Chapter 6　Inner Product Spaces

6.1 Inner Products

◎ We want to impose further structure on v.s. so that we can talk about norm of a vector and orthogonality between two vectors.

◇ norm $\equiv$ length

◇ orthogonal $\equiv$ perpendicular

◎ *Def.* An inner product (IP) on a real/complex vector space V is a function that associates a real/complex number $\langle \mathbf{u}, \mathbf{v} \rangle$ with each pair of vectors **u** and **v** in *V* in such a way that the following axioms are satisfied for all vectors **u**, **v**, and **w** in *V* and all scalars *k*:

◇1 $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle^*$ (i.e. complex conjugate of $\langle \mathbf{v}, \mathbf{u} \rangle$) (= $\langle \mathbf{v}, \mathbf{u} \rangle$ in the 'real' case)

◇2 $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$

◇3 $\langle k\mathbf{u}, \mathbf{v} \rangle = k \langle \mathbf{u}, \mathbf{v} \rangle$

◇4 $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$, and $\langle \mathbf{u}, \mathbf{u} \rangle = 0$ iff $\mathbf{u} = \mathbf{0}$

◎ ***Def.*** A real vector space with an inner product is called a real inner product space.

◇ Discussions on real IP space can be

generalized to complex case.

◇ A complex vector space with an IP is called a complex inner product space.

◎ *Def.* Norm of a vector **u**: $\|\mathbf{u}\| = \langle \mathbf{u}, \mathbf{u} \rangle^{1/2}$

◎ *Def.* Distance bet. **u** and **v**: $d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|$

◎ Properties of inner products:

◇ $\langle \mathbf{0}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{0} \rangle = 0$

◇ $\langle \mathbf{u}, \mathbf{v} + \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle$

◇ $\langle \mathbf{u}, k\mathbf{v} \rangle = k \langle \mathbf{u}, \mathbf{v} \rangle$

◇ $\langle \mathbf{u} - \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle - \langle \mathbf{v}, \mathbf{w} \rangle$

◇ $\langle \mathbf{u}, \mathbf{v} - \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle - \langle \mathbf{u}, \mathbf{w} \rangle$

6.2 Angle and Orthogonality in IP Spaces

◎ ***Thm.*** (Cauchy-Schwarz inequality) Let **u**, **v** be vectors in an IP space. Then,

$|\langle \mathbf{u}, \mathbf{v} \rangle| \leq \|\mathbf{u}\| \, \|\mathbf{v}\|$

◇ For memorization, recall that

$\langle \mathbf{u},\mathbf{v} \rangle = \|\mathbf{u}\| \, \|\mathbf{v}\| \cos(\theta)$ in Eucidean R^n.

◎ Properties of norm:

◇ $\| \mathbf{u} \| \geq 0$

◇ $\| \mathbf{u} \| = 0$ if and only if $\mathbf{u} = \mathbf{0}$

◇ $\| k\mathbf{u} \| = |k| \, \| \mathbf{u} \|$

◇ $\| \mathbf{u} + \mathbf{v} \| \leq \| \mathbf{u} \| + \| \mathbf{v} \|$ (Triangle inequality)

◎ Properties of distance:

◇ $d(\mathbf{u}, \mathbf{v}) \geq 0$

◇ $d(\mathbf{u}, \mathbf{v}) = 0$ if and only if $\mathbf{u} = \mathbf{v}$

◇ $d(\mathbf{u}, \mathbf{v}) = d(\mathbf{v}, \mathbf{u})$

◇ $d(\mathbf{u}, \mathbf{v}) \leq d(\mathbf{u}, \mathbf{w}) + d(\mathbf{w}, \mathbf{v})$ (Triangle inequality)

◎ *Def.* Angle between **u** and **v**:

$\cos^{-1}(\langle \mathbf{u}, \mathbf{v}\rangle/(\ \|\mathbf{u}\|\ \|\mathbf{v}\|))$ (range: $0 \sim \pi$)

◇ The definition was inspired by

$\langle \mathbf{u},\mathbf{v}\rangle = \|\mathbf{u}\|\ \|\mathbf{v}\| \cos(\theta)$ in Eucidean R^n.

