MATH2111 Tutorial 12 & 13

T1A&T1B QUAN Xueyang
T1C&T2A SHEN Yinan
T2B&T2C ZHANG Fa

1 Inner Product, Length, and Orthogonality

1. **Definition (Inner Product)**. If $\mathbf{u}$ and $\mathbf{v}$ are vectors in $\mathbb{R}^n$, then the number $\mathbf{u}^T\mathbf{v}$ is called the inner product (or dot product) of $\mathbf{u}$ and $\mathbf{v}$, and often it is written as $\langle \mathbf{u}, \mathbf{v} \rangle$ (or $\mathbf{u} \cdot \mathbf{v}$).

2. **Theorem**. Let $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ be vectors in $\mathbb{R}^n$, and let c be a scalar. Then
   (a) $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$
   (b) $(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w}$
   (c) $(c\mathbf{u}) \cdot \mathbf{v} = c(\mathbf{u} \cdot \mathbf{v}) = \mathbf{u} \cdot (c\mathbf{v})$
   (d) $\mathbf{u} \cdot \mathbf{u} \geq 0$, and $\mathbf{u} \cdot \mathbf{u} = 0$ if and only if $\mathbf{u} = \mathbf{0}$

3. **Definition**. The norm (or length) of $\mathbf{v}$ is the nonnegative scalar $||\mathbf{v}||$ defined by

   $$||\mathbf{v}|| = \sqrt{\mathbf{v} \cdot \mathbf{v}} = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}$$

   therefore we have $||\mathbf{v}||^2 = \mathbf{v} \cdot \mathbf{v}$.

4. **Definition**. For $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$, the distance between $\mathbf{u}$ and $\mathbf{v}$, written as $\text{dist}(\mathbf{u}, \mathbf{v})$, is the length of the vector $\mathbf{u} - \mathbf{v}$. i.e.

   $$\text{dist}(\mathbf{u}, \mathbf{v}) = ||\mathbf{u} - \mathbf{v}||$$

5. **Definition**. Two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ are orthogonal (to each other) if $\mathbf{u} \cdot \mathbf{v} = 0$.

6. **Theorem (The Pythagorean Theorem)**. Two vectors $\boldsymbol{u}$ and $\mathbf{v}$ are orthogonal if and only if $||\mathbf{u} + \mathbf{v}||^2 = ||\mathbf{u}||^2 + ||\mathbf{v}||^2$.

7. **Definition**.
   (a) If a vector $\mathbf{z}$ is orthogonal to every vector in a subspace W of $\mathbb{R}^n$, then $\mathbf{z}$ is said to be orthogonal to W.
   (b) The set of all vectors $\mathbf{z}$ that are orthogonal to W is called the orthogonal complement of W and is denoted by $W^\perp$ and read as " W perpendicular" or simply "W perp".

8. **Theorem**.
   (a) A vector $\mathbf{x}$ is in $W^\perp$ if and only if $\mathbf{x}$ is orthogonal to every vector in a set that spans W.
   (b) $W^\perp$ is a subspace of $\mathbb{R}^n$.

9. **Theorem**. Let A be an $m \times n$ matrix. The orthogonal complement of the row space of A is the null space of A, and the orthogonal complement of the column space of A is the null space of A^T:
   (a) $(\text{Row } A)^\perp = \text{Nul } A$
   (b) $(\text{Col } A)^\perp = \text{Nul } A^\top$

2 Orthogonal Sets

1. **Definition**.
   (a) A set of vectors $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ in $\mathbb{R}^n$ is said to be an orthogonal set if each pair of distinct vectors from the set is orthogonal. i.e. $\mathbf{u}_i \cdot \mathbf{u}_j = 0$ whenever $i \neq j$
   (b) An orthogenal basis for a subspace W of $\mathbb{R}^n$ is a basis for W that is also an orthogonal set.

2. **Theorem**. If $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is an orthogonal set of nonzero vectors in $\mathbb{R}^n$, then S is linearly independent and hence is a basis for the subspace spanned by S.

3. **Theorem**. Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ be an orthogonal basis for a subspace W of $\mathbb{R}^n$. For each $\mathbf{y}$ in W,

$$y = \frac{y \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1 + \cdots + \frac{y \cdot \mathbf{u}_p}{\mathbf{u}_p \cdot \mathbf{u}_p}\mathbf{u}_p$$

4. **Definition**. A set $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is an orthonormal set if it is an orthogonal set of unit vectors. If W is the subspace spanned by such a set, then $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is an orthonormal basis for W.

