Chapter 11: Inner Product

11.1 Basic Properties of Dot Product

Definition 11.1.1: For $\alpha, \beta \in \mathbb{R}^m$, we define

$$\langle \alpha, \beta \rangle = \sum_{i=1}^{m} [\alpha]_i [\beta]_i = [\alpha]_1 [\beta]_1 + \cdots + [\alpha]_m [\beta]_m.$$

It is called an *inner product* (*dot product*) of $\mathbb{R}^m$ and $\mathbb{R}^m$ is called an *inner product space*. If we regard α, β as $m \times 1$ matrix, then

$$\langle \alpha, \beta \rangle = \alpha^t \beta.$$

Example 11.1.1:

$$\left\langle \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 3 \\ 4 \end{pmatrix} \right\rangle = 1 \times 3 + 2 \times 4 = 11,$$

$$\left\langle \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} \right\rangle = 1 \times 4 + 2 \times 5 + 3 \times 6 = 32.$$

Proposition 11.1.2: *For any $\alpha, \beta, \gamma \in \mathbb{R}^m$, $a \in \mathbb{R}$, we have*

(1) $\langle \alpha, \beta \rangle = \langle \beta, \alpha \rangle$. Symmetric

(2) $\langle a\alpha + \beta, \gamma \rangle = a \langle \alpha, \gamma \rangle + \langle \beta, \gamma \rangle$. Bilinear

(3) $\langle \alpha, \alpha \rangle \geq 0$. *Moreover,* $\langle \alpha, \alpha \rangle = 0$ *if and only if* $\alpha = \mathbf{0}$. Positive Definite

Proof: (1) and (2) are properties of matrix algebra.

(3). Let $\alpha = (x_1, \ldots, x_m)^t$. Then $\langle \alpha, \alpha \rangle = \sum\limits_{i=1}^{m} x_i^2 \geq 0$. So $\langle \alpha, \alpha \rangle = 0$ if and only if $x_i = 0$ for all i if and only if $\alpha = \mathbf{0}$.

Remark 11.1.3: Dot product is a very special inner product. In general, an inner product is a function, $\langle \cdot, \cdot \rangle$, of two vector-variables satisfying properties (1), (2) and (3) in Proposition 11.1.2. All results in the following sections, you may see that they hold since the proof do not involve the actually formula of an inner product.

Following is an example of inner product even though it may be out of syllabus.

Example 11.1.2: Consider the vector space $C^0[a, b]$ (or $P_n(\mathbb{R})$), the set of all continuous real-valued functions. Let $f, g \in C^0[a, b]$. Define

$$\langle f, g \rangle = \int_a^b f(t)g(t)dt.$$

It is easy to check that this function satisfies the properties listed in Proposition 11.1.2. ■

Proposition 11.1.4: *Let $a, b \in \mathbb{R}$, $\alpha, \beta, \gamma \in \mathbb{R}^m$. Then*

(a) $\langle \mathbf{0}, \alpha \rangle = \langle \alpha, \mathbf{0} \rangle = 0$.

(b) $\langle a\alpha + b\beta, \gamma \rangle = a \langle \alpha, \gamma \rangle + b \langle \beta, \gamma \rangle$.

(c) $\langle \gamma, a\alpha + b\beta \rangle = a \langle \gamma, \alpha \rangle + b \langle \gamma, \beta \rangle$.

(d) *If* $\langle \alpha, \theta \rangle = 0$ *for all* $\theta \in \mathbb{R}^m$*, then* $\alpha = \mathbf{0}$.

(e) *If* $\langle \alpha, \theta \rangle = \langle \beta, \theta \rangle$ *for all* $\theta \in \mathbb{R}^m$*, then* $\alpha = \beta$.

Proof:

Definition 11.1.5: The *norm* (or *length*) of $\alpha \in \mathbb{R}^m$ is defined by $\|\alpha\| = \sqrt{\langle \alpha, \alpha \rangle}$. Since $\langle \alpha, \alpha \rangle \geq 0$, $\sqrt{\langle \alpha, \alpha \rangle}$ is meaningful.

