MMAT5390: Mathematical Image Processing

Chapter 2: Review of Basic Mathematical Concepts

As described in the last chapter, a digital image can be considered as a big two-dimensional matrix. As a image resolution increases, a digital image can be regarded as a discretization of a continuous function defined on a two-dimensional rectangular domain. Mathematical image processing are mainly based on two mathematical tools, namely, advanced linear algebra and advanced calculus. In this chapter, we will review some basic mathematical concepts necessary for our discussions.

1 Basic linear algebra

We will first review some basic mathematical concepts related to linear algebra, which will be used in our discussions about mathematical imaging. The proofs for the theorems are omitted.

1.1 Determinant of a matrix

Definition 1.1. Let A be a $n \times n$ matrix. If $n = 1$, so that $A = (A_{11})$, we define $\det(A) = A_{11}$. For $n \geq 2$, we define $\det(A)$ recursively as

$$\det(A) = \sum_{j=1}^{n} (-1)^{1+j} A_{1j} \det(\tilde{A}_{1j}),$$

where $\tilde{A}_{ij}$ is the $(n-1) \times (n-1)$ matrix obtained from removing row i and column j of A (called the **minor** of the entry of A in row i, column j). The scalar $\det(A)$ is called the **determinant** of A and is also denoted by $|A|$. The scalar

$$(-1)^{i+j} \det(\tilde{A}_{ij})$$

is called the **cofactor** of the entry of A in row i, column j.

Theorem 1.2. *The determinant of a $n \times n$ matrix is a linear function of each row when the remaining rows are held fixed. That is, for $1 \leq r \leq n$, we have*

$$\det \begin{pmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_{r-1} \\ \mathbf{u} + k\mathbf{v} \\ \mathbf{a}_{r+1} \\ \vdots \\ \mathbf{a}_n \end{pmatrix} = \det \begin{pmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_{r-1} \\ \mathbf{u} \\ \mathbf{a}_{r+1} \\ \vdots \\ \mathbf{a}_n \end{pmatrix} + k \det \begin{pmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_{r-1} \\ \mathbf{v} \\ \mathbf{a}_{r+1} \\ \vdots \\ \mathbf{a}_n \end{pmatrix}$$

whenever k is a scalar and $\mathbf{u}, \mathbf{v}$ and each $\mathbf{a}_i$ are row vectors in $\mathbb{R}^n$.

Theorem 1.3. *The determinant of a square matrix can be evaluated by cofactor expansion along any row. That is, if A is a $n \times n$ matrix, then for any integer i ($1 \leq i \leq n$),*

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} A_{ij} \det(\tilde{A}_{ij}).$$

Corollary. *If a square matrix A has two identical rows, then $\det(A) = 0$.*

Theorem 1.4. *If A is a square matrix and B is a matrix obtained from A by interchanging any two rows of A, then $\det(B) = -\det(A)$.*

Theorem 1.5. *Let A be a square matrix, and let B be a matrix obtained by adding a multiple of one row of A to another row of A. Then* $\det(B) = \det(A)$.

Corollary. *If a $n \times n$ matrix has rank less than n, then* $\det(A) = 0$.

Theorem 1.6. *For any two $n \times n$ matrices A and B,* $\det(AB) = \det(A)\det(B)$.

Corollary. *A square matrix is invertible if and only if* $\det(A) \neq 0$. *Furthermore, if A is invertible, then* $\det(A^{-1}) = \dfrac{1}{\det(A)}$.

Theorem 1.7. *For any square matrix A,* $\det(A^T) = \det(A)$.

1.2 Eigenvalues and Eigenvectors

Definition 1.8. Let A be a $n \times n$ matrix. A nonzero vector $\mathbf{v} \in \mathbb{R}^n$ is called an **eigenvector** of A if there exists a scalar λ such that $A\mathbf{v} = \lambda\mathbf{v}$. The scalar λ is called the **eigenvalue** of A corresponding to the eigenvector $\mathbf{v}$.