5. **Theorem**. An $m \times n$ matrix U has orthonormal columns if and only if $U^T U = I$.

6. **Theorem**. Let U be an $m \times n$ matrix with orthonormal columns, and let $\mathbf{x}$ and $\mathbf{y}$ be in $\mathbb{R}^n$. Then
   (a) $\|U\mathbf{x}\| = \|\mathbf{x}\|$
   (b) $(U\mathbf{x}) \cdot (U\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$
   (c) $(U\mathbf{x}) \cdot (U\mathbf{y}) = 0$ if and only if $\mathbf{x} \cdot \mathbf{y} = 0$

3 Orthogonal Projections

1. **Theorem (Orthogonal Projection)**. The orthogonal projection of $\mathbf{y}$ onto $\mathbf{u}$ can be found by

$$\text{proj}_\text{u}\, \mathbf{y} = \frac{\mathbf{y} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}}\mathbf{u}$$

2. **Theorem (The Orthogonal Decomposition Theorem)**. Let W be a subspace of $\mathbb{R}^n$. Then each $\mathbf{y}$ in $\mathbb{R}^n$ can be written uniquely in the form

$$y = \hat{\mathbf{y}} + z$$

   where $\hat{\mathbf{y}}$ is in W and z is in $W^\perp$. In fact, $\hat{y}$ is called the orthogonal projection of $\mathbf{y}$ onto W, and if $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is any orthogonal basis of W, then

$$\hat{\mathbf{y}} = \text{proj}_W\, y = \frac{y \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1 + \cdots + \frac{y \cdot \mathbf{u}_p}{\mathbf{u}_p \cdot \mathbf{u}_p}\mathbf{u}_p$$

   and $z = y - \text{proj}_W\, y$.

3. **Theorem (The Best Approximation Theorem)**. Let W be a subspace of $\mathbb{R}^n$ and $\mathbf{y}$ be any vector in $\mathbb{R}^n$. Then $\text{proj}_W\,\mathbf{y}$ is the closest point in W to $\mathbf{y}$, in the sense that

$$\left\|\mathbf{y} - \text{proj}_W\,\mathbf{y}\right\| < \|\mathbf{y} - \mathbf{v}\|$$

for all $\mathbf{v}$ in W distinct from $\text{proj}_W\,\mathbf{y}$.

4. **Theorem**. If $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is an orthonormal basis for a subspace W of $\mathbb{R}^n$, then

$$\text{proj}_W\,\mathbf{y} = (\mathbf{y}\cdot\mathbf{u}_1)\,\mathbf{u}_1 + (\mathbf{y}\cdot\mathbf{u}_2)\,\mathbf{u}_2 + \cdots + (\mathbf{y}\cdot\mathbf{u}_p)\,\mathbf{u}_p$$

If $U = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_p \end{bmatrix}$, then

$$\text{proj}_W\,\mathbf{y} = UU^T\mathbf{y}$$

for all $\mathbf{y}$ in $\mathbb{R}^n$.

4 The Gram-Schmidt Process

1. **Theorem (The Gram-Schmidt Process)**. Given a basis $\{\mathbf{x}_1, \ldots, \mathbf{x}_p\}$ for a nonzero subspace W of $\mathbb{R}^n$, define

$$\begin{aligned}
\mathbf{v}_1 &= \mathbf{x}_1\\
\mathbf{v}_2 &= \mathbf{x}_2 - \frac{\mathbf{x}_2\cdot\mathbf{v}_1}{\mathbf{v}_1\cdot\mathbf{v}_1}\mathbf{v}_1\\
\mathbf{v}_3 &= \mathbf{x}_3 - \frac{\mathbf{x}_3\cdot\mathbf{v}_1}{\mathbf{v}_1\cdot\mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_3\cdot\mathbf{v}_2}{\mathbf{v}_2\cdot\mathbf{v}_2}\mathbf{v}_2\\
&\vdots\\
\mathbf{v}_p &= \mathbf{x}_p - \frac{\mathbf{x}_p\cdot\mathbf{v}_1}{\mathbf{v}_1\cdot\mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_p\cdot\mathbf{v}_2}{\mathbf{v}_2\cdot\mathbf{v}_2}\mathbf{v}_2 - \cdots - \frac{\mathbf{x}_p\cdot\mathbf{v}_{p-1}}{\mathbf{v}_{p-1}\cdot\mathbf{v}_{p-1}}\mathbf{v}_{p-1}
\end{aligned}$$

Then $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is an orthogonal basis for W. In addition

$$\text{Span}\,\{\mathbf{v}_1, \ldots, \mathbf{v}_k\} = \text{Span}\,\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$$

for $1 \le k \le p$.