◎ *Def.* **u** and **v** in an IP space are said to be orthogonal (denoted as $\mathbf{u} \perp \mathbf{v}$) if $\langle \mathbf{u}, \mathbf{v}\rangle = 0$.

◎ **0** is orthogonal to any vector.

◎ *Thm.* (Generalized Theorem of Pythagoras)

If $\mathbf{u} \perp \mathbf{v}$, then $\|\ \mathbf{u} + \mathbf{v}\ \|^2 = \|\ \mathbf{u}\ \|^2 + \|\ \mathbf{v}\ \|^2$.

◎ Consider W : a subspace of an IP space V.

◇ *Def.* A vector **u** in V is said to be orthogonal to *W* if it is orthogonal to every vector in *W*.

◇ ***Def.*** The set of all vectors in *V* that are orthogonal to *W* is called the orthogonal complement of *W* (denoted as $W^{\perp}$).

◎ Properties of orthogonal complements:

◇ $W^{\perp}$ is a subspace of *V*.

◇ $W \cap W^{\perp} = \boldsymbol{0}$.

◇ $(W^{\perp})^{\perp} = W$.

◎ Let A be an $m\times n$ matrix. We have the following geometric links between subspaces:

◇ nullspace(A) and row-space(A) are orthogonal complements in R^n ($=R^{1\times n}$) wrt the Euclidean IP.

◇ nullspace(A^T) and column-space(A) are orthogonal complements in $R^{m\times 1}$ wrt the Euclidean IP.

◎ In any IP space V, the zero space $\{\mathbf{0}\}$ and V are orthogonal complements.

◎ Two more equivalent statements to an n×n matrix **A** being invertible (derived from the fact that $\mathbf{Ax}=\mathbf{0}$ has only trivial solution):

◇ The orthogonal complement of nullspace(**A**) is R^n.

◇ The orthogonal complement of row-space(**A**) is {**0**}.

6.3 Orthonormal Bases

◎ *Def.* $S=\{\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_n\}$ is an orthogonal set iff $\mathbf{v}_i \perp \mathbf{v}_j$ for all $i \neq j$. If, in addition, $\|\mathbf{v}_j\|=1$ for all i, then, S is an orthonormal (o.n.) set.

◎ *Thm.* If $S=\{\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_n\}$ is an orthogonal set, then it is l.i.

<u>*Prf.*</u> Consider $k_1\mathbf{v}_1+k_2\mathbf{v}_2+\ldots+k_n\mathbf{v}_n=\mathbf{0}$ – (#)

$<$LHS(#),$\mathbf{v}_1> = <$ RHS(#),$\mathbf{v}_1> = 0$ – (@)

In the meantime, $<$LHS(#),$\mathbf{v}_1>$

$= k_1< \mathbf{v}_1,\mathbf{v}_1> + k_2< \mathbf{v}_2,\mathbf{v}_1> + \ldots + k_n< \mathbf{v}_n,\mathbf{v}_1>$

$= k_1< \mathbf{v}_1,\mathbf{v}_1> + k_2\cdot 0 + \ldots + k_n\cdot 0$ – ($)

(and $< \mathbf{v}_1,\mathbf{v}_1> \neq 0$, unless $\mathbf{v}_1=\mathbf{0}$)

(@) and ($) $\rightarrow$ $k_1=0$

Similarly, we can show $k_2=0$, $k_3=0$, …

◎ <u>*Thm.*</u> If $B=\{\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_n\}$ is an o.n. basis for an IP space V, and $\mathbf{u}$ is any vector in V, then

$$\mathbf{u} = \langle \mathbf{u}, \mathbf{v}_1\rangle\, \mathbf{v}_1 + \langle \mathbf{u}, \mathbf{v}_2\rangle\, \mathbf{v}_2 + \cdots + \langle \mathbf{u}, \mathbf{v}_n\rangle\, \mathbf{v}_n$$

◇ The scalars $\langle \mathbf{u}, \mathbf{v}_1 \rangle, \langle \mathbf{u}, \mathbf{v}_2 \rangle, \ldots, \langle \mathbf{u}, \mathbf{v}_n \rangle$ are the coordinates of the vector **u** wrt B.