Definition 1.9. Let A be a $n \times n$ matrix. The polynomial $f(t) = det(A - tI_n)$ is called the **characteristic polynomial** of A.

Definition 1.10. Let A be a $m \times n$ matrix. We define the **conjugate transpose** or **adjoint** of A to be the $n \times m$ matrix A^* such that $(A^*)_{ij} = \overline{A_{ji}}$ for all i, j.

Definition 1.11. Let A be a $n \times n$ matrix. A is said to be **normal** if $AA^* = A^*A$; in particular, A is said to be **unitary** if $AA^* = A^*A = I_n$.

Definition 1.12. Let A and B be $n \times n$ matrices. A is said to be **similar** to B if there exists an invertible matrix C such that $B = C^{-1}AC$. In particular, A is said to be **unitarily equivalent** to B if C is unitary.

1.3 Diagonalization and Jordan Canonical Form

Definition 1.13. Let A be a $n \times n$ matrix. A is said to be **diagonalizable** if A is similar to a diagonal matrix; in particular, A is said to be **unitarily diagonalizable** if A is unitarily equivalent to a diagonal matrix.

Theorem 1.14. *A square matrix is normal if and only if it is unitarily diagonalizable.*

Theorem 1.15. *All eigenvalues of a real symmetric matrix are real.*

Theorem 1.16. *For every $n \times n$ complex matrix A, there exists a $n \times n$ complex invertible matrix B such that $B^{-1}AB$ is of the form* $\begin{pmatrix} A_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & A_2 & \cdots & \mathbf{0} \\ \vdots & & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & A_k \end{pmatrix}$, *where each matrix block A_i is of the form* $\begin{pmatrix} \lambda_i & 1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda_i & 1 & \cdots & 0 & 0 \\ 0 & 0 & \lambda_i & \cdots & 0 & 0 \\ \vdots & & & \ddots & & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_i & 1 \\ 0 & 0 & 0 & \cdots & 0 & \lambda_i \end{pmatrix}$.

Definition 1.17. The product $B^{-1}AB$ in Theorem 1.16 is said to be a **Jordan canonical form** of A.

2 Vector calculus in Euclidean space $\mathbb{R}^n$

In this short section, we shall introduce some basic concepts about the vector calculus, which are widely used in engineering, physics and other areas requiring mathematics.
Consider the Euclidean space $\mathbb{R}^n$, with a rectangular coordinate system formed by the x_1-, x_2-, $\cdots$, and x_n-coordinate axis. The vector calculus is about some basic operations of a vector-valued function in $\mathbb{R}^n$.
For any given vector-valued function $\mathbf{v}(\mathbf{x})$ in $\mathbb{R}^n$ with n components, we shall write $\mathbf{v}(\mathbf{x})$ as

$$\mathbf{v}(\mathbf{x}) = (v_1(\mathbf{x}), v_2(\mathbf{x}), \cdots, v_n(\mathbf{x}))^T \quad \text{with} \quad \mathbf{x} = (x_1, x_2, \cdots, x_n)^T.$$

2.1 Gradient and divergence

<u>**Gradient vector**</u>. For each scalar function $u(\mathbf{x})$ in $\mathbb{R}^n$, we can define a vector $\mathbf{v} = \big(\frac{\partial u}{\partial x_1}, \frac{\partial u}{\partial x_2}, \cdots, \frac{\partial u}{\partial x_n}\big)^T$. This function is called the gradient vector of $u(\mathbf{x})$. We often write

$$\mathbf{v} = \text{grad } u = \nabla u\,.$$

Example 2.1. Find the gradient of u for $u(x_1, x_2, x_3) = x_1^2 + x_2^2 + \sin(\pi x_3)$.

