Optimization

Lecture 13

27 December 2024

Index

Unconstrained Optimization

Introduction

1. A point $x^* \in U$ is a **maximum** of F on U if:

$$F(x^*) \geq F(x) \quad \text{for all } x \in U.$$

2. $x^* \in U$ is a **strict maximum** if x^* is a maximum and:

$$F(x^*) > F(x) \quad \text{for all } x \neq x^* \text{ in } U.$$

3. $x^* \in U$ is a **local (or relative) maximum** of F if there exists a ball $B_r(x^*)$ about x^* such that:

$$F(x^*) \geq F(x) \quad \text{for all } x \in B_r(x^*) \cap U.$$

4. $x^* \in U$ is a **strict local maximum** of F if there exists a ball $B_r(x^*)$ about x^* such that:

$$F(x^*) > F(x) \quad \text{for all } x \neq x^* \text{ in } B_r(x^*) \cap U.$$

Unconstrained Optimization

Introduction

- A point x^* is a **local maximum** if there are no nearby points where F takes on a larger value.
- A **global maximum** or **absolute maximum** occurs when x^* is the maximum of F over the entire domain U, not just locally.
- To emphasize precision:
  - We say x^* is a **maximizer** or **maximum point** of F.
  - Alternatively, F has its **maximum value** at x^*.
- The term ”max” is often used as a convenient shorthand for ”maximum.”

First Order Condition

Critical Points and First Order Conditions

- For a function f of one variable, the first-order condition for x^* to be a maximum or minimum is:

$$f'(x^*) = 0,$$

  meaning x^* is a **critical point** of f.
- x^* must lie in the **interior** of the domain of f (not at the endpoints).
- For a function F of n variables, this extends to the partial derivatives:

$$\frac{\partial F}{\partial x_i}(x^*) = 0 \quad \text{for each } i.$$

- x^* is an interior point of the domain of F if there exists a whole ball $B_r(x^*)$ about x^* within the domain.

Theorem 17.1

Let $F : U \to \mathbb{R}$ be a C^1 function defined on a subset U of $\mathbb{R}^n$. If x^* is a local maximum or minimum of F in U and x^* is an interior point of U, then:

$$\frac{\partial F}{\partial}(x^*) = 0 \quad \text{for } i = 1, \ldots, n. \tag{1}$$

First Order Condition

Example: First Order Condition

To find the local maxima and minima of $F(x, y) = x^3 - y^3 + 9xy$, compute the first-order partial derivatives and set them to zero:

$$\frac{\partial F}{\partial x} = 3x^2 + 9y = 0, \quad \frac{\partial F}{\partial y} = -3y^2 + 9x = 0. \tag{2}$$

Solution

- From $\frac{\partial F}{\partial x} = 0$, we find:

  $$y = -\frac{1}{3}x^2.$$

  Substitute this into $\frac{\partial F}{\partial y} = 0$:

  $$0 = -3y^2 + 9x = -3\left(-\frac{1}{3}x^2\right)^2 + 9x = -\frac{1}{3}x^4 + 9x.$$

First Order Condition

Solution (cont.)

- Simplify and solve:

$$x(27 - x^3) = 0 \implies x = 0 \quad \text{or} \quad x = 3.$$

- For $x = 0$, $y = -\frac{1}{3}(0)^2 = 0$, so one critical point is $(0, 0)$.
- For $x = 3$, $y = -\frac{1}{3}(3)^2 = -3$, so the other critical point is $(3, -3)$.

Conclusion

At this stage, the candidates for local maxima or minima of F are:

$$(0, 0) \quad \text{and} \quad (3, -3).$$

Further analysis is required to determine whether these points are maxima or minima.

Second Order Conditions

Definition

As with functions of one variable, the n-vector x^* is a **critical point** of a function $F(x_1, \ldots, x_n)$ if x^* satisfies:

$$\frac{\partial F}{\partial x_i}(x^*) = 0 \quad \text{for } i = 1, \ldots, n. \tag{3}$$

Hessian Matrix and Critical Points

The critical points of $F(x, y) = x^3 - y^3 + 9xy$ in Example 17.1 are $(0, 0)$ and $(3, -3)$. To determine whether these points are maxima or minima, we use the second derivatives of F.