Example 11.1.3: Let $V = \mathbb{R}^3$ with the dot product. Let

$$\alpha = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \text{ and } \beta = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}.$$

Then

$$\|\alpha\| = \sqrt{\langle \alpha, \alpha \rangle} = \sqrt{1^2 + 2^2 + 3^2} = \sqrt{14}.$$
$$\|\beta\| = \sqrt{\langle \beta, \beta \rangle} = \sqrt{1^2 + 0^2 + 1^2} = \sqrt{2}.$$

■

Proposition 11.1.6: *Let* $a \in \mathbb{R}$ *and* $\alpha \in \mathbb{R}^m$.

1. $\|a\alpha\| = |a|\|\alpha\|$.
2. *Suppose* $\alpha \neq \mathbf{0}$*. Let* $a = \frac{1}{\|\alpha\|}$*, then* $\|a\alpha\| = 1$.

Proof: Consider $\|a\alpha\|^2 = \langle a\alpha, a\alpha \rangle = a^2 \langle \alpha, \alpha \rangle = a^2\|\alpha\|^2$. Since $\|\alpha\| \geq 0$, $\|a\alpha\| = |a|\|\alpha\|$.
Suppose $\alpha \neq \mathbf{0}$ and let $a = \frac{1}{\|\alpha\|}$. Now $\|a\alpha\| = \left|\frac{1}{\|\alpha\|}\right| \|\alpha\| = \frac{\|\alpha\|}{\|\alpha\|} = 1$. □

Definition 11.1.7: A vector $\alpha \in \mathbb{R}^m$ is said to be a *unit vector* if $\|\alpha\| = 1$.
A non-zero vector α can be *normalized* to a unit vector $\frac{1}{\|\alpha\|}\alpha$.

Example 11.1.4: In Example 11.1.3, α and β can be normalized to

$$\frac{1}{\|\alpha\|}\alpha = \frac{1}{\sqrt{14}}\alpha = \begin{pmatrix} \frac{1}{\sqrt{14}} \\ \frac{2}{\sqrt{14}} \\ \frac{3}{\sqrt{14}} \end{pmatrix},$$

$$\frac{1}{\|\beta\|}\beta = \frac{1}{\sqrt{2}}\beta = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{pmatrix}$$

respectively. ■

11.2 Orthogonal Sets

In this section, V denotes a subspace of $\mathbb{R}^m$ with inner product.

Definition 11.2.1: Suppose $\alpha, \beta \in V$. α and β are said to be *orthogonal* or *perpendicular* if $\langle \alpha, \beta \rangle = 0$. In this case, it is denoted by $\alpha \perp \beta$.

Example 11.2.1:

1. Let $V = \mathbb{R}^3$. Then

$$\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \perp \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}$$

as

$$\left\langle \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix} \right\rangle = 1 \times (-1) + 2 \times (-1) + 3 \times 1 = 0.$$

2. Let $V = \mathbb{R}^m$, then $\boldsymbol{e}_i \perp \boldsymbol{e}_j$ if $i \neq j$.

■

Definition 11.2.2: A subset $S = \{\alpha_1, \ldots, \alpha_k\}$ of V is said to be *orthogonal* if

1. $\mathbf{0} \notin S$, i.e., $\alpha_i \neq \mathbf{0}$ for $i = 1, \ldots, k$.
2. $\alpha_i \perp \alpha_j$ for $i \neq j$, i.e., $\langle \alpha_i, \alpha_j \rangle = 0$ for $i \neq j$.

Example 11.2.2:

1. $S = \left\{ \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix} \right\}$ is orthogonal subset of $\mathbb{R}^3$.
2. $S = \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix} \right\}$ is orthogonal subset of $\mathbb{R}^3$. Check!
3. $S = \{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_k\}$ in $\mathbb{R}^m$ with $k \leq m$ is orthogonal.

