > 0$ tel que

$$f(\vec{x}^*) \leq f(\vec{x}), \qquad \forall\, \vec{x} \in D \cap B_\epsilon(\vec{x}^*),$$

où

$$B_\epsilon(\vec{x}^*) = \{\vec{x} \in \mathbb{R}^n \ : \ \|\vec{x} - \vec{x}^*\| < \epsilon\}$$

est la boule de rayon ϵ centrée en $\vec{x}^*$.

Remarques

1. Un minimum absolu est aussi un minimum local.
2. Un minimum peut aussi être un maximum.
3. Il n'existe pas toujours de minimum absolu, ni de minimum local.
4. En général, il est difficile de trouver un minimum absolu. Dans le cours, on se concentre surtout sur la recherche d'un minimum local.
5. Résoudre un problème de maximisation est équivalent à résoudre un problème de minimisation, i.e.,

$$\max_{\vec{x}\in D} f(\vec{x}) \quad \equiv \quad -\min_{\vec{x}\in D} -f(\vec{x}).$$

On peut donc se concentrer que sur les problèmes de minimisation.

Table des matières

En optimisation sans contraintes, on considère $D = \mathbb{R}^n$. On cherche donc à résoudre

$$\min_{\vec{x} \in \mathbb{R}^n} \quad f(\vec{x}).$$

Condition nécessaire d'optimalité du premier ordre

Si $\vec{x}^*$ est un minimum local, alors

$$\nabla f(\vec{x}^*) = \vec{0}.$$

Définition : point critique

$\vec{x}^*$ est un point critique de $f(\vec{x})$ si $\nabla f(\vec{x}^*) = \vec{0}$.

Un point critique peut être

- un minimum local
- un maximum local
- un point de selle (i.e., $\forall\, \epsilon > 0$, il existe $\vec{a}, \vec{b} \in B_\epsilon(\vec{x}^*)$ tels que $f(\vec{a}) < f(\vec{x}^*) < f(\vec{b})$).

Définition : matrice hessienne

Soit $f(\vec{x})$ une fonction de n variables. La matrice hessienne de $f(\vec{x})$, notée $\nabla^2 f(\vec{x})$, est donnée par

$$\nabla^2 f(\vec{x}) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(\vec{x}) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(\vec{x}) & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(\vec{x}) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(\vec{x}) & \frac{\partial^2 f}{\partial x_2^2}(\vec{x}) & \dots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(\vec{x}) \\ \dots & \dots & \ddots & \dots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(\vec{x}) & \frac{\partial^2 f}{\partial x_n \partial x_2}(\vec{x}) & \dots & \frac{\partial^2 f}{\partial x_n^2}(\vec{x}) \end{pmatrix}.$$

Cette matrice est symétrique si toutes les dérivées sont continues.

Conditions nécessaires du second ordre

Si $\vec{x}^*$ est un minimum local de $f(\vec{x})$, alors

$$\nabla f(\vec{x}^*) = \vec{0} \quad \text{et} \quad \vec{y}^{\mathsf{T}} \nabla^2 f(\vec{x}^*) \vec{y} \geq 0, \ \forall\, \vec{y} \in \mathbb{R}^n$$

(i.e., la matrice hessienne $\nabla^2 f(\vec{x}^*)$ est semi-définie positive).

Conditions suffisantes du second ordre

Soit $\vec{x}^* \in \mathbb{R}^n$. Si

$$\nabla f(\vec{x}^*) = \vec{0} \quad \text{et} \quad \vec{y}^{\mathsf{T}} \nabla^2 f(\vec{x}^*) \vec{y} > 0, \ \forall\, \vec{y} \in \mathbb{R}^n$$

(i.e., la matrice hessienne $\nabla^2 f(\vec{x}^*)$ est définie positive), alors $\vec{x}^*$ est un minimum local de $f(\vec{x})$.

Une matrice $n \times n$ symétrique A est définie positive si

$$\det\left([a_{11}]\right) > 0, \det\left(\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\right) > 0,$$

$$\det\left(\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}\right) > 0, \quad \dots \quad , \det(A) > 0.$$

Une matrice $n \times n$ symétrique A est définie négative si la matrice $-A$ est définie positive.

