CHAPTER 7

SYMMETRIC MATRICES AND QUADRATIC FORMS

Symmetric matrices arise more often in applications than any other major class of matrices. They are diagonalizable and this property is related to the quadratic forms, and optimization problems.

7.1 Orthogonal Matrices

In Chapter 6 we observed that is a square matrix of size n has exactly n linearly independent eigenvectors, then it is diagonalizable. In this section we consider classes of matrices that are actually diagonalizable.

- **Orthogonal matrix**

  A square matrix A is said to be orthogonal if

$$AA^T = A^T A = I$$

  or equivalently $A^T = A^{-1}$.

By definition it is easy to conclude that if A is orthogonal, then $\det(A) = \mp 1$.

Example: The standard matrix

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

for the rotation operator in $\mathbb{R}^2$ is orthogonal since

$$\begin{aligned}
\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}^T \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} &= \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \\
&= \begin{bmatrix} \cos^2\theta + \sin^2\theta & -\sin\theta\cos\theta + \sin\theta\cos\theta \\ -\sin\theta\cos\theta + \sin\theta\cos\theta & \cos^2\theta + \sin^2\theta \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I. \blacksquare
\end{aligned}$$

- **Orthogonally diagonalizable matrix**

  A matrix A is said to be orthogonally diagonalizable if there exists an orthogonal matrix S such that $S^{-1}AS = S^T AS$ is diagonalizable.

By definition, if A is orthogonally diagonalizable matrix, then

$$S^{-1}AS = D \Rightarrow A = SDS^{-1} = SDS^T$$

for an orthogonal matrix S and a diagonal matrix D. This implies that

$$A^T = \left(SDS^T\right)^T = SD^TS^T = SDS^T = A.$$

Thus, A is a symmetric matrix. Surprisingly, the converse is true as well.

- **Spectral theorem**

  A matrix A is orthogonally diagonalizable if and only if A is a symmetric matrix.

Example: Show that for every symmetric matrix A of size n there is a symmetric matrix B of the same size such that $B^3 = A$.

Solution: The spectral theorem says that there exist an orthogonal matrix S and a diagonal matrix D such that $S^{-1}AS = D$. Let $D_1^3 = D$. Then, since

$$B^3 = \left(S^{-1}D_1S\right)^3 = S^{-1}D_1^3S = S^{-1}DS = A,$$

$B = S^{-1}D_1S$ is the desired matrix.■

Example: Orthogonally diagonalize the matrix

$$A = \begin{bmatrix} 6 & -2 & -1 \\ -2 & 6 & -1 \\ -1 & -1 & 5 \end{bmatrix}$$

if possible.

Solution: We first find an eigenbasis. We have

$$\det(\lambda I - A) = \begin{vmatrix} \lambda - 6 & 2 & 1 \\ 2 & \lambda - 6 & 1 \\ 1 & 1 & \lambda - 5 \end{vmatrix} = \lambda^3 - 17\lambda^2 + 90\lambda - 144 = (\lambda - 3)(\lambda - 6)(\lambda - 8),$$

so $\lambda = 3$, $\lambda = 6$, and $\lambda = 8$ are eigenvalues. For $\lambda = 3$, we have

$$3I - A = \begin{bmatrix} -3 & 2 & 1 \\ 2 & -3 & 1 \\ 1 & 1 & 2 \end{bmatrix} \overset{\text{E.R.O.}}{\to} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}$$

so a corresponding eigenvector is

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

For $\lambda = 6$, we have

$$6I - A = \begin{bmatrix} 0 & 2 & 1 \\ 2 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix} \overset{\text{E.R.O.}}{\to} \begin{bmatrix} 1 & 0 & \frac{1}{2} \\ 0 & 1 & \frac{1}{2} \\ 0 & 0 & 0 \end{bmatrix}$$

so a corresponding eigenvector is

$$\mathbf{v}_2 = \begin{bmatrix} -1 \\ -1 \\ 2 \end{bmatrix}.$$

For $\lambda = 8$, we have

$$8I - A = \begin{bmatrix} 2 & 2 & 1 \\ 2 & 2 & 1 \\ 1 & 1 & 3 \end{bmatrix} \overset{\text{E.R.O.}}{\to} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