■

Proposition 11.2.3: *Let $S = \{\alpha_1, \ldots, \alpha_k\}$ be an orthogonal subset of V. Let*

$$\alpha = a_1\alpha_1 + \cdots + a_k\alpha_k,$$
$$\beta = b_1\alpha_1 + \cdots + b_k\alpha_k,$$

for some $a_i, b_i \in \mathbb{R}$, $i = 1, \ldots, k$. Then

$$\langle \alpha, \beta \rangle = a_1b_1\|\alpha_1\|^2 + \cdots + a_kb_k\|\alpha_k\|^2 = \sum_{i=1}^{k} a_ib_i\|\alpha_i\|^2. \tag{11.1}$$

Proof:

Theorem 11.2.4: *Let $S = \{\alpha_1, \ldots, \alpha_k\}$ be an orthogonal subset of V. Then S is linearly independent.*

Proof: Suppose we have the relation of linear dependence:

$$a_1\alpha_1 + \cdots + a_k\alpha_k = \mathbf{0}.$$

Theorem 11.2.5: *Let $S = \{\alpha_1, \ldots, \alpha_k\}$ be an orthogonal subset of V. Suppose $\alpha \in \langle S \rangle$. So*

$$\alpha = a_1\alpha_1 + \cdots + a_k\alpha_k,$$

for some $a_i \in \mathbb{R}$, $i = 1, \ldots, k$. Then

$$a_i = \frac{\langle \alpha, \alpha_i \rangle}{\|\alpha_i\|^2},$$

i.e.,

$$\alpha = \frac{\langle \alpha, \alpha_1 \rangle}{\|\alpha_1\|^2}\alpha_1 + \cdots + \frac{\langle \alpha, \alpha_k \rangle}{\|\alpha_k\|^2}\alpha_k = \sum_{i=1}^{k} \frac{\langle \alpha, \alpha_i \rangle}{\|\alpha_i\|^2}\alpha_i.$$

Proof:

Remark 11.2.6: The advantage of using the above method is that we do not need to solve linear equations to find the linear combination.
In order to use the theorem, we need to ensure that $\alpha \in \langle S \rangle$.

Example 11.2.3: Let S be the orthogonal subset in Example 11.2.2, item 2. Given

$$\alpha = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \in \langle S \rangle .$$

Write α as a linear combination of $\alpha_1, \alpha_2, \alpha_3$.
Answer:

Definition 11.2.7: Let V be a subspace of $\mathbb{R}^m$. A subset S of V is said to be an *orthogonal basis* for V if S is a basis for V and is orthogonal.

If S is an orthogonal subset of V, then by Theorem 11.2.4, it is automatically linearly independent. So in order to check if S is an orthogonal basis, we only need to check if $\langle S \rangle = V$. So we have

Theorem 11.2.8: *Let V be a subspace of $\mathbb{R}^m$. Suppose S is an orthogonal subset of V. Then S is an orthogonal basis if and only if $\langle S \rangle = V$.*

Corollary 11.2.9: *Suppose S is an orthogonal subset of $\mathbb{R}^m$. Then S is a basis for $\langle S \rangle$.*

Corollary 11.2.10: *Let V be a subspace of $\mathbb{R}^m$. Suppose $\mathscr{B} = \{\alpha_1, \ldots, \alpha_n\}$ is an orthogonal basis for V. Then for any $\alpha \in V$,*

$$\alpha = \frac{\langle \alpha, \alpha_1 \rangle}{\|\alpha_1\|^2}\alpha_1 + \cdots + \frac{\langle \alpha, \alpha_n \rangle}{\|\alpha_n\|^2}\alpha_n.$$

Proof: Follows from Theorem 11.2.5. □

Example 11.2.4:

1. $S = \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix} \right\}$ is orthogonal basis of $\mathbb{R}^3$, since $\dim(\langle S \rangle) = 3$ and $\langle S \rangle \subseteq \mathbb{R}^3$.

2. $S = \{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_m\}$ in $\mathbb{R}^m$ is an orthogonal basis. It is called the *standard basis* for $\mathbb{R}^m$. ■

Definition 11.2.11: A subset $S = \{\alpha_1, \ldots, \alpha_k\}$ of $\mathbb{R}^m$ is said to be *orthonormal* if it is orthogonal and every vector in S is a unit vector, i.e.,

$$\langle \alpha_i, \alpha_j \rangle = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

Let V be a subspace of $\mathbb{R}^m$. The subset $\mathscr{B}$ is said to be an *orthonormal basis* for V if it is orthonormal and is a basis for V.