A C^2 function of n variables has n^2 second-order partial derivatives at each point in its domain. These derivatives are combined into the **Hessian matrix** of F:

$$D^2F(x^*) = \begin{pmatrix} \frac{\partial^2 F}{\partial x_1^2}(x^*) & \cdots & \frac{\partial^2 F}{\partial x_1 \partial x_n}(x^*) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 F}{\partial x_n \partial x_1}(x^*) & \cdots & \frac{\partial^2 F}{\partial x_n^2}(x^*) \end{pmatrix}. \tag{4}$$

The Hessian Matrix

Definition

The **Hessian matrix** is a square matrix of second-order partial derivatives of a scalar-valued function. For a function $F(x_1, x_2, \ldots, x_n)$, the Hessian matrix at a point x^* is:

$$D^2F(x^*) = \begin{pmatrix} \frac{\partial^2 F}{\partial x_1^2} & \frac{\partial^2 F}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 F}{\partial x_1 \partial x_n} \\ \frac{\partial^2 F}{\partial x_2 \partial x_1} & \frac{\partial^2 F}{\partial x_2^2} & \cdots & \frac{\partial^2 F}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 F}{\partial x_n \partial x_1} & \frac{\partial^2 F}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 F}{\partial x_n^2} \end{pmatrix}.$$

Optimization

Properties of Hessian Matrix

- The Hessian is **symmetric** if F is C^2 (twice continuously differentiable):

$$\frac{\partial^2 F}{\partial x_i \partial x_j} = \frac{\partial^2 F}{\partial x_j \partial x_i}.$$

- It is an $n \times n$ matrix for a function of n variables.

Hessian Matrix in Optimization

At a critical point x^* (where $\nabla F(x^*) = 0$):

- If $D^2 F(x^*)$ is **positive definite**, x^* is a **local minimum**.
- If $D^2 F(x^*)$ is **negative definite**, x^* is a **local maximum**.
- If $D^2 F(x^*)$ has both positive and negative eigenvalues, x^* is a **saddle point**.

Optimization

Example: Hessian Matrix

Consider the function $F(x, y) = x^3 - y^3 + 9xy$. The first-order partial derivatives are:

$$\frac{\partial F}{\partial x} = 3x^2 + 9y, \quad \frac{\partial F}{\partial y} = -3y^2 + 9x.$$

For the function $F(x, y) = x^3 - y^3 + 9xy$, the Hessian matrix is:

$$D^2F(x, y) = \begin{pmatrix} \frac{\partial^2 F}{\partial x^2} & \frac{\partial^2 F}{\partial x \partial y} \\ \frac{\partial^2 F}{\partial y \partial x} & \frac{\partial^2 F}{\partial y^2} \end{pmatrix} = \begin{pmatrix} 6x & 9 \\ 9 & -6y \end{pmatrix}.$$

Optimization

Example 2: Hessian Matrix

Consider the function $F(x, y) = x^2 + y^2 - 4xy$. The first-order partial derivatives are:

$$\frac{\partial F}{\partial x} = 2x - 4y, \quad \frac{\partial F}{\partial y} = 2y - 4x.$$

For the function $F(x, y) = x^2 + y^2 - 4xy$, the Hessian matrix is:

$$D^2F(x, y) = \begin{pmatrix} \frac{\partial^2 F}{\partial x^2} & \frac{\partial^2 F}{\partial x \partial y} \\ \frac{\partial^2 F}{\partial y \partial x} & \frac{\partial^2 F}{\partial y^2} \end{pmatrix} = \begin{pmatrix} 2 & -4 \\ -4 & 2 \end{pmatrix}.$$

Unconstrained Optimization

Sufficient Conditions

1. The second-order condition for a critical point x^* of a function f on $\mathbb{R}^1$ to be a **maximum** is that the second derivative $f''(x^*)$ be negative.
2. For a function F of n variables, the second-order condition is that the second derivative $D^2F(x^*)$ be **negative definite** as a symmetric matrix at the critical point x^*.
3. Similarly, the second-order condition for a critical point x^* of a function f of one variable to be a **local minimum** is that $f''(x^*)$ be positive.
4. The analogous second-order condition for an n-dimensional critical point x^* to be a **local minimum** is that the Hessian of F at x^*, $D^2F(x^*)$, be **positive definite**.

