3 Approximation and Optimization in Several Variables

3.1 Taylor's Theorem

Let $f : X \to \mathbb{R}$ be a function where $X \subset \mathbb{R}$ is open. Suppose f is differentiable up to order at least k. For any $a \in X$, define the following *Taylor polynomial* of degree k.

$$P_k(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2}(x-a)^2 + \cdots + \frac{f^{(k)}(a)}{k!}(x-a)^k$$

Taylor's theorem states that

$$f(x) = P_k(x) + R_k(x, a)$$

where

$$\lim_{x \to a} \frac{R_k(x,a)}{(x-a)^k} = 0.$$

The *remainder term* $R_k(x,a)$ can be expressed as follows. If f is differentiable up to order at least $k+1$, then there exists ξ between a and x such that

$$R_k(x,a) = \frac{f^{(k+1)}(\xi)}{(k+1)!}(x-a)^{k+1}.$$

Theorem 3.1. (Taylor's theorem) Let $f : X \to \mathbb{R}$ be a differentiable function where $X \subset \mathbb{R}^n$ is open. For any $\mathbf{a} \in X$, define

$$P_1(\mathbf{x}) = f(\mathbf{a}) + Df(\mathbf{a})(\mathbf{x} - \mathbf{a}).$$

Then

$$f(\mathbf{x}) = P_1(\mathbf{x}) + R_1(\mathbf{x}, \mathbf{a})$$

where

$$\lim_{\mathbf{x} \to \mathbf{a}} \frac{R_1(\mathbf{x}, \mathbf{a})}{\|\mathbf{x} - \mathbf{a}\|} = 0.$$

Definition 3.1. Let $f : X \to \mathbb{R}$ be a differentiable function where $X \subset \mathbb{R}^n$ is open. Let $\mathbf{a} \in X$ and let $\mathbf{h} \in \mathbb{R}^n$. The *incremental change* of f is defined by

$$\Delta f = f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a}).$$

The *total differential* of f is defined by

$$df(\mathbf{a}, \mathbf{h}) = Df(\mathbf{a})\mathbf{h}.$$

The first-order Taylor's theorem says that when $\mathbf{h}$ has a small norm, then $\Delta f \approx df$.

Example 3.1. A cuboid of size $1 \times 2 \times 3$ m^3 is to be made. Suppose there is an error of at most 0.02 m for each side in the final product. Then the maximum error of the volume V of the cuboid is approximately $dV = 0.22$ m^3.

Theorem 3.2. (Taylor's theorem) Let $f : X \to \mathbb{R}$ be a function of class C^k where $X \subset \mathbb{R}^n$ is open. For any $\mathbf{a} \in X$, define

$$P_k(\mathbf{x}) = f(\mathbf{a}) + \sum_{j=1}^{k} \frac{1}{j!}\Big(\sum_{i_1,i_2,\ldots,i_j=1}^{n} f_{x_{i_1}x_{i_2}\cdots x_{i_j}}(\mathbf{a})(x_{i_1}-a_{i_1})(x_{i_2}-a_{i_2})\cdots(x_{i_j}-a_{i_j})\Big).$$

Then

$$f(\mathbf{x}) = P_k(\mathbf{x}) + R_k(\mathbf{x},\mathbf{a})$$

where

$$\lim_{\mathbf{x}\to\mathbf{a}} \frac{R_k(\mathbf{x},\mathbf{a})}{\|\mathbf{x}-\mathbf{a}\|^k} = 0.$$

Proposition 3.1. Let $f : X \to \mathbb{R}$ be a function of class C^{k+1} where $X \subset \mathbb{R}^n$ is open. For any $\mathbf{a} \in X$, define

$$P_k(\mathbf{x}) = f(\mathbf{a}) + \sum_{j=1}^{k} \frac{1}{j!}\Big(\sum_{i_1,i_2,\ldots,i_j=1}^{n} f_{x_{i_1}x_{i_2}\cdots x_{i_j}}(\mathbf{a})(x_{i_1}-a_{i_1})(x_{i_2}-a_{i_2})\cdots(x_{i_j}-a_{i_j})\Big).$$

Then there exists ξ lying on the segment between $\mathbf{a}$ and $\mathbf{x}$ such that

$$f(\mathbf{x}) = P_k(\mathbf{x}) + \frac{1}{(k+1)!} \sum_{i_1,i_2,\ldots,i_{k+1}=1}^{n} f_{x_{i_1}x_{i_2}\cdots x_{i_{k+1}}}(\xi)(x_{i_1}-a_{i_1})(x_{i_2}-a_{i_2})\cdots(x_{i_{k+1}}-a_{i_{k+1}}).$$

