Optimisation continue

Farida Meghatria

University of Khemis Miliana

Cours et exercices, Master 1, Analyse Mathématique et Applications.
Chapitre2: Rappels de calcul différentiel

Contents

1. Fonctions différentiables : rappels essentiels
2. Différentielle de Fréchet
3. La matrice hessienne
   - Exemple
4. Théorème des fonctions implicites
5. Conclusion

Fonctions différentiables : rappels essentiels

Fonctions de $\mathbb{R}^n$ dans $\mathbb{R}$

Soit $f : \mathbb{R}^n \to \mathbb{R}$, une fonction réelle à plusieurs variables. Elle associe à chaque vecteur $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ une valeur réelle $f(x)$.

Exemple :

$$f(x_1, x_2) = x_1^2 + 3x_1x_2 + 2x_2^2$$

Dérivée directionnelle

La dérivée directionnelle de f en un point $x \in \mathbb{R}^n$ dans la direction d'un vecteur unitaire $v \in \mathbb{R}^n$ est définie par :

$$D_v f(x) = \lim_{t \to 0} \frac{f(x + tv) - f(x)}{t}$$

Elle mesure le taux de variation de f dans la direction de v.

Gradient

Le **gradient** de f, noté $\nabla f(x)$, est le vecteur des dérivées partielles premières :

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(x) \\ \frac{\partial f}{\partial x_2}(x) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x) \end{pmatrix}$$

Propriété : la dérivée directionnelle s'exprime à l'aide du gradient :

$$D_v f(x) = \nabla f(x)^T v$$

Le gradient indique donc la direction de la plus forte croissance de la fonction f

Exemple

Pour $f(x_1, x_2) = x_1^2 + 3x_1x_2 + 2x_2^2$, on a :

$$\frac{\partial f}{\partial x_1} = 2x_1 + 3x_2, \quad \frac{\partial f}{\partial x_2} = 3x_1 + 4x_2$$

$$\Rightarrow \nabla f(x) = \begin{pmatrix} 2x_1 + 3x_2 \\ 3x_1 + 4x_2 \end{pmatrix}$$

Différentielle

La **différentielle** de f au point x, notée $df(x)$, est une application linéaire :

$$df(x) : \mathbb{R}^n \to \mathbb{R}, \quad df(x)(h) = \nabla f(x)^T h$$

Cela signifie que, pour une petite perturbation $h \in \mathbb{R}^n$, on a :

$$f(x+h) \approx f(x) + df(x)(h) = f(x) + \nabla f(x)^T h$$

Cette expression correspond à une approximation linéaire de la fonction autour du point x.

Interprétation géométrique

- Le gradient est orthogonal aux courbes (ou surfaces) de niveau de f.
- La différentielle permet d'approximer localement f par une fonction affine.
- Cela est à la base de nombreuses méthodes d'optimisation (ex. : descente de gradient).

Conditions de différentiabilité

Une fonction $f : \mathbb{R}^n \to \mathbb{R}$ est dite **différentiable** en un point $x \in \mathbb{R}^n$ si :

- $\surd$ toutes ses dérivées partielles premières existent autour de x,
- $\surd$ et sont continues dans un voisinage de x.

Dans ce cas, l'approximation $f(x+h) \approx f(x) + \nabla f(x)^T h$ est valide pour h petit.

Mini-exercice

Soit $f(x,y) = \ln(x^2+y^2)$, avec $(x,y) \neq (0,0)$.

1. Calculer le gradient $\nabla f(x,y)$.
2. Donner une approximation de $f(x+h_1, y+h_2)$ à l'aide de la différentielle.

Correction

Soit $f(x,y) = \ln(x^2+y^2)$, avec $x \neq 0$, $y \neq 0$.
[1] Calcul du gradient $\nabla f(x,y)$
On calcule les dérivées partielles de f :

$$\frac{\partial f}{\partial x} = \frac{1}{x^2+y^2} \cdot 2x = \frac{2x}{x^2+y^2}$$

$$\frac{\partial f}{\partial y} = \frac{1}{x^2+y^2} \cdot 2y = \frac{2y}{x^2+y^2}$$

Le gradient est donc :

$$\nabla f(x,y) = \begin{pmatrix} \dfrac{2x}{x^2+y^2} \\ \dfrac{2y}{x^2+y^2} \end{pmatrix}$$

[2] Approximation de $f(x+h_1, y+h_2)$ à l'aide de la différentielle
La différentielle de f en (x,y) appliquée à $h=(h_1,h_2)$ est :

$$df(x,y)(h) = \nabla f(x,y)^T \cdot h = \frac{2x}{x^2+y^2}h_1 + \frac{2y}{x^2+y^2}h_2$$

L'approximation affine de f est donc :

$$f(x+h_1, y+h_2) \approx f(x,y) + df(x,y)(h)$$

$$\Rightarrow f(x+h_1, y+h_2) \approx \ln(x^2+y^2) + \frac{2x}{x^2+y^2}h_1 + \frac{2y}{x^2+y^2}h_2$$

Différentielle de Fréchet

Différentielle de Fréchet

La notion de différentielle de Fréchet généralise la dérivée classique à des fonctions définies entre espaces vectoriels normés. Elle permet de formaliser la linéarisation optimale d'une application au voisinage d'un point donné. Cette notion est essentielle dans le cadre de l'analyse fonctionnelle et de l'optimisation en dimension finie ou infinie.