so a corresponding eigenvector is

$$\mathbf{v}_3 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}.$$

These vectors are mutually orthogonal. For calculation concerns we normalize them by dividing them to their norms. Thus, we have

$$\mathbf{u}_1 = \begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} -1/\sqrt{6} \\ -1/\sqrt{6} \\ 2/\sqrt{6} \end{bmatrix}, \quad \mathbf{u}_3 = \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix}$$

respectively. Let

$$S = \begin{bmatrix} 1/\sqrt{3} & -1/\sqrt{6} & -1/\sqrt{2} \\ 1/\sqrt{3} & -1/\sqrt{6} & 1/\sqrt{2} \\ 1/\sqrt{3} & 2/\sqrt{6} & 0 \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 8 \end{bmatrix}.$$

Then, $A = S^{-1}DS$, where S is an orthogonal matrix.■

Orthogonality of eigenvectors in the above example is not a coincidence.

- **Orthogonality of eigenvectors**

  If A is symmetric, then any two eigenvectors from different eigenspaces are orthogonal.

Proof: We compute the product $\mathbf{v}_1^T A \mathbf{v}_2$ in two different ways:

$$\mathbf{v}_1^T A \mathbf{v}_2 = \mathbf{v}_1^T (\lambda_2 \mathbf{v}_2) = \lambda_2 (\mathbf{v}_1 \cdot \mathbf{v}_2)$$
$$\mathbf{v}_1^T A \mathbf{v}_2 = \mathbf{v}_1^T A^T \mathbf{v}_2 = (A\mathbf{v}_1)^T \mathbf{v}_2 = (\lambda_1 \mathbf{v}_1)^T \mathbf{v}_2 = \lambda_1 (\mathbf{v}_1 \cdot \mathbf{v}_2)$$

Thus

$$(\lambda_1 - \lambda_2)(\mathbf{v}_1 \cdot \mathbf{v}_2) = 0 \Rightarrow \mathbf{v}_1 \cdot \mathbf{v}_2 = 0$$

since $\lambda_1 \neq \lambda_2$.■

7.2 Quadratic Forms

Quadratic forms occur frequently in applications of linear algebra to engineering and signal processing. Some of the mathematical background for such applications follows from symmetric matrices.

- **Quadratic form**

  *A **quadratic form** on R^n is a function defined by*

  $$Q(\mathbf{x}) = \mathbf{x}^T A \mathbf{x}, \quad \mathbf{x} \in \mathbb{R}^n,$$

  *where A is an $n \times n$ symmetric matrix. The matrix A is called the **matrix of the quadratic form**.*

The simplest example of a nonzero quadratic form is

$$Q(\mathbf{x}) = \mathbf{x}^T I \mathbf{x} = \|\mathbf{x}\|^2.$$

Example: Let $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$. Compute $\mathbf{x}^T A \mathbf{x}$ for

(a) $A = \begin{bmatrix} 4 & 0 \\ 0 & 3 \end{bmatrix}$ **(b)** $A = \begin{bmatrix} 3 & -2 \\ -2 & 7 \end{bmatrix}$

Solution:

(a) We have

$$\mathbf{x}^T A \mathbf{x} = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 4x_1^2 + 3x_2^2.$$

(b) We have

$$\mathbf{x}^T A \mathbf{x} = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 3 & -2 \\ -2 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 3x_1^2 + 7x_2^2 - 4x_1x_2.$$■

From the above example we note that the quadratic form associated with the diagonal matrix has no **cross-product** terms like x_1x_2.

Example: For $x \in \mathbb{R}^3$ let $Q(\mathbf{x}) = 5x_1^2 + 3x_2^2 + 2x_3^2 - x_1x_2 + 8x_2x_3$. Write this quadratic form as $\mathbf{x}^T A\mathbf{x}$.