Since an orthonormal set S is orthogonal, the above theorems regarding orthogonal sets are also true for orthonormal sets. In particular

Theorem 11.2.12: *Let $S = \{\alpha_1, \ldots, \alpha_k\}$ be an orthonormal subsets of $\mathbb{R}^m$ and $\alpha \in \langle S \rangle$. Then*

$$\alpha = \langle \alpha, \alpha_1 \rangle \alpha_1 + \cdots + \langle \alpha, \alpha_k \rangle \alpha_k.$$

If $S = \{\alpha_1, \ldots, \alpha_k\}$ is an orthogonal subset of $\mathbb{R}^m$, then $\left\{ \frac{1}{\|\alpha_1\|}\alpha_1, \ldots, \frac{1}{\|\alpha_k\|}\alpha_k \right\}$ is an orthonormal subset. The process is called *normalization*.

Example 11.2.5:

1. $S = \{(1, 2)^t, (-2, 1)^t\}$ is an orthogonal basis for $\mathbb{R}^2$. Normalizing it, we obtain an orthonormal basis.

$$\mathscr{B} = \left\{ \left(\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}} \right)^t, \left(\frac{-2}{\sqrt{5}}, \frac{1}{\sqrt{5}} \right)^t \right\}.$$

2. $S = \{(1, 1, 1)^t, (1, -1, 0)^t, (1, 1, -2)^t\}$ is an orthogonal basis for $\mathbb{R}^3$. Normalizing it, we obtain an orthonormal basis

$$\mathscr{B} = \left\{ \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}} \right)^t, \left(\frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}, 0 \right)^t, \left(\frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}, \frac{-2}{\sqrt{6}} \right)^t \right\}.$$

■

11.3 Gram-Schmidt Orthogonalization Process

Let $S = \{\alpha_1, \ldots, \alpha_k\}$ be an orthogonal subset of $\mathbb{R}^m$. Suppose $\beta \in \langle S \rangle$. From Theorem 11.2.5 we have

$$\beta = \frac{\langle \beta, \alpha_1 \rangle}{\|\alpha_1\|^2}\alpha_1 + \cdots + \frac{\langle \beta, \alpha_k \rangle}{\|\alpha_k\|^2}\alpha_k.$$

But what if β is not in $\langle S \rangle$? Let us compare the difference. We have the following theorem.

Theorem 11.3.1: *Let $S = \{\alpha_1, \ldots, \alpha_k\}$ be an orthogonal subset of $\mathbb{R}^m$ and let $\beta \in \mathbb{R}^m$. Then*

$$\alpha = \beta - \frac{\langle \beta, \alpha_1 \rangle}{\|\alpha_1\|^2}\alpha_1 - \cdots - \frac{\langle \beta, \alpha_k \rangle}{\|\alpha_k\|^2}\alpha_k$$

is orthogonal to α_i for $i = 1, \ldots, k$.

Proof:

Theorem 11.3.2: *Let $S = \{\beta_1, \beta_2, \ldots, \beta_k\}$ be a linearly independent subset of V. Let $\alpha_1 = \beta_1$ and*

$$\alpha_\ell = \beta_\ell - \frac{\langle \beta_\ell, \alpha_1 \rangle}{\|\alpha_1\|^2}\alpha_1 - \cdots - \frac{\langle \beta_\ell, \alpha_{\ell-1} \rangle}{\|\alpha_{\ell-1}\|^2}\alpha_{\ell-1}, \quad \text{for } 2 \le \ell \le k.$$

Then $T = \{\alpha_1, \ldots, \alpha_k\}$ is an orthogonal set.
Also $\langle \beta_1, \ldots, \beta_\ell \rangle = \langle \alpha_1, \ldots, \alpha_\ell \rangle$ for $\ell = 1, \ldots, k$. In particular $\langle S \rangle = \langle T \rangle$.
The process of obtaining T by the above procedure is called the Gram-Schmidt Orthogonalization Process.