Définition

Soient E et F deux espaces vectoriels normés, et soit $f : E \to F$ une application. On dit que f est **différentiable au sens de Fréchet** en un point $x_0 \in E$ s'il existe une application linéaire continue $L : E \to F$ telle que :

$$\lim_{h \to 0} \frac{\|f(x_0 + h) - f(x_0) - L(h)\|_F}{\|h\|_E} = 0.$$

Dans ce cas, on note $Df(x_0) = L$ la **différentielle de Fréchet** de f en x_0. L'application L est l'approximation linéaire optimale de f au voisinage de x_0.

Remarques

- La différentielle de Fréchet, si elle existe, est unique.
- Si f est différentiable au sens de Fréchet en x_0, alors f est continue en x_0.
- En dimension finie, la différentielle de Fréchet coïncide avec la matrice jacobienne (vue comme application linéaire).

Lien avec la dérivabilité classique

Soit $f : \mathbb{R}^n \to \mathbb{R}^m$ une fonction. Alors, f est différentiable au sens de Fréchet en x_0 si et seulement si elle est différentiable au sens classique en x_0, et la différentielle de Fréchet est alors donnée par la matrice jacobienne de f en x_0 :

$$Df(x_0)(h) = J_f(x_0) \cdot h,$$

où $J_f(x_0)$ est la matrice jacobienne de f en x_0, et $h \in \mathbb{R}^n$.

Exemple

Soit la fonction $f : \mathbb{R}^2 \to \mathbb{R}$ définie par :

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

Calculons la différentielle de Fréchet en un point (x_0, y_0).
La différentielle est définie par :

$$Df(x_0, y_0)(h_1, h_2) = \frac{d}{dt} f(x_0 + th_1, y_0 + th_2)\Big|_{t=0}.$$

On dérive :

$$f(x_0 + th_1, y_0 + th_2) = (x_0 + th_1)^2 + (x_0 + th_1)(y_0 + th_2) + (y_0 + th_2)^2.$$

On obtient:

$$Df(x_0, y_0)(h_1, h_2) = 2x_0h_1 + y_0h_1 + x_0h_2 + 2y_0h_2 = (2x_0 + y_0)h_1 + (x_0 + 2y_0)h_2.$$

Donc, la différentielle de Fréchet en (x_0, y_0) est l'application linéaire :

$$Df(x_0, y_0) : (h_1, h_2) \mapsto (2x_0 + y_0)h_1 + (x_0 + 2y_0)h_2.$$

Ce qui correspond au produit scalaire :

$$Df(x_0, y_0)(h) = \nabla f(x_0, y_0)^T \cdot h,$$

avec

$$\nabla f(x_0, y_0) = \begin{pmatrix} 2x_0 + y_0 \\ x_0 + 2y_0 \end{pmatrix}.$$

La matrice hessienne

Définition

Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction deux fois dérivable. La **matrice hessienne** de f en un point $x \in \mathbb{R}^n$ est la matrice carrée $H_f(x) \in \mathbb{R}^{n \times n}$ des dérivées partielles secondes, définie par :

$$H_f(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(x) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(x) & \frac{\partial^2 f}{\partial x_2^2}(x) & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(x) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(x) & \frac{\partial^2 f}{\partial x_n \partial x_2}(x) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(x) \end{pmatrix}$$

Symétrie de la hessienne

Si f est de classe $\mathcal{C}^2$, c'est-à-dire que toutes ses dérivées partielles secondes sont continues, alors la matrice hessienne est **symétrique** :

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i} \quad \text{(théorème de Schwarz)}$$

Interprétation quadratique

La hessienne intervient dans le développement de Taylor à l'ordre 2 :

$$f(x+h) \approx f(x) + \nabla f(x)^T h + \frac{1}{2} h^T H_f(x) h$$

Ce terme quadratique est essentiel pour analyser la nature des points critiques.

Rôle en optimisation : test de la hessienne

Soit $x^* \in \mathbb{R}^n$ tel que $\nabla f(x^*) = 0$ (point critique). Alors :

Si $H_f(x^*)$ est définie **positive** (toutes ses valeurs propres sont > 0), alors x^* est un **minimum local**.