Unconstrained Optimization

Theorem 17.2

Let $F : U \to \mathbb{R}^1$ be a C^2 function whose domain is an open set U in $\mathbb{R}^n$. Suppose that x^* is a critical point of F. Then:

1. If the Hessian $D^2F(x^*)$ is a **negative definite symmetric matrix**, then x^* is a **strict local maximum** of F.
2. If the Hessian $D^2F(x^*)$ is a **positive definite symmetric matrix**, then x^* is a **strict local minimum** of F.
3. If $D^2F(x^*)$ is **indefinite**, then x^* is neither a local maximum nor a local minimum of F.

Definition: Saddle Point

A critical point x^* of F for which the Hessian $D^2F(x^*)$ is **indefinite** is called a **saddle point** of F.

It is graph is saddle shaped represented by:

$$F(x_1, x_2) = x_1^2 - x_2^2,$$

Unconstrained Optimization

Theorem 17.3

Let $F : U \to \mathbb{R}^1$ be a C^2 function whose domain is an open set U in $\mathbb{R}^n$. Suppose that:

$$\frac{\partial F}{\partial x_i}(x^*) = 0 \quad \text{for } i = 1, \ldots, n,$$

and that the n leading principal minors of $D^2F(x^*)$ alternate in sign:

$$\begin{vmatrix} F_{x_1x_1} \end{vmatrix} < 0,$$

$$\begin{vmatrix} F_{x_1x_1} & F_{x_1x_2} \\ F_{x_2x_1} & F_{x_2x_2} \end{vmatrix} > 0,$$

$$\begin{vmatrix} F_{x_1x_1} & F_{x_1x_2} & F_{x_1x_3} \\ F_{x_2x_1} & F_{x_2x_2} & F_{x_2x_3} \\ F_{x_3x_1} & F_{x_3x_2} & F_{x_3x_3} \end{vmatrix} < 0, \ldots$$

at x^*. Then, x^* is a **strict local maximum** of F.

Unconstrained Optimization

Theorem 17.4

Let $F : U \to \mathbb{R}^1$ be a C^2 function whose domain is an open set U in $\mathbb{R}^n$. Suppose that:

$$\frac{\partial F}{\partial x_i}(x^*) = 0 \quad \text{for } i = 1, \dots, n,$$

and that the n leading principal minors of $D^2F(x^*)$ are all positive:

$$\begin{vmatrix} F_{x_1x_1} \end{vmatrix} > 0,$$

$$\begin{vmatrix} F_{x_1x_1} & F_{x_1x_2} \\ F_{x_2x_1} & F_{x_2x_2} \end{vmatrix} > 0,$$

$$\begin{vmatrix} F_{x_1x_1} & F_{x_1x_2} & F_{x_1x_3} \\ F_{x_2x_1} & F_{x_2x_2} & F_{x_2x_3} \\ F_{x_3x_1} & F_{x_3x_2} & F_{x_3x_3} \end{vmatrix} > 0, \dots$$

at x^*. Then, x^* is a **strict local minimum** of F.

Unconstrained Optimization

Theorem 17.5

Let $F : U \to \mathbb{R}^1$ be a C^2 function whose domain is an open set U in $\mathbb{R}^n$. Suppose that:

$$\frac{\partial F}{\partial x_i}(x^*) = 0 \quad \text{for } i = 1, \ldots, n,$$

and that some nonzero leading principal minors of $D^2F(x^*)$ violate the sign patterns in the hypotheses of Theorems 17.3 and 17.4. Then, x^* is a **saddle point** of F; it is neither a local maximum nor a local minimum.

Principal Minors

Definition

Let A be an $n \times n$ matrix. A $k \times k$ submatrix of A formed by deleting $n - k$ columns, say columns $i_1, i_2, \ldots, i_{n-k}$, and the same $n - k$ rows, rows $i_1, i_2, \ldots, i_{n-k}$, from A is called a k**-th order principal submatrix** of A. The determinant of a $k \times k$ principal submatrix is called a k**-th order principal minor** of A.