Proof: Firstly, we have $\langle \beta_1 \rangle = \langle \alpha_1 \rangle$. We are going to add one vector at a time.
Suppose $\langle \alpha_1, \ldots, \alpha_{\ell-1} \rangle = \langle \beta_1, \ldots, \beta_{\ell-1} \rangle$ and the set $\{\alpha_1, \ldots, \alpha_{\ell-1}\}$ is orthogonal.

Thus $\langle \alpha_1, \ldots, \alpha_\ell \rangle = \langle \beta_1, \ldots, \beta_{\ell-1}, \alpha_\ell \rangle = \langle \beta_1, \ldots, \beta_{\ell-1}, \beta_\ell \rangle$, by Lemma 7.3.4.

Finally by Theorem 11.3.1, α_ℓ is orthogonal to α_i for $i = 1, \ldots, \ell - 1$. Thus $\{\alpha_1, \ldots, \alpha_\ell\}$ is orthogonal.

We repeat the process by increasing ℓ till $\ell = k$. □

Corollary 11.3.3: *Suppose V is a subspace of $\mathbb{R}^m$. There exists an orthogonal (orthonormal) basis for V.*

Proof: By Corollary 8.2.10, V has a basis S. Applying Gram-Schmidt orthogonalization process to S, we obtain an orthogonal set $\mathscr{B}$. By Theorem 11.3.2, $V = \langle S \rangle = \langle \mathscr{B} \rangle$. By Corollary 11.2.9, $\mathscr{B}$ is a basis of V.

Normalizing $\mathscr{B}$, we can obtain an orthonormal basis. □

Example 11.3.1: Let $V = \mathbb{R}^4$ with the standard inner product. Let

$$\beta_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \beta_2 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \beta_3 = \begin{pmatrix} 0 \\ 1 \\ 2 \\ 1 \end{pmatrix}.$$

Then $S = \{\beta_1, \beta_2, \beta_3\}$ is linearly independent. Applying Gram-Schmidt Orthogonalization Process find an orthonormal subset T such that $\langle S \rangle = \langle T \rangle$.

Answer: Take $\alpha_1 = \beta_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}$.

Example 11.3.2: Let $A = \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}$ be a 1×4 matrix. Let $V = \mathcal{N}(A)$. Find an orthonormal basis for V.

Answer: We can obtain that

$$S = \left\{\beta_1 = (-1, 1, 0, 0)^t, \beta_2 = (-1, 0, 1, 0)^t, \beta_3 = (-1, 0, 0, 1)^t\right\}$$

is a basis of $\mathcal{N}(A)$.

We apply Gram-Schmidt Orthogonalization Process to the set.

11.4 Some Useful Inequalities (Optional)

Theorem 11.4.1 (Cauchy-Schwarz's inequality)**:** *Let V be an inner product space. Then for $\alpha, \beta \in V$, $|\langle \alpha, \beta \rangle| \le \|\alpha\| \|\beta\|$.*

Proof:

Remark 11.4.2: *The equality holds if and only if β is a multiple of α or $\alpha = \mathbf{0}$.*
For, in the proof of Theorem 11.4.1, we see that the equality holds only if $\alpha = \mathbf{0}$ or $\frac{\langle \alpha, \beta \rangle}{\|\alpha\|^2}\alpha - \beta = \mathbf{0}$.
Conversely, if $\alpha = \mathbf{0}$ or $\beta = c\alpha$ for some scalar c, then the equality holds.

Theorem 11.4.3 (Triangle inequality)**:** *Let V be an inner product space. Then for $\alpha, \beta \in V$, $\|\alpha + \beta\| \le \|\alpha\| + \|\beta\|$. The equality holds if and only if $\alpha = \mathbf{0}$ or $\beta = c\alpha$ for some non-negative real number c.*

Proof:

Definition 11.4.4: Let $S = \{\xi_1, \xi_2, \ldots, \xi_m\}$ be an orthonormal set in an inner product space V and let $\alpha \in V$. The scalars $a_i = \langle \xi_i, \alpha\rangle$, $1 \le i \le m$, are called the *Fourier coefficients of α with respect to S.*