2. **Theorem (The QR Factorization)**.
If A is an $m \times n$ matrix with linearly independent columns, then A can be factored as $A = QR$, where Q is an $m \times n$ matrix whose columns form an orthonormal basis for $\text{Col}\,A$ and R is an $n \times n$ upper triangular invertible matrix with positive entries on its diagonal.

5 Least-Squares Problems

1. **Definition**. If A is a $m \times n$ matrix and $\mathbf{b}$ is in $\mathbb{R}^m$, a least-squares solution of $A\mathbf{x} = \mathbf{b}$ is an $\hat{\mathbf{x}}$ in $\mathbb{R}^n$ such that

$$\|\mathbf{b} - A\hat{\mathbf{x}}\| \le \|\mathbf{b} - A\mathbf{x}\|$$

for all $\mathbf{x}$ in $\mathbb{R}^n$.

2. **Theorem**. The set of least-squares solutions of $A\mathbf{x} = \mathbf{b}$ coincides with the nonempty set of solutions of the normal equations $A^TA\mathbf{x} = A^T\mathbf{b}$.

3. **Theorem**. Let A be an $m \times n$ matrix. The following statements are logically equivalent:
   (a) The equation $A\mathbf{x} = \mathbf{b}$ has a unique least-squares solution for each $\mathbf{b}$ in $\mathbb{R}^m$.
   (b) The columns of A are linearly indpendent.
   (c) The matrix $A^T A$ is invertible.
   When these statements are true, the least-squares solution $\hat{\mathbf{x}}$ is given by

$$\hat{\mathbf{x}} = \left(A^T A\right)^{-1} A^T \mathbf{b}$$

4. **Theorem**. Given an $m \times n$ matrix A with linearly independent columns, let $A = QR$ be a QR factorization of A. Then, for each $\mathbf{b} \in \mathbb{R}^m$, the equation $A\mathbf{x} = \mathbf{b}$ has a unique least-squares solution, given by

$$\hat{\mathbf{x}} = R^{-1} Q^T \boldsymbol{b}$$

6 Applications to Linear Models

1. **Definition (Least-Squares Lines)**.
   Given experimental data $(x_1, y_1), \ldots, (x_n, y_n)$, the least-squares line is the line $y = \beta_0 + \beta_1 x$ that minimizes the sum of the squares of the residuals. And $\hat{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$ is the least-square solution of

$$X\beta = \mathbf{y}$$

   where

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \ \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

2. **Definition (Least-Squares Fitting Curves)**.
   Given experimental data $(x_1, y_1), \ldots, (x_n, y_n)$, the least-squares fitting curve is the curve $y = \beta_0 f_0(x) + \beta_1 f_1(x) + \cdots + \beta_k f_k(x)$ that minimizes the sum of the squares of the residuals, where $f_0, \ldots, f_k$ are known functions and $\hat{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}$ is the least-square solution of

$$X\beta = \mathbf{y}$$

   where

$$X = \begin{bmatrix} f_0(x_1) & f_1(x_1) & \cdots & f_k(x_1) \\ f_0(x_2) & f_1(x_2) & \cdots & f_k(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ f_0(x_n) & f_1(x_n) & \cdots & f_k(x_n) \end{bmatrix}, \ \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

7 Diagonalization of Symmetric Matrices

1. **Definition (Symmetric Matrix)**. A symmetric matrix is a square matrix A such that $A^T = A$.

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

3. **Definition**. An $n \times n$ matrix A is said to be orthogonally diagonalizable if there are an orthogonal matrix P (with $P^{-1} = P^T$) and a diagonal matrix D such that

$$A = PDP^T = PDP^{-1}$$

4. **Theorem (The Spectral Theorem for Symmetric Matrices)**.
An $n \times n$ symmetric matrix A has the following properties:
(a) A has n real eigenvalues, counting multiplicities.
(b) The dimension of the eigenspace for each eigenvalue equals the multiplicity of as a root of the characteristic equation.
(c) The eigenspaces are mutually orthogonal, in the sense that eigenvectors corresponding to different eigenvalues are orthogonal.
(d) A is orthogonally diagonalizable.

5. **Theorem**. An $n \times n$ matrix A is orthogonally diagonalizable if and only if A is a symmetric matrix.