Solution: The coefficients of x_1^2, x_2^2, x_3^2 are on the main diagonal of A. To make A symmetric, the coefficient x_ix_j for $i \neq j$ must be split evenly between the (i,j)th and (j,i)th entries in A. The coefficient of x_1x_3 is 0, so

$$Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x} = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} 5 & -\frac{1}{2} & 0 \\ -\frac{1}{2} & 3 & 4 \\ 0 & 4 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}. \blacksquare$$

If $\mathbf{x}$ represents a variable vector in $\mathbb{R}^n$, then a **change of variable** is an equation of the form

$$\mathbf{x} = S\mathbf{y} \quad \text{or equivalently} \quad \mathbf{y} = S^{-1}\mathbf{x},$$

where S is an invertible matrix and $\mathbf{y}$ is a new variable vector in $\mathbb{R}^n$. If this change of variable is made in a quadratic form $\mathbf{x}^T A\mathbf{x}$, then

$$\mathbf{x}^T A\mathbf{x} = (S\mathbf{y})^T A(S\mathbf{y}) = \mathbf{y}^T S^T AS\mathbf{y} = \mathbf{y}^T \left(S^T AS\right)\mathbf{y},$$

and the new variable matrix of the quadratic form is $S^T AS$. Since A is symmetric, we have S is orthogonal and $S^T AS$ is a diagonal matrix D. Thus, we obtain that

$$\mathbf{x}^T A\mathbf{x} = \mathbf{y}^T D\mathbf{y},$$

where D is diagonal. In particular, if the quadratic form $\mathbf{x}^T A\mathbf{x}$ has cross-product terms, then by the change of variable $\mathbf{x} = S\mathbf{y}$, we have a quadratic form without any cross-product terms.

Example: Consider the matrix

$$A = \begin{bmatrix} 1 & -4 \\ -4 & 5 \end{bmatrix}.$$

Its eigenvalues are $\lambda = 3$ and $\lambda = -7$. Associated normalized eigenvectors are

$$\begin{bmatrix} 2/\sqrt{5} \\ -1/\sqrt{5} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix},$$

respectively. These eigenvalues are orthogonal. Let

$$S = \begin{bmatrix} 2/\sqrt{5} & 1/\sqrt{5} \\ -1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} 3 & 0 \\ 0 & -7 \end{bmatrix}.$$

Then $A = SDS^{-1}$. A suitable change of variable is

$$\mathbf{x} = S\mathbf{y}, \quad \text{where } \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad \text{and} \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}.$$

Then

$$x_1^2 - 8x_1x_2 - 5x_2^2 = \mathbf{x}^T A\mathbf{x} = (S\mathbf{y})^T A\,(S\mathbf{y}) = \mathbf{y}^T S^T AS\mathbf{y} = \mathbf{y}^T D\mathbf{y} = 3y_1^2 - 7y_2^2.\blacksquare$$

- **Positive definite, negative definite, and indefinite quadratic forms** *A quadratic form Q is*

  (a) ***positive definite*** *if $Q(\mathbf{x}) > 0$ for all $\mathbf{x} \neq \mathbf{0}$*

  (b) ***negative definite*** *if $Q(\mathbf{x}) < 0$ for all $\mathbf{x} \neq \mathbf{0}$*

  (c) ***indefinite*** *if $Q(\mathbf{x})$ assumes both positive and negative values.*

Also, Q is said to be **positive semidefinite** if $Q(\mathbf{x}) \geq 0$ for all $\mathbf{x}$, and **negative semidefinite** if $Q(\mathbf{x}) \leq 0$ for all $\mathbf{x}$.

Example: Let A be a skew-symmetric matrix. Show that A^2 is symmetric and determine definiteness of A^2.