Optimization

Example: Principal Minors

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}.$$

- There is one **third-order principal minor**: $\det(A)$.
- There are three **second-order principal minors**:
  1. $\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$, formed by deleting column 3 and row 3 from A.
  2. $\begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix}$, formed by deleting column 2 and row 2 from A.
  3. $\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}$, formed by deleting column 1 and row 1 from A.
- There are three **first-order principal minors**:
  1. $|a_{11}|$, formed by deleting the last 2 rows and columns.
  2. $|a_{22}|$, formed by deleting the first and third rows and the first and third columns.
  3. $|a_{33}|$, formed by deleting the first 2 rows and columns.

Leading Principal Minors

Definition

Let A be an $n \times n$ matrix. The k-th order principal submatrix of A obtained by deleting the **last** $n-k$ **rows** and the **last** $n-k$ **columns** from A is called the **k-th order leading principal submatrix** of A.
Its determinant is called the **k-th order leading principal minor** of A. We will denote the k-th order leading principal submatrix by A_k and the corresponding leading principal minor by $|A_k|$.
An $n \times n$ matrix has n leading principal submatrices:

- the top-leftmost 1×1 submatrix,
- the top-leftmost 2×2 submatrix,
- etc.

For the general 3×3 matrix of Example 16.2, the three leading principal minors are:

$$|a_{11}|, \quad \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}, \quad \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}.$$

Leading Principal Minors

Definition

Let A be an $n \times n$ matrix. The k-th order principal submatrix of A obtained by deleting the **last** $n - k$ **rows** and the **last** $n - k$ **columns** from A is called the **k-th order leading principal submatrix** of A.
Its determinant is called the **k-th order leading principal minor** of A. We will denote the k-th order leading principal submatrix by A_k and the corresponding leading principal minor by $|A_k|$.
An $n \times n$ matrix has n leading principal submatrices:

- the top-leftmost 1×1 submatrix,
- the top-leftmost 2×2 submatrix,
- etc.

For the general 3×3 matrix of Example 16.2, the three leading principal minors are:

$$|a_{11}|, \quad \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}, \quad \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}.$$

Unconstrained Optimization

Example: Critical Points

1. Compute the first-order partial derivatives:

$$\frac{\partial F}{\partial x} \quad \text{and} \quad \frac{\partial F}{\partial y}.$$

2. Set the partial derivatives to zero:

$$\frac{\partial F}{\partial x} = 0, \quad \frac{\partial F}{\partial y} = 0.$$

Consider the function $F(x, y) = x^3 - y^3 + 9xy$:

$$\frac{\partial F}{\partial x} = 3x^2 + 9y, \quad \frac{\partial F}{\partial y} = -3y^2 + 9x.$$

$$3x^2 + 9y = 0, \quad -3y^2 + 9x = 0.$$

Critical points are:

$$(0, 0) \quad \text{and} \quad (3, -3).$$

Unconstrained Optimization

Leading Principal Minors (LPMs)

To find the k-th order Leading Principal Minor (LPM):

1. Extract the top-left $k \times k$ submatrix from the Hessian matrix H.
2. Compute the determinant of this $k \times k$ submatrix.

Example (cont.)

At $(0,0)$:

$$H(0,0) = \begin{pmatrix} 0 & 9 \\ 9 & 0 \end{pmatrix}.$$

- **1st LPM**: Determinant of the top-left 1×1 submatrix:

$$\text{1st LPM} = \det\begin{pmatrix} 0 \end{pmatrix} = 0.$$

- **2nd LPM**: Determinant of the entire 2×2 matrix:

$$\text{2nd LPM} = \det\begin{pmatrix} 0 & 9 \\ 9 & 0 \end{pmatrix} = (0)(0) - (9)(9) = -81.$$

Unconstrained Optimization

Example (cont.)

- The Hessian is **indefinite**, so $(0,0)$ is a **saddle point**.

At $(3,-3)$:

$$H(3,-3) = \begin{pmatrix} 18 & 9 \\ 9 & 18 \end{pmatrix}.$$

- **1st LPM**: Determinant of the top-left 1×1 submatrix:

$$\text{1st LPM} = \det\left(18\right) = 18.$$

- **2nd LPM**: Determinant of the entire 2×2 matrix:

$$\text{2nd LPM} = \det \begin{pmatrix} 18 & 9 \\ 9 & 18 \end{pmatrix} = (18)(18) - (9)(9) = 243.$$

- The Hessian is **positive definite**, so $(3,-3)$ is a **strict local minimum**.