Definition 3.2. Let $f : X \to \mathbb{R}$ be a function of class C^2 where $X \subset \mathbb{R}^n$ is open. The *Hessian* of f is defined by

$$Hf = \begin{pmatrix} f_{x_1x_1} & f_{x_1x_2} & \cdots & f_{x_1x_n} \\ f_{x_2x_1} & f_{x_2x_2} & \cdots & f_{x_2x_n} \\ \vdots & \vdots & \ddots & \vdots \\ f_{x_nx_1} & f_{x_nx_2} & \cdots & f_{x_nx_n} \end{pmatrix}.$$

Using this notation, we can rewrite the formula for $P_2(\mathbf{x})$ in Taylor's theorem as follows:

$$P_2(\mathbf{x}) = f(\mathbf{a}) + Df(\mathbf{a})(\mathbf{x}-\mathbf{a}) + \frac{1}{2}(\mathbf{x}-\mathbf{a})^T Hf(\mathbf{a})(\mathbf{x}-\mathbf{a}).$$

Example 3.2. Define $f : \mathbb{R}^2 \to \mathbb{R}$ by $f(x, y) = e^x \sin(x + y)$ and consider $\mathbf{a} = \mathbf{0}$. Then

$$Hf = \begin{pmatrix} 2e^x \cos(x+y) & e^x(-\sin(x+y) + \cos(x+y)) \\ e^x(-\sin(x+y) + \cos(x+y)) & -e^x \sin(x+y) \end{pmatrix}$$

and

$$P_2(x, y) = x + y + x^2 + xy.$$

Example 3.3. Define $f : \mathbb{R}^2 \to \mathbb{R}$ by $f(x, y) = \ln(x^2 + y^2 + 1)$ and consider $\mathbf{a} = \mathbf{0}$. For any (x, y) with $|x|, |y| \leqslant 0.1$, we have

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

where $|E(x, y)| \leqslant 0.00083$.

3.2 Extrema of Functions of Several Variables

Definition 3.3. Let $f : X \to \mathbb{R}$ be a function where $X \subset \mathbb{R}^n$. We say that f has a *global minimum* (plural: *global minima*) at $\mathbf{a} \in X$ if $f(\mathbf{x}) \geqslant f(\mathbf{a})$ for all $\mathbf{x} \in X$.

Similarly, we say that f has a *global maximum* (plural: *global maxima*) at $\mathbf{a} \in X$ if $f(\mathbf{x}) \leqslant f(\mathbf{a})$ for all $\mathbf{x} \in X$.

Definition 3.4. Let $f : X \to \mathbb{R}$ be a function where $X \subset \mathbb{R}^n$. We say that f has a *local minimum* (plural: *local minima*) at $\mathbf{a} \in X$ if there exists an open set U containing $\mathbf{a}$ such that $f(\mathbf{x}) \geqslant f(\mathbf{a})$ for all $\mathbf{x} \in (X \cap U)$.

Similarly, we say that f has a *local maximum* (plural: *local maxima*) at $\mathbf{a} \in X$ if there exists an open set U containing $\mathbf{a}$ such that $f(\mathbf{x}) \leqslant f(\mathbf{a})$ for all $\mathbf{x} \in (X \cap U)$.

If $\mathbf{a}$ is a point of minimum or maximum, then we say that it is a point of (local or global) *extremum* (plural: *extrema*).

Theorem 3.3. Let $f : X \to \mathbb{R}$ be a function where $X \subset \mathbb{R}^n$, and let $\mathbf{a}$ be an interior point of X. If f has a local extremum at $\mathbf{a}$ and $\nabla f(\mathbf{a})$ exists, then $\nabla f(\mathbf{a}) = \mathbf{0}$.

The converse of this theorem does not hold. This means $\nabla f(\mathbf{a}) = \mathbf{0}$ does not imply f has a local extremum at $\mathbf{a}$.

Definition 3.5. Let $f : X \to \mathbb{R}$ be a function where $X \subset \mathbb{R}^n$, and let $\mathbf{a}$ be an interior point of X. We say that $\mathbf{a}$ is a *critical point* of f if $\nabla f(\mathbf{a})$ does not exist or $\nabla f(\mathbf{a}) = \mathbf{0}$.

It follows that an interior local extremum point must be a critical point.

Definition 3.6. Let $f : X \to \mathbb{R}$ be a function where $X \subset \mathbb{R}^n$, and let $\mathbf{a}$ be an interior point of X. We say that $\mathbf{a}$ is a *saddle point* of f if $\nabla f(\mathbf{a}) = \mathbf{0}$ but $\mathbf{a}$ is not a point of local extremum.