Solution: Since

$$\left(A^2\right)^T = \left(A^T\right)^2 = (-A)^2 = A^2,$$

A^2 is symmetric. Now,

$$Q(\mathbf{x}) = \mathbf{x}^T A^2\mathbf{x} = \mathbf{x}^T AA\mathbf{x} = -\mathbf{x}^T A^T A\mathbf{x} = -(A\mathbf{x})^T A\mathbf{x} = -(A\mathbf{x}) \cdot (A\mathbf{x}) \leq 0$$

for all $\mathbf{x}$, so A^2 is negative semidefinite.$\blacksquare$

- **Quadratic forms and eigenvalues** *Let A be an $n \times n$ symmetric matrix. A quadratic form $\mathbf{x}^T A\mathbf{x}$ is*

  (a) *positive definite if and only if the eigenvalues of A are all positive.*

  (b) *negative definite if and only if the eigenvalues of A are all negative.*

  (c) *indefinite if and only if A has both positive and negative eigenvalues.*

Proof: Since A is symmetric, there exists an orthogonal change of variable $\mathbf{x} = S\mathbf{y}$ such that

$$Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x} = \mathbf{y}^T D\mathbf{y} = \lambda_1 x_1^2 + \lambda_2 x_2^2 + \cdots + \lambda_n x_n^2,$$

where $\lambda_1, \lambda_2, \ldots, \lambda_n$ are eigenvalues of A.$\blacksquare$

Example: Show that for any positive definite matrix A of size n can be written as $A = BB^T$, where B is of size n with orthogonal columns.

Solution: Since A is positive definite, its eigenvalues are all positive. Thus, $A = SDS^T$, where S is an orthogonal matrix and D is a diagonal matrix with positive diagonal entries. Then there exists a diagonal matrix D_1 such that $D_1^2 = D$. Now,

$$A = SDS^T = SD_1^2S^T = SD_1D_1S^T = SD_1\left(SD_1\right)^T.$$

Letting $B = SD_1$ gives the desired result.■

7.3 Constrained Optimization

It is often needed to find the maximum or minimum value of a quadratic for $Q(\mathbf{x})$ for $\mathbf{x}$ in some specified set. Typically, the problem can be arranged so that $\mathbf{x}$ varies over the set of unit vectors. This problem is called the **constrained optimization** problem that has lots of applications.

The requirement that a vector $\mathbf{x}$ in $\mathbb{R}^n$ be a unit vector can be stated in several equivalent ways:

$$\|\mathbf{x}\| = 1, \quad \|\mathbf{x}\|^2 = 1, \quad \mathbf{x}^T\mathbf{x} = 1, \quad x_1^2 + x_2^2 + \cdots + x_n^2 = 1.$$

The last one is commonly used in applications.

When a quadratic form Q has no cross-product terms, it is easy to find the extrema of $Q(\mathbf{x})$ for $\mathbf{x}^T\mathbf{x} = 1$.

Example: Find the maximum and minimum values of $Q(\mathbf{x}) = 9x_1^2 + 4x_2^2 + 3x_3^2$ subject to the constraint $\mathbf{x}^T\mathbf{x} = 1$.

Solution: We note that

$$4x_2^2 \le 9x_2^2 \quad \text{and} \quad 3x_3^2 \le 9x_3^2.$$

Hence

$$Q(\mathbf{x}) = 9x_1^2 + 4x_2^2 + 3x_3^2 \le 9x_1^2 + 9x_2^2 + 9x_3^2 = 9\left(x_1^2 + x_2^2 + x_3^2\right) = 9$$

whenever $x_1^2 + x_2^2 + x_3^2 = 1$. So the maximum value of $Q(\mathbf{x})$ cannot be larger than 9 when $\mathbf{x}$ is a unit vector. Furthermore, $Q(\mathbf{x}) = 9$ when $\mathbf{x} = (1, 0, 0)$. Thus, 9 is the maximum value of $Q(\mathbf{x})$ for $\mathbf{x}^T\mathbf{x} = 1$.