◇ $[\mathbf{u}]_B = [\langle \mathbf{u}, \mathbf{v}_1 \rangle, \langle \mathbf{u}, \mathbf{v}_2 \rangle, \ldots, \langle \mathbf{u}, \mathbf{v}_n \rangle]^T$ denotes the coordinate vector of **u** wrt B.

◇ Equivalently,

$(\mathbf{u})_B = (\langle \mathbf{u}, \mathbf{v}_1 \rangle, \langle \mathbf{u}, \mathbf{v}_2 \rangle, \ldots, \langle \mathbf{u}, \mathbf{v}_n \rangle)$

◎ *Thm.* If B is an o.n. basis for an *n*-dim IP space, and if $[\mathbf{u}]_B = [u_1, u_2, \ldots, u_n]^T$ and $[\mathbf{v}]_B = [v_1, v_2, \ldots, v_n]^T$, then

$$\|\mathbf{u}\| = \sqrt{u_1^2 + u_2^2 + \cdots + u_n^2}$$

$$d(\mathbf{u}, \mathbf{v}) = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2 + \cdots + (u_n - v_n)^2}$$

$$\langle \mathbf{u}, \mathbf{v} \rangle = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n$$

◇ With o.n. bases, the computation of general norms and IP's can be reduced to the computation of Euclidean norms and IP's of the coordinate vectors.

◎ *Thm.* (Projection Theorem): If W is a finite-dim subspace of an IP space V, then every vector $\mathbf{u}$ in V can be expressed in exactly one way as $\mathbf{u} = \mathbf{w}_1 + \mathbf{w}_2$, where $\mathbf{w}_1$ is in W and $\mathbf{w}_2$ is in $W^{\perp}$.

◇ The vector $\mathbf{w}_1$ is called the (orthogonal) projection of $\mathbf{u}$ on(to) W, and is denoted as $\text{proj}_W\,\mathbf{u}$.

◇ The vector $\mathbf{w}_2$ is called the component of $\mathbf{u}$ orthogonal to W, and is denoted as $\text{proj}_{W\perp}\,\mathbf{u}$.

◇ If $\{\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_n\}$ is an o.n. basis of W, then $\text{proj}_W\,\mathbf{u}=\langle\mathbf{u},\mathbf{v}_1\rangle\mathbf{v}_1+\langle\mathbf{u},\mathbf{v}_2\rangle\,\mathbf{v}_2+\ldots+\langle\mathbf{u},\mathbf{v}_n\rangle\mathbf{v}_n$.

◇ If $\{\mathbf{w}_1,\mathbf{w}_2,\ldots,\mathbf{w}_n\}$ is an orthogonal basis of W, then
$\mathrm{proj}_W\,\mathbf{u}=\langle\mathbf{u},\mathbf{v}_1\rangle\mathbf{v}_1+\langle\mathbf{u},\mathbf{v}_2\rangle\,\mathbf{v}_2+\ldots+\langle\mathbf{u},\mathbf{v}_n\rangle\mathbf{v}_n$, where $\mathbf{v}_i= \mathbf{w}_i/\|\mathbf{w}_i\|$ (i.e. normalized vector (aka. unit vector) along the $\mathbf{w}_i$ direction).

◎ *Thm.* Every nonzero finite-dim IP space V has an o.n. basis.

◇ Given any basis (or more loosely, any spanning set) of V, an o.n. (or more loosely, orthogonal) basis (not unique) can be constructed by the Gram-Schmidt process.

- ◇ Key idea: We sequentially modify each basis vector so that it is orthogonal to all of the previously established basis vectors.
- ◇ More exactly speaking, we subtract all projections onto previous basis vectors from the current basis vector.