Une matrice $n \times n$ symétrique A est semi-définie positive si $\det(A) \geq 0$ et toutes ses sous-matrices carrées obtenues en éliminant des lignes et les colonnes correspondantes ont un déterminant non négatif.

Par exemple, une matrice 2×2 symétrique A est semi-définie positive si

$$\det\left([a_{11}]\right) \geq 0, \det\left([a_{22}]\right) \geq 0 \quad \text{et} \quad \det(A) \geq 0.$$

Une matrice $n \times n$ symétrique A est semi-définie négative si la matrice $-A$ est semi-définie positive.

En pratique, pour résoudre un problème d'optimisation sans contraintes, on trouve un point critique $\vec{x}^*$ et on étudie sa nature :

1. Si $\nabla^2 f(\vec{x}^*)$ est définie positive, alors $\vec{x}^*$ est un minimum local
2. Si $\nabla^2 f(\vec{x}^*)$ est définie negative, alors $\vec{x}^*$ est un maximum local
3. Si $\nabla^2 f(\vec{x}^*)$ n'est pas semi-définie positive, alors $\vec{x}^*$ n'est pas un minimum local
4. Si $-\nabla^2 f(\vec{x}^*)$ n'est pas semi-définie négative, alors $\vec{x}^*$ n'est pas un maximum local
5. Si les conditions 3 et 4 sont satisfaites, alors $\vec{x}^*$ est un point de selle.

Table des matières

Méthode du gradient dans $\mathbb{R}^2$

- Cette méthode permet de trouver, en général, un minimum local d'une fonction $f(x, y)$.
- C'est une méthode itérative
  - (x_k, y_k) : point courant au début de l'itération k
  - $\vec{d}_k = -\nabla f(x_k, y_k)$: direction de descente au point (x_k, y_k)
  - $h(\alpha) = f((x_k, y_k) + \alpha \vec{d}_k)$: fonction à une variable à minimiser à l'itération k

Pseudo-code de l'algorithme

1: Soit (x_0, y_0) un point initial.
2: Poser $k = 0$.
3: Calculer $\vec{d}_k = -\nabla f(x_k, y_k)$.
4: **Si** $\|\vec{d}_k\| = 0$ **alors**
5: Arrêter car (x_k, y_k) est un point critique.
6: **Sinon**
7: Résoudre le problème de recherche linéaire

$$\min_{\alpha \geq 0} \; h(\alpha) = f((x_k, y_k) + \alpha \vec{d}_k).$$

Dénoter par α^* le minimum obtenu.
8: Poser $(x_{k+1}, y_{k+1}) = (x_k, y_k) + \alpha^* \vec{d}_k$.
9: Poser $k = k + 1$.
10: Aller à l'étape 3.

Remarques

- En général, l'algorithme trouve un point critique qui est un minimum local.
- L'algorithme peut ne pas converger (i.e., aller vers $-\infty$).
- S'il y a plusieurs minima locaux et que l'algorithme converge, on ne sait pas vers quel minimum local.
- Plusieurs méthodes peuvent être utilisées pour résoudre le problème de recherche linéaire. Dans le cours, il sera résolu de façon analytique.
- Pour un problème de maximisation, on choisit $\vec{d}_k = \nabla f(x_k, y_k)$ et on maximise pour la recherche linéaire.
- L'algorithme se généralise au cas de plus de deux variables.
- L'algorithme s'approche d'un point critique en zigzaguant.

Table des matières

Soit $D \subseteq \mathbb{R}^n$ un domaine et $f : D \to \mathbb{R}$ une fonction de n variables. Considérons P, le problème d'optimisation avec contraintes

$$\min_{\vec{x} \in D} f(\vec{x}).$$

Définitions : ensemble fermé, ensemble borné

- Un ensemble S est fermé si tout point $\vec{x}$ de la frontière de S appartient à S.
- Un ensemble S est borné s'il existe une constante $M \in \mathbb{R}$ telle que $\|\vec{x}\| \leq M$, $\forall\, \vec{x} \in S$.

Théorème

Si D est fermé et borné et $f(\vec{x})$ est une fonction continue sur D, alors le problème P a au moins un minimum absolu $\vec{x}^* \in D$.