Theorem 11.4.5: *Let V be an inner product space. Suppose $\{\xi_1, \ldots, \xi_k\}$ is an orthonormal set in V. For $\alpha \in V$, we have $\min\limits_{\substack{x_i\in\mathbb{R}\\ 1\le i\le k}} \left\|\alpha - \sum\limits_{i=1}^{k} x_i\xi_i\right\| = \left\|\alpha - \sum\limits_{i=1}^{k} a_i\xi_i\right\|$, where $a_i = \langle \xi_i, \alpha\rangle$ are the Fourier coefficients of α.*

Moreover, $\left\|\alpha - \sum\limits_{i=1}^{k} x_i\xi_i\right\| = \left\|\alpha - \sum\limits_{i=1}^{k} a_i\xi_i\right\|$ if and only if $x_i = a_i$ for all i. Also $\sum\limits_{i=1}^{k} |a_i|^2 \le \|\alpha\|^2$ and $\left\langle \xi_j, \alpha - \sum\limits_{i=1}^{k} a_i\xi_i \right\rangle = 0$, $\forall j = 1, \ldots, k$. Hence, we see that $\alpha - \sum\limits_{i=1}^{k} a_i\xi_i$ is orthogonal to each vector in $\langle \xi_1, \ldots, \xi_k\rangle$.

Proof: Consider

$$\begin{aligned}\left\|\alpha - \sum_{i=1}^{k} x_i\xi_i\right\|^2 &= \left\langle \alpha - \sum_{i=1}^{k} x_i\xi_i, \alpha - \sum_{i=1}^{k} x_i\xi_i \right\rangle = \langle\alpha,\alpha\rangle - \sum_{i=1}^{k} x_i a_i - \sum_{i=1}^{k} x_i a_i + \sum_{i=1}^{k} x_i x_i \\ &= \|\alpha\|^2 + \sum_{i=1}^{k}(a_i - x_i)(a_i - x_i) - \sum_{i=1}^{k}|a_i|^2 = \|\alpha\|^2 + \sum_{i=1}^{k}|a_i - x_i|^2 - \sum_{i=1}^{k}|a_i|^2.\end{aligned}$$

11.5 Least Squares Problem (Optional)

Let V be an inner product space with W as its finite dimensional subspace. Given $\alpha \in V \setminus W$. We want to find $\beta \in W$ such that $\|\alpha - \beta\| = \min\limits_{\xi\in W} \|\alpha - \xi\|$. One way of doing this is to find an orthonormal basis of W and proceed as above.

An alternative method is to pick any basis, say $\mathscr{A} = \{\eta_1, \ldots, \eta_k\}$, of W. Suppose $\beta = \sum\limits_{j=1}^{k} x_j\eta_j \in W$ is the required vector, its existence is asserted in Theorem 11.4.5 yet to be determined.

By Theorem 11.4.5, we have $\langle \eta_i, \alpha - \beta\rangle = 0$ $\forall i = 1, \ldots, k$. So $\left\langle \eta_i, \alpha - \sum\limits_{j=1}^{k} x_j\eta_j \right\rangle = 0$ and then

$\langle \eta_i, \alpha\rangle = \sum_{j=1}^{k} \langle \eta_i, \eta_j\rangle x_j$. This system has the matrix form $G\boldsymbol{x} = N$, where $\boldsymbol{x} = \begin{bmatrix} x_1 & \cdots & x_k \end{bmatrix}^t$, $N = \begin{bmatrix} \langle \eta_1, \alpha\rangle & \cdots & \langle \eta_k, \alpha\rangle \end{bmatrix}^t$ and $G = (\langle \eta_i, \eta_j\rangle)$.

The matrix G is called the *Gram matrix* of $\mathscr{A}$. It is known that G is invertible if $\mathscr{A}$ is linearly independent.

Now, we have a unique solution β. Such vector β is called the *best approximation to* α in the least squares sense.

Example 11.5.1: Let $W = \{(x, y, z, w) \in \mathbb{R}^4 \mid x - y - z + w = 0\}$ be a subspace of $\mathbb{R}^4$ under the usual inner product (i.e., the dot product). Find a vector $\beta \in W$ that is closest to $\alpha = (-2, 1, 2, 1)^t$.