Example 3.4. Suppose there is a mountain whose surface is the graph of

$$f(x, y) = -x^2 - 4y^2 + 6x - 8y + 24$$

with $f(x, y) \geqslant 0$. The highest point of this mountain is $(3, -1, 37)$.

Example 3.5. Define $f : \mathbb{R}^3 \to \mathbb{R}$ by

$$f(x, y, z) = x^2 + 2y^2 - z^2 + 2yz + x + 3z - 5.$$

Then $\nabla f(x, y, z) = (2x + 1, 4y + 2z, -2z + 2y + 3)$. The only critical point of f is $\left(-\frac{1}{2}, -\frac{1}{2}, 1\right)$, which is a saddle point. Also, f does not have global extrema.

Definition 3.7. Let A be an $n \times n$ real symmetric matrix. We say that A is *positive definite* (*negative definite*) if $\mathbf{x}^T A\mathbf{x} > 0$ ($\mathbf{x}^T A\mathbf{x} < 0$, respectively) for all nonzero $\mathbf{x} \in \mathbb{R}^n$.

If A is not positive or negative definite, it is called an *indefinite matrix*.

Theorem 3.4. (Second partial derivative test) Let $f : X \to \mathbb{R}$ be a function of class C^2 where $X \subset \mathbb{R}^n$. Let $\mathbf{a}$ be an interior point of X which is a critical point of f. Then the following hold.

(a) If $Hf(\mathbf{a})$ is positive definite, then f has a local minimum at $\mathbf{a}$.

(b) If $Hf(\mathbf{a})$ is negative definite, then f has a local maximum at $\mathbf{a}$.

(c) If $\det Hf(\mathbf{a}) \neq 0$ and $Hf(\mathbf{a})$ is indefinite, then $\mathbf{a}$ is a saddle point of f.

Definition 3.8. Let A be an $n \times n$ matrix. For $k = 1, 2, \ldots, n$, let A_k be the $k \times k$ submatrix of A in the top left corner. Then $d_k = \det A_k$ is called a *leading principal minor* of A.

Theorem 3.5. (Sylvester's criterion) Let A be an $n \times n$ real symmetric matrix. Let $d_1, d_2, \ldots, d_n$ be the leading principal minors of A. Then the following hold.

(a) A is positive definite if and only if $d_k > 0$ for $k = 1, 2, \ldots, n$

(b) A is negative definite if and only if $d_k < 0$ for odd k and $d_k > 0$ for even k

Example 3.6. Define

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}, \quad B = \begin{pmatrix} -3 & 1 & 2 \\ 1 & -1 & -1 \\ 2 & -1 & -4 \end{pmatrix}, \quad C = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 3 & -1 \\ 0 & -1 & -2 \end{pmatrix}.$$

Then A is positive definite, B is negative definite, and C is indefinite.

Example 3.7. Define $f : \mathbb{R}^3 \to \mathbb{R}$ by

$$f(x, y, z) = x^3 + xy^2 + x^2 + y^2 + 3z^2.$$

Then $\nabla f(x, y, z) = (3x^2 + y^2 + 2x, 2xy + 2y, 6z)$. The only critical points of f are $(0, 0, 0)$ and $\left(-\frac{2}{3}, 0, 0\right)$. By the second partial derivative test, we find that f has a local minimum at $(0, 0, 0)$, while $\left(-\frac{2}{3}, 0, 0\right)$ is a saddle point.

Definition 3.9. Let X be a subset of $\mathbb{R}^n$. We say that X is *bounded* if there exists $M > 0$ such that $\|\mathbf{x}\| \leqslant M$ for all $\mathbf{x} \in X$.

Definition 3.10. Let X be a subset of $\mathbb{R}^n$. We say that X is *compact* if it is closed and bounded.

Theorem 3.6. (Extreme value theorem) Let $f : X \to \mathbb{R}$ be a continuous function where $X \subset \mathbb{R}^n$ is compact. Then f attains its global maximum and global minimum.

Example 3.8. An experiment is carried out in a device of size 8×8. We use the Cartesian coordinates to represent each point by (x, y) where $-4 \leqslant x, y \leqslant 4$. Suppose the temperature at the point (x, y) is given by

$$x^2 - xy + y^2$$

degree Celsius. Then the point $(0, 0)$ has the lowest temperature, while the points $(4, -4)$ and $(-4, 4)$ have the highest temperature.

Example 3.9. Define $f : \mathbb{R}^2 \to \mathbb{R}$ by

$$f(x,y) = e^{x^2+y^2}\left(x^2 + y^2 - \frac{8}{3}x + 2\right).$$

Then f has a global minimum at $(x,y) = (1,0)$ and it has no global maximum.