◎ QR decomposition: If $\boldsymbol{A}$ is an $m \times n$ matrix with l.i. columns, then $\boldsymbol{A}$ can be factored as $\boldsymbol{A} = \boldsymbol{QR}$, where $\boldsymbol{Q}$ is an $m \times n$ matrix with o.n. column vectors, and $\boldsymbol{R}$ is an $n \times n$ invertible upper triangular matrix.

6.4 Best Approximation and Least Squares

◎ Projection in 3-D Euclidean space onto a plane:

- ◇ If P is a point in 3-space and W is a plane through the origin, then the point Q in W closest to P is obtained by dropping a perpendicular from P to W.
- ◇ Therefore, if we let $\mathbf{u}$ = OP, the distance between P and W is $\| \mathbf{u} - \text{proj}_W\mathbf{u} \|$.

- In other words, among all vectors w in W, the vector $\mathbf{w} = \text{proj}_W\mathbf{u}$ minimizes the distance $\| \mathbf{v} - \mathbf{w} \|$.
- In the sense of minimum distance of error (vector), we can regard $\text{proj}_W\mathbf{u}$ as the "best approximation" to $\mathbf{u}$ by the vectors in W.

◎ Best approximation theorem for general IP space: If W is a subspace of an IP space V, and if $\mathbf{u}$ is a vector in V, then $\text{proj}_W\mathbf{u}$ is the best approximation to $\mathbf{u}$ from W in the sense that $\| \mathbf{u} - \text{proj}_W \mathbf{u} \| < \| \mathbf{u} - \mathbf{w} \|$, where $\mathbf{w}$ is any other vector in W.

◎ Least squares problem:

- ◇ Given $\mathbf{Ax} = \mathbf{b}$ of m equations in n unknowns, find a vector $\mathbf{x}$, if possible, so that $\| \mathbf{Ax} - \mathbf{b} \|$ (wrt the Euclidean IP) is minimized.
- ◇ Such a vector is called a least squares solution (denoted as $\mathbf{x}_{LS}$).
- ◇ Denote col-space($\mathbf{A}$) as W. The least squares problem is (geometrically)

equivalent to finding a vector **x** in $R^{n\times 1}$ such that **Ax** is the closest vector in W to **b**. In other words, $\mathbf{Ax}_{LS} = \text{proj}_W\mathbf{b}$.

◇ Orthogonality requirement:

error (i.e. $\mathbf{b} - \mathbf{Ax}_{LS}$) $\perp$ W

◇ Equivalently, we require

$\mathbf{b} - \mathbf{Ax}_{LS} \perp$ every column of **A**

→ We solve for **x** in $\mathbf{A}^T(\mathbf{b} - \mathbf{Ax}) = \mathbf{0}$.

→ $\mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{b}$ (normal equations)

→ $\mathbf{x}_{LS} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$, if $\mathbf{A}^T\mathbf{A}$ is invertible.

◇ *Thm.* $\mathbf{A}^T\mathbf{A}$ is invertible iff **A** has l.i.

columns.

◇ $\text{proj}_W \mathbf{b} = \mathbf{A}\mathbf{x}_{LS} = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$

◎ Orthogonal projection:

◇ *Def.* If W is a subspace of $R^{m\times 1}$, then the transformation P: $R^{m\times 1} \to W$ that maps each vector $\mathbf{x}$ in $R^{m\times 1}$ into $\text{proj}_W \mathbf{x}$ is called orthogonal projection of $R^{m\times 1}$ on(to) W.

◇ Matrix $\mathbf{A}$: m×n. Then, $\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$ performs the orthogonal projection of $R^{m\times 1}$ onto col-space($\mathbf{A}$).

◎ More equivalent statement for invertibility of an n$\times$n matrix **A**:

- ◇ The orthogonal complement of the nullspace of **A** is R^n.
- ◇ The orthogonal complement of the row space of **A** is $\{\mathbf{0}\}$.
- ◇ $\mathbf{A}^T\mathbf{A}$ is invertible.