To find the minimum value of $Q(\mathbf{x})$, we observe that

$$9x_1^2 \ge 3x_1^2 \quad \text{and} \quad 4x_2^2 \ge 3x_2^2,$$

and hence

$$Q(\mathbf{x}) = 9x_1^2 + 4x_2^2 + 3x_3^2 \geq 3x_1^2 + 3x_2^2 + 3x_3^2 = 3\left(x_1^2 + x_2^2 + x_3^2\right) = 3$$

whenever $x_1^2 + x_2^2 + x_3^2 = 1$. Also, $Q(\mathbf{x}) = 3$ when $\mathbf{x} = (0, 0, 1)$. So 3 is the minimum value of $Q(\mathbf{x})$ when $\mathbf{x}^T\mathbf{x} = 1$.■

For any symmetric matrix A, the set of all possible values of $\mathbf{x}^T A\mathbf{x}$ for $\|\mathbf{x}\| = 1$ is a closed interval on the real axis. We denote the left and right endpoints of this interval by m and M, respectively. That is, let

$$m = \min\left\{\mathbf{x}^T A\mathbf{x} : \|\mathbf{x}\| = 1\right\} \quad \text{and} \quad M = \max\left\{\mathbf{x}^T A\mathbf{x} : \|\mathbf{x}\| = 1\right\}.$$

If λ is an eigenvalue of A, then $m \leq \lambda \leq M$. In fact, m and M are themselves eigenvalues.

- **Extrema and eigenvalues**

  Let A be a symmetric matrix and define

  $$m = \min\left\{\mathbf{x}^T A\mathbf{x} : \|\mathbf{x}\| = 1\right\} \quad \textit{and} \quad M = \max\left\{\mathbf{x}^T A\mathbf{x} : \|\mathbf{x}\| = 1\right\}.$$

  Then M is the greatest eigenvalue and m is the least eigenvalue of A.

Example: Find the maximum value of the quadratic form $\mathbf{x}^T A\mathbf{x}$ subject to the constraint $\mathbf{x}^T\mathbf{x} = 1$ when

$$A = \begin{bmatrix} 3 & 2 & 1 \\ 2 & 3 & 1 \\ 1 & 1 & 4 \end{bmatrix}.$$

Find also a unit vector at which this maximum value is assumed.

Solution: The desired maximum value is the greatest eigenvalue of A. Since

$$\det(\lambda I - A) = \begin{vmatrix} \lambda - 3 & -2 & -1 \\ -2 & \lambda - 3 & -1 \\ -1 & -1 & \lambda - 4 \end{vmatrix} = (\lambda - 6)(\lambda - 3)(\lambda - 1),$$

the greatest eigenvalue is 6. Since

$$6I - A = \begin{bmatrix} 3 & -2 & -1 \\ -2 & 3 & -1 \\ -1 & -1 & 2 \end{bmatrix} \overset{\text{E.R.O.}}{\longrightarrow} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}$$

so a corresponding eigenvector is

$$\mathbf{v} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

Thus the unit vector at which the maximum value 6 attained is

$$\mathbf{u} = \begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}. \blacksquare$$

In the next result the values $\mathbf{x}^T A\mathbf{x}$ are computed with additional constraints on the unit vector $\mathbf{x}$.

- **Two constraints**

  Let A be a symmetric matrix, λ_1 be its greatest eigenvalue, and $\mathbf{u}_1$ be the corresponding unit eigenvector. The maximum value of $x^T Ax$ subject to the constraints

  $$\mathbf{x}^T\mathbf{x} = 1 \quad \textit{and} \quad \mathbf{x}^T\mathbf{u}_1 = 0$$

  is the second greatest eigenvalue of A.

Example: Find the maximum value of $Q(\mathbf{x}) = 9x_1^2 + 4x_2^2 + 3x_3^2$ subject to the constraints $\mathbf{x}^T\mathbf{x} = 1$ and $\mathbf{x}^T\mathbf{u}_1 = 0$, where $\mathbf{u}_1 = (1, 0, 0)$.