Premier cas (une contrainte d'égalité) :

$$D = \{\vec{x} \in \mathbb{R}^n \ : \ g(\vec{x}) = k\}.$$

Conditions nécessaires d'optimalité du premier ordre

Si $\vec{x}^*$ est un minimum local de $f(\vec{x})$ sur $D \in \mathbb{R}^n$ et $\nabla g(\vec{x}^*) \neq \vec{0}$, alors il existe $\lambda \in \mathbb{R}$ tel que

$$\begin{aligned} \nabla f(\vec{x}^*) &= \lambda \nabla g(\vec{x}^*) & (1) \\ g(\vec{x}^*) &= k. & (2) \end{aligned}$$

Définition : point critique

Les points $\vec{x}^*$ pour lesquels il existe λ tel que les conditions (1) et (2) sont satisfaites sont appelés des points critiques.

Chacun de ces points peut être un minimum local, un maximum local ou un point de selle.

Définition : multiplicateur de Lagrange

λ est appelé le multiplicateur de Lagrange associé à la contrainte $g(\vec{x}) = k$.

Soit

$$\begin{aligned} v(k) = \min \quad & f(\vec{x}) \\ \text{s.à} \quad & g(\vec{x}) = k \end{aligned}$$

une fonction qui donne la valeur minimale de ce problème d'optimisation en fonction de k.

Proposition

Soit λ le multiplicateur de Lagrange associé à $g(\vec{x}) = k$. Alors

$$\lambda = v'(k).$$

Pour un minimum local, λ indique le taux de variation de la valeur du minimum local.

Le polynôme de Taylor de degré 1 de $v(k)$ autour de k_0 est

$$T_1(k) = v(k_0) + \lambda_0(k - k_0),$$

où λ_0 est le multiplicateur de Lagrange pour $k = k_0$. Ce polynôme peut être utilisé pour estimer les valeurs optimales $v(k)$ pour k près de k_0, i.e.,

$$v(k) \approx v(k_0) + \lambda_0(k - k_0).$$

Second cas (plusieurs contraintes d'égalité) :

$$D = \{\vec{x} \in \mathbb{R}^n \ : \ g_j(\vec{x}) = k_j, j = 1, 2, \ldots, m\}.$$

Conditions nécessaires d'optimalité du premier ordre

Si $\vec{x}^*$ est un minimum local de $f(\vec{x})$ sur $D \in \mathbb{R}^n$ et l'ensemble des gradients $\{\nabla g_j(\vec{x}^*) : \ j = 1, 2, \ldots, m\}$ est linéairement indépendant, alors il existe $\lambda_j \in \mathbb{R}$, $j = 1, 2, \ldots, m$, tels que

$$\begin{aligned} \nabla f(\vec{x}^*) &= \sum_{j=1}^{m} \lambda_j \nabla g_j(\vec{x}^*) && (3) \\ g_j(\vec{x}^*) &= k_j, \quad j = 1, 2, \ldots, m. && (4) \end{aligned}$$

Troisième cas (une contrainte d'inégalité) :

$$D = \{\vec{x} \in \mathbb{R}^n \ : \ g(\vec{x}) \leq k\}.$$

Il y a deux possibilités :

1. un minimum local $\vec{x}^*$ est dans l'intérieur strict de D et alors $\nabla f(\vec{x}^*) = 0$.
2. un minimum local $\vec{x}^*$ est sur la frontière de D et alors il satisfait les conditions (1) et (2) :

$$\begin{aligned} \nabla f(\vec{x}^*) &= \lambda \nabla g(\vec{x}^*) \\ g(\vec{x}^*) &= k. \end{aligned}$$

Méthode d'optimisation (si D est fermé et borné)

1. Trouver tous les points tels que

$$\nabla f(\vec{x}) = 0 \quad \text{et} \quad g(\vec{x}) < k$$

   ou

$$\nabla f(\vec{x}) = \lambda \nabla g(\vec{x}) \quad \text{avec} \quad \lambda \in \mathbb{R}, \;\; g(\vec{x}) = k \quad \text{et} \quad \nabla g(\vec{x}) \neq 0.$$

2. Évaluer $f(\vec{x})$ en tous ces points et comparer leurs valeurs pour trouver un minimum absolu.