<u>**Divergence**</u>. For any given vector-valued function $\mathbf{v} = (v_1, v_2, \cdots, v_n)^T$, we can define a scalar function $w(\mathbf{x}) = \frac{\partial v_1}{\partial x_1} + \frac{\partial v_2}{\partial x_2} + \cdots + \frac{\partial v_n}{\partial x_n}$. This function is called the divergence of function $\mathbf{v}$. And we often write

$$\text{div } \mathbf{v} = \frac{\partial v_1}{\partial x_1} + \frac{\partial v_2}{\partial x_2} + \cdots + \frac{\partial v_n}{\partial x_n} \equiv \nabla \cdot \mathbf{v}$$

It is easy to verify that

$$\text{div grad } u = \nabla \cdot \nabla u = \frac{\partial^2 u}{\partial x_1^2} + \frac{\partial^2 u}{\partial x_2^2} + \cdots + \frac{\partial^2 u}{\partial x_n^2},$$

and we often write

$$\Delta u = \frac{\partial^2 u}{\partial x_1^2} + \frac{\partial^2 u}{\partial x_2^2} + \cdots + \frac{\partial^2 u}{\partial x_n^2}\,.$$

Δ is called the **Laplacian operator**. The notation Δ is often extended to vector-valued functions: for a vector-valued function $\mathbf{v}(\mathbf{x})$ in $\mathbb{R}^n$, we write

$$\Delta \mathbf{v} = (\Delta v_1, \Delta v_2, \cdots, \Delta v_n)^T.$$

2.2 Vorticity and curl operations

Three dimensions
<u>**Curl vector (vorticity)**</u>. For a given vector-valued function $\mathbf{v} = (v_1, v_2, v_3)^T$ in $\mathbb{R}^3$, we can define the following special vector-valued function

$$\text{curl } \mathbf{v} \equiv \nabla \times \mathbf{v} = \begin{pmatrix} \frac{\partial v_3}{\partial x_2} - \frac{\partial v_2}{\partial x_3} \\ \frac{\partial v_1}{\partial x_3} - \frac{\partial v_3}{\partial x_1} \\ \frac{\partial v_2}{\partial x_1} - \frac{\partial v_1}{\partial x_2} \end{pmatrix}$$

This vector is called the **curl** vector of $\mathbf{v}$ or the **vorticity** of $\mathbf{v}$. It may help remember the curl operation if we compare it with the vector cross-product $\mathbf{a} \times \mathbf{b}$.
Two dimensions
In two dimensions, curl operation is also frequently used. But its operation is very different from three dimensions. For any given scalar function $v(x, y)$, we define

$$\mathbf{curl}\, v = (\frac{\partial v}{\partial y}, -\frac{\partial v}{\partial x})^T\,.$$

Sometimes, we write

$$\mathbf{curl}\, v = \nabla^{\perp} v\,,$$

as ∇ and $\nabla^\perp$ are orthogonal in the sense that

$$\nabla v \cdot \nabla^\perp v = 0\,.$$

It is interesting to notice that the Euclidean norm of $\mathbf{curl}\, v$ is the same as the gradient ∇v. But this is only valid in two dimensions. (Why?)
One can also define another curl operation in two dimensions. For any vector-valued function $\mathbf{v}(x, y)$, we define

$$\operatorname{curl} \mathbf{v} = \frac{\partial v_2}{\partial x} - \frac{\partial v_1}{\partial y}\,.$$

This maps a vector-valued function into a scalar function.

2.3 Some relations between gradient, divergence and curl operations

Example 2.2. Verify that the curl of the gradient of any function $u(x_1, x_2, x_3)$ is zero.