Solution: If $\mathbf{x} = (x_1, x_2, x_3)$, then $\mathbf{x}^T\mathbf{u}_1 = 0$ implies that $x_1 = 0$. Since $\mathbf{x}$ is a unit vector, we must then have $x_2^2 + x_3^2 = 1$. So

$$9x_1^2 + 4x_2^2 + 3x_3^2 = 4x_2^2 + 3x_3^2 \leq 4x_2^2 + 4x_3^2 = 4.$$

Thus, the constrained maximum of the quadratic form is not larger than 4. This value is assumed when $\mathbf{x} = (0, 1, 0)$. Note that 4 is the second largest eigenvalue of the matrix of the quadratic form.$\blacksquare$

7.4 Applications in Multivariate Calculus

In multivariate calculus, an important problem is to find extremum values of a function of several variables. In particular, consider the function $z = f(x, y)$ of two variables (this is in fact a transformation from $\mathbb{R}^2$ to $\mathbb{R}$). We want to find conditions under which f has a local extremum value at (x_0, y_0). By considering $f(x + x_0, y + y_0)$ we may take this point to be $(0, 0)$. Then, replacing $f(x, y)$ by $f(x, y) - f(0, 0)$, we may assume that $f(0, 0) = 0$.

Polynomials are easy to differentiate and evaluate, and we like to use them to approximate other functions. This is the content of Taylor's expansion:

$$f(x) = \sum_{k=1}^{n} \frac{f^{(k)}(a)}{k!}(x-a)^k + \frac{f^{(n+1)}(t)}{(n+1)!}(x-a)^{n+1}, \quad t \in (a, x).$$

In multivariate calculus we also study Taylor's expansion. If $f(x,y)$ has Taylor's expansion at $(0,0)$, then we write

$$f(x,y) = a_{00} + a_{10}x + a_{01}y + a_{20}x^2 + a_{11}xy + a_{02}y^2 + \{\text{higher-order terms}\},$$

where

$$\begin{aligned}
a_{00} &= f(0,0) = 0 \\
a_{10} &= f_x(0,0) = \frac{\partial f}{\partial x}(0,0), \quad a_{01} = f_y(0,0) = \frac{\partial f}{\partial y}(0,0), \\
a_{20} &= \frac{1}{2} f_{xx}(0,0) = \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(0,0), \quad a_{11} = f_{xy}(0,0) = \frac{\partial^2 f}{\partial x \partial y}(0,0), \quad a_{02} = f_{yy}(0,0) = \frac{\partial^2 f}{\partial y^2}(0,0).
\end{aligned}$$

Here, $f_x(x,y)$ means that we differentiate f with respect to x by assuming that y is a constant with respect to x. This is called **partial differentiation**. On the other hand, the second-order mixed partial derivative $f_{xy}(x,y)$ means that we differentiate f first with respect to x then with respect to y (all the mentioned derivatives are assumed to exist). The higher-order terms involve products of powers of x and y, so they tend to 0 faster near $(0,0)$, and usually do not affect the local behavior of f.

If $(0,0)$ is a critical point, then f_x and f_y are zero at $(0,0)$, that is, $a_{10} = a_{01} = 0$. Thus, the problem of understanding the extremum values of f is equivalent to understanding the quadratic forms

$$f(x,y) = ax^2 + 2bxy + cy^2,$$

where $a = 2a_{20} = f_{xx}(0,0)$, $b = a_{11} = f_{xy}(0,0)$, and $c = 2a_{02} = f_{yy}(0,0)$.

Since $f(0,0) = 0$, the point $(0,0)$ is

· a **local minimum** when $f(x,y) > f(0,0) = 0$ for all $(x,y) \neq (0,0)$, that is, when the quadratic form is positive definite.

· a **local maximum** when $f(x,y) < f(0,0) = 0$ for all $(x,y) \neq (0,0)$, that is, when the quadratic form is negative definite.

· a **saddle point** when $f(x,y) > 0$ for some (x,y) and $f(x,y) < 0$ for others, that is, when the quadratic form is indefinite.

The quadratic form

$$f(x,y) = ax^2 + 2bxy + cy^2$$

can be written as

$$f(x,y) = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \mathbf{x}^T A \mathbf{x},$$

where

$$\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix} \quad \text{and} \quad A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}.$$

Since $a = f_{xx}(0,0)$, $b = f_{xy}(0,0)$, and $c = f_{yy}(0,0)$, the matrix A for the quadratic form f is

$$A = \begin{bmatrix} f_{xx}(0,0) & f_{xy}(0,0) \\ f_{xy}(0,0) & f_{yy}(0,0) \end{bmatrix}.$$

Note that A is symmetric.