3.3 Lagrange Multipliers

Example 3.10. Define $f : X \to \mathbb{R}$ by

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

where $X = \{(x,y) \in \mathbb{R}^2 : x + y = 1\}$. Then the global minimum of f is attained at $(x,y) = (1,0)$.

Theorem 3.7. (Lagrange multiplier rule) Let $f, g : X \to \mathbb{R}$ be functions of class C^1 where $X \subset \mathbb{R}^n$ is open. Let $S = \{\mathbf{x} \in X : g(\mathbf{x}) = c\}$ be the level set of g at height c. If the restriction $f|_S$ of f on S has a local extremum at $\mathbf{a} \in S$ and $\nabla g(\mathbf{a}) \neq \mathbf{0}$, then there exists $\lambda \in \mathbb{R}$ such that

$$\nabla f(\mathbf{a}) = \lambda \nabla g(\mathbf{a}).$$

The restriction $f|_S$ means the function $f|_S : S \to \mathbb{R}$ defined by $f|_S(\mathbf{x}) = f(\mathbf{x})$. This means we want to optimize $f(\mathbf{x})$ given that $g(\mathbf{x}) = c$.

Example 3.11. Define $f : S \to \mathbb{R}$ by

$$f(x,y) = x^3 + y^3$$

where $S = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$. Then the global minimum of f is attained at $(x,y) = (0,-1), (-1,0)$ and the global maximum is attained at $(x,y) = (0,1), (1,0)$.

Example 3.12. A box of volume 4 without lid is to be made. The cost is proportional to the surface area of the box. To minimize the cost, it is the same as minimizing the function

$$f(x,y,z) = xy + 2xz + 2yz$$

under the constraints $xyz = 4$ and $x, y, z > 0$. The global minimum is attained at $(x,y,z) = (2,2,1)$.

Theorem 3.8. (Lagrange multiplier rule) Let $f, g_1, g_2, \dots, g_k : X \to \mathbb{R}$ be functions of class C^1 where $X \subset \mathbb{R}^n$ is open and $k < n$. Let

$$S = \{\mathbf{x} \in X : g_j(\mathbf{x}) = c_j \text{ for } j = 1, 2, \dots, k\}$$

where $c_1, c_2, \dots, c_k$ are constants. If the restriction $f|_S$ of f on S has a local extremum at $\mathbf{a} \in S$ and $\{\nabla g_1(\mathbf{a}), \nabla g_2(\mathbf{a}), \dots, \nabla g_k(\mathbf{a})\}$ is linearly independent, then there exist $\lambda_1, \lambda_2, \dots, \lambda_k \in \mathbb{R}$ such that

$$\nabla f(\mathbf{a}) = \lambda_1 \nabla g_1(\mathbf{a}) + \lambda_2 \nabla g_2(\mathbf{a}) + \cdots + \lambda_k \nabla g_k(\mathbf{a}).$$

A set $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ is said to be *linearly independent* if the only solution to

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_k\mathbf{v}_k = \mathbf{0}$$

where $a_1, a_2, \dots, a_k \in \mathbb{R}$ is $a_1 = a_2 = \cdots = a_k = 0$. In particular, when $k = 2$, this means each of $\mathbf{v}_1$ and $\mathbf{v}_2$ is not a multiple of the other.

Example 3.13. In the Euclidean space, there is a ring which is the intersection of the cylinder $x^2 + y^2 = 3$ and the plane $x + z = 1$. The charge density of the ring at the point (x, y, z) is yz coulombs per cubic unit. Then the charge density is maximized when $(x, y, z) = (-1, \sqrt{2}, 2)$, and is minimized when $(x, y, z) = (-1, -\sqrt{2}, 2)$.

Links

Theorems

- 3.1: First-order Taylor's theorem
- 3.2: Taylor's theorem
- 3.3: A necessary condition for a point to be a local extremum point
- 3.4: Second partial derivative test
- 3.5: Sylvester's criterion
- 3.6: Extreme value theorem
- 3.7: Lagrange multiplier rule (one constraint)
- 3.8: Lagrange multiplier rule (multiple constraints)

Propositions

- 3.1: Taylor's theorem with the remainder term

Terminologies and Notations

- bounded set
- compact set
- critical point
- extrema
- extremum
- global maxima
- global maximum
- global minima
- global minimum
- Hessian Hf
- incremental change Δf
- indefinite matrix
- leading principal minor
- linearly independent
- local maxima
- local maximum
- local minima

- local minimum
- negative definite
- positive definite
- remainder term
- saddle point
- Taylor polynomial
- total differential df