Si $H_f(x^*)$ est définie **négative**, alors x^* est un **maximum local**.

Si $H_f(x^*)$ a des valeurs propres de signes opposés, alors x^* est un **point selle**.

Si $H_f(x^*)$ est semi-définie, il faut une analyse plus fine.

Soit $f(x,y) = x^2 + 4xy + y^2$. On calcule la hessienne :

$$\frac{\partial f}{\partial x} = 2x + 4y, \quad \frac{\partial f}{\partial y} = 4x + 2y$$

$$H_f(x,y) = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{pmatrix} = \begin{pmatrix} 2 & 4 \\ 4 & 2 \end{pmatrix}$$

Calculons les valeurs propres :

$$\det \begin{pmatrix} 2-\lambda & 4 \\ 4 & 2-\lambda \end{pmatrix} = (2-\lambda)^2 - 16 = \lambda^2 - 4\lambda - 12 \Rightarrow \lambda_1 = 6,\ \lambda_2 = -2$$

Les valeurs propres sont de signes opposés implique (f) admet un **point selle** en tout point critique.

Remarque

La matrice hessienne est centrale dans les méthodes d'optimisation numérique de type Newton, où elle sert à corriger l'approximation de la direction de descente.

Théorème des fonctions implicites

Motivation

Dans de nombreux problèmes d'optimisation avec contraintes, on cherche à exprimer certaines variables comme fonctions d'autres à partir d'une équation implicite :

$$F(x, y) = 0$$

où $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ est une fonction différentiable. Le théorème des fonctions implicites garantit, sous certaines conditions, qu'il est possible localement d'écrire y comme une fonction de x.

Énoncé du théorème (cas général)

Soit $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ une fonction de classe $\mathcal{C}^1$ (continûment différentiable).
Soit $(x_0, y_0) \in \mathbb{R}^n \times \mathbb{R}^m$ tel que :

$$F(x_0, y_0) = 0 \quad \text{et} \quad \det\left(\frac{\partial F}{\partial y}(x_0, y_0)\right) \neq 0$$

Alors, il existe un voisinage $U \subset \mathbb{R}^n$ de x_0, un voisinage $V \subset \mathbb{R}^m$ de y_0, et une fonction $g : U \to V$, de classe $\mathcal{C}^1$, telle que :

$$\forall x \in U, \quad F(x, g(x)) = 0 \quad \text{et} \quad g(x_0) = y_0$$

Autrement dit, on peut résoudre localement $F(x, y) = 0$ sous la forme $y = g(x)$.

Conséquences en optimisation

Le théorème permet :

- d'éliminer des variables dans les contraintes,
- de reformuler un problème sous contraintes comme un problème sans contrainte,
- de justifier l'utilisation de multiplicateurs de Lagrange (voir conditions de KKT).

Exemple simple

Soit la contrainte $F(x,y) = x^2 + y^2 - 1 = 0$. On souhaite exprimer y en fonction de x, près de $x_0 = 0$, $y_0 = 1$.
On a :

$$\frac{\partial F}{\partial y}(x,y) = 2y, \quad \text{et} \quad \frac{\partial F}{\partial y}(0,1) = 2 \neq 0$$

Donc, d'après le théorème, il existe une fonction $y = g(x)$ définie dans un voisinage de 0 telle que $x^2 + g(x)^2 = 1$, avec $g(0) = 1$. On retrouve ici $g(x) = \sqrt{1 - x^2}$ (localement).

Formule pour la dérivée de la fonction implicite

Si $y = g(x)$ satisfait $F(x, g(x)) = 0$, alors la dérivée de g est donnée par :

$$Dg(x) = -\left(\frac{\partial F}{\partial y}(x, g(x))\right)^{-1} \cdot \frac{\partial F}{\partial x}(x, g(x))$$

Cette formule est utile pour calculer les gradients de fonctions contraintes.

Quelques références

[1] R. H. Eckel, R. Kahn, R. M. Robertson and R. A. Rizza, Preventing cardiovascular disease and diabetes: a call to action from the American Diabetes Association and the American Heart Association. Circulation. 2006; 113(25): 2943-2946.

[2] F.Meghatria, O.Belhamiti.Predictive model for the risk of cardiovascular disease and type 2. Chaos, Solitons and Fractals. 2021; 146: 110834

[3] ZHU, Jun, WANG, Sibo, WU, Yujie, et al. Smartphone addiction habit is positively associated with coronary artery disease and its severity in Chinese adults: a case-control study. Frontiers in Cardiovascular Medicine, 2024; vol. 11, p. 1374797.

[4] Z. M. Nabhan and T. S. Hannon, Cardiometabolic risk factors in American children: what can be learned from current trends?. Journal of Adolescent Health. 2012; 50(2): 107-109.