Now, if $a = 0$, then, for example, $f(1,0) = 0$, so f is neither positive definite nor negative definite nor indefinite. Thus, we assume that $a \neq 0$. By completing the square, we write

$$f(x,y) = a\left(a + \frac{b}{a}y\right)^2 + \left(c - \frac{b^2}{a}\right)y^2.$$

Therefore, we conclude that

· f has a local minimum if and only if $a > 0$ and $ac - b^2 > 0$.

· f has a local maximum if and only if $a < 0$ and $ac - b^2 > 0$.

· f has a saddle point if and only if $ac - b^2 < 0$.

This gives the **second derivative test** for extremum values for $f(x,y)$:

· f has a local minimum if and only if $f_{xx} > 0$ and $f_{xx}f_{yy} - f_{xy}^2 > 0$.

· f has a local maximum if and only if $f_{xx} < 0$ and $f_{xx}f_{yy} - f_{xy}^2 > 0$.

· f has a saddle point if and only if $f_{xx}f_{yy} - f_{xy}^2 < 0$.

Example: Find extremum points of $f(x,y) = 2x^3 + y^3 - 3x^2 - 12x - 3y$.

Solution: We have

$$f_x = 6x^2 - 6x - 12 = 6(x+1)(x-2) = 0 \Rightarrow x = -1 \quad \text{and} \quad x = 2$$

$$f_y = 3y^2 - 3 = 3(y+1)(y-1) = 0 \Rightarrow y = -1 \quad \text{and} \quad y = 1$$

so $(-1,-1)$, $(-1,1)$, $(2,-1)$, and $(2,1)$ are critical points. $f_{xx} = 12x - 6$, $f_{yy} = 6y$, and $f_{xy} = 0$, so $(-1,-1)$ is a local maximum point, $(2,1)$ is a local minimum point, and $(-1,1)$ and $(2,-1)$ are saddle points by the second derivative test.■

In general, for a function $f : \mathbb{R}^n \to \mathbb{R}$, we have Taylor's expansion whose terms are of degree two or less of the form

$$T_2(\mathbf{x}) = f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a}) + \frac{1}{2}(\mathbf{x} - \mathbf{a})^T \nabla^2 f(\mathbf{a})(\mathbf{x} - \mathbf{a}),$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and $\mathbf{a} = (a_1, a_2, \ldots, a_n)$, and

$$\nabla f(\mathbf{a}) = \begin{bmatrix} f_{x_1}(\mathbf{a}) \\ f_{x_2}(\mathbf{a}) \\ \vdots \\ f_{x_n}(\mathbf{a}) \end{bmatrix} \quad \text{and} \quad \nabla^2 f(\mathbf{a}) = \begin{bmatrix} f_{x_1x_1}(\mathbf{a}) & f_{x_1x_2}(\mathbf{a}) & f_{x_1x_3}(\mathbf{a}) & \cdots & f_{x_1x_n}(\mathbf{a}) \\ f_{x_2x_1}(\mathbf{a}) & f_{x_2x_2}(\mathbf{a}) & f_{x_2x_3}(\mathbf{a}) & \cdots & f_{x_2x_n}(\mathbf{a}) \\ f_{x_3x_1}(\mathbf{a}) & f_{x_3x_2}(\mathbf{a}) & f_{x_3x_3}(\mathbf{a}) & \cdots & f_{x_3x_n}(\mathbf{a}) \\ \vdots & \vdots & \vdots & & \vdots \\ f_{x_nx_1}(\mathbf{a}) & f_{x_nx_2}(\mathbf{a}) & f_{x_nx_3}(\mathbf{a}) & \cdots & f_{x_nx_n}(\mathbf{a}) \end{bmatrix}.$$

We call ∇f and $\nabla^2 f$ the **gradient vector** and the **Hessian matrix** of f, respectively. Note that the Hessian matrix is symmetric provided that $f_{x_ix_j}(\mathbf{a}) = f_{x_jx_i}(\mathbf{a})$ for all $i \neq j$. This is always the case when f has continuous second-order mixed partial derivatives.