Solution. By the definition, we can check

$$\nabla \times \nabla u = \begin{pmatrix} \frac{\partial}{\partial x_1} \\ \frac{\partial}{\partial x_2} \\ \frac{\partial}{\partial x_3} \end{pmatrix} \times \begin{pmatrix} \frac{\partial u}{\partial x_1} \\ \frac{\partial u}{\partial x_2} \\ \frac{\partial u}{\partial x_3} \end{pmatrix} = \mathbf{0}\ .$$

The student may work out the detail. $\square$

- Find $\nabla \times \nabla u$ when

$$u(x_1, x_2, x_3) = e^{-x_1^2 + x_2^2 + \sin(\pi x_3)} \sin(\pi x_1 x_2 x_3).$$

- Think about the operation $\nabla\,(\nabla \times u)$.

Example 2.3. Verify that the divergence of the curl of any vector-valued function $\mathbf{v}(x_1, x_2, x_3)$ is zero.

Solution. The student may work out the detail for

$$\operatorname{div}\, \mathbf{curl}\ v = \nabla \cdot (\nabla \times v) = 0\ .$$

$\square$

- Think about the operation $\operatorname{curl}(\operatorname{div} u)$. It does not make sense in 3D. What about this operation in 2D ?

Example 2.4. For any vector-valued function $\mathbf{v}(x_1, x_2, x_3)$, verify the following relation

$$\nabla \times (\nabla \times \mathbf{v}) = \nabla(\nabla \cdot \mathbf{v}) - \Delta \mathbf{v}\,.$$

The following conclusions are widely used to simplify some mathematical models:

Theorem 2.5. *If a vector-valued function $\mathbf{v}(x_1, x_2, x_3)$ is divergence-free, that is, $\nabla \cdot \mathbf{v} = 0$, then there exists a vector-valued function $\mathbf{w}(x_1, x_2, x_3)$ such that*

$$\mathbf{v}(x_1, x_2, x_3) = \nabla \times \mathbf{w}(x_1, x_2, x_3)\,.$$

On the other hand, if a vector-valued function $\mathbf{v}(x_1, x_2, x_3)$ is vorticity-free, that is, $\nabla \times \mathbf{v} = 0$, then there exists a scalar field $\phi(x_1, x_2, x_3)$ such that

$$\mathbf{v}(x_1, x_2, x_3) = \nabla \phi(x_1, x_2, x_3)\,.$$

Finally, the students may try themselves to verify the following property:

Example 2.6. For any scalar function $u(x_1, x_2, \cdots, x_n)$ and any vector-valued function $\mathbf{v}(x_1, x_2, \cdots, x_n)$, we have

$$\nabla \cdot (u\, \mathbf{v}) = \nabla u \cdot \mathbf{v} + u(\nabla \cdot \mathbf{v})\,.$$

What is the operation $\nabla \times (u\,\mathbf{v})$? Work out a convenient formula for evaluating such operation.

The following integration by parts formulae are widely used in deriving variational formulations for various partial differential equations:

For two scalar functions u and v, we have

$$\int_\Omega u_{x_i}\, v\, dx = -\int_\Omega u\, v_{x_i}\, dx + \int_{\partial\Omega} u\, v\, n_i\, ds\,.$$

where $\Omega \subset R^k$, $\mathbf{n} = (n_1, n_2, \cdots, n_k)^T$ is the unit outward normal direction to the boundary $\partial\Omega$ of Ω. Hence for a vector-valued function $\mathbf{u}$ and a scalar function v, we have

$$\int_\Omega (\nabla \cdot \mathbf{u})\, v\, dx = -\int_\Omega (\mathbf{u} \cdot \nabla v)\, dx + \int_{\partial\Omega} (\mathbf{u} \cdot \mathbf{n})\, v\, ds\,.$$

For two vector-valued functions $\mathbf{u}$ and $\mathbf{v}$, we have

$$\int_\Omega (\nabla \times \mathbf{u}) \cdot \mathbf{v}\, dx = \int_\Omega \mathbf{u} \cdot (\nabla \times \mathbf{v})\, dx - \int_{\partial\Omega} (\mathbf{u} \times \mathbf{n}) \cdot \mathbf{v}\, ds\,.$$