Solution:

Given $A \in M_{m,k}$ and $B \in M_{m,1}$. We want to find a matrix $X_0 \in M_{k,1}$ such that $\|AX_0 - B\|$ is minimum among all $k \times 1$ matrices X. Let $W = \mathcal{C}(A)$ be the column space of A, i.e., $W = \{AX \mid X \in \mathbb{R}^k\}$. By Theorem 11.4.5, we know that there is an X_0 such that $AX_0 - B$ is orthogonal to each vector in W. Thus using the dot product as our inner product, we must have

$$(AX)^t(AX_0 - B) = 0 \text{ for all } X \in \mathbb{R}^k.$$

That is, $X^tA^t(AX_0 - B) = 0$ for all X. This could happen only if $A^t(AX_0 - B) = \mathbf{0}$ (by item (d) of Proposition 11.1.4). That means X_0 is a solution of the so-called *normal equation*

$$A^tAX = A^tB.$$

Moreover, it is known that if $\operatorname{rank}(A) = k$, then A^tA is non-singular. So the normal equation has unique solution in $X_0 = (A^tA)^{-1}(A^tB)$.

Following example is a well-known problem in statistics called regression.

Example 11.5.2: We would like to find a polynomial of degree n such that it fits given m points $(x_1, y_1), \ldots, (x_m, y_m)$ in the plane in the least squares sense, in general, $m > n + 1$. We put $y = \sum_{j=0}^{n} c_j x^j$, where c_j's are to be determined such that $\sum_{i=1}^{m} (\hat{y}_i - y_i)^2$ is the least. Put $\hat{y}_i = \sum_{j=0}^{n} c_j x_i^j$, $i = 1, \ldots, m$. Then we have to solve the system

$$\begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^n \\ 1 & x_2 & x_2^2 & \cdots & x_2^n \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_m & x_m^2 & \cdots & x_m^n \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_m \end{bmatrix}$$

in the least squares sense. Thus we need to find an X_0 such that

$$\|AX_0 - B\| = \min_{X\in\mathbb{R}^{n+1}} \|AX - B\|$$

with

$$A = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^n \\ 1 & x_2 & x_2^2 & \cdots & x_2^n \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_m & x_m^2 & \cdots & x_m^n \end{bmatrix}, \quad X = \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_n \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}.$$

When $n = 1$, the problem is called the *linear regression problem.* In this case

$$A = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix}, \quad X = \begin{bmatrix} c_0 \\ c_1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}.$$

Now $W = \left\langle (1, 1, \ldots, 1)^t,\ (x_1, x_2, \ldots, x_m)^t \right\rangle$. In practice as, $x_1, \ldots, x_m$ are not all equal, so $\dim W = 2$. Then

$$A^tA = \begin{bmatrix} m & \sum\limits_{i=1}^m x_i \\ \sum\limits_{i=1}^m x_i & \sum\limits_{i=1}^m x_i^2 \end{bmatrix} = \begin{bmatrix} m & m\bar{x} \\ m\bar{x} & \sum\limits_{i=1}^m x_i^2 \end{bmatrix}, \quad A^tB = \begin{bmatrix} \sum\limits_{i=1}^m y_i \\ \sum\limits_{i=1}^m x_iy_i \end{bmatrix} = \begin{bmatrix} m\bar{y} \\ \sum\limits_{i=1}^m x_iy_i \end{bmatrix},$$

where $\bar{x} = \frac{1}{m}\sum\limits_{i=1}^m x_i$ and $\bar{y} = \frac{1}{m}\sum\limits_{i=1}^m y_i$.

Then $\det(A^tA) = m\sum\limits_{i=1}^m x_i^2 - m^2\bar{x}^2 = m\sum\limits_{i=1}^m (x_i - \bar{x})^2$. Hence

$$X_0 = (A^tA)^{-1}(A^tB) = \frac{1}{m\sum\limits_{i=1}^m (x_i - \bar{x})^2} \begin{bmatrix} m\bar{y}\sum\limits_{i=1}^m x_i^2 - m\bar{x}\sum\limits_{i=1}^m x_iy_i \\ m\sum\limits_{i=1}^m x_iy_i - m^2\bar{x}\bar{y} \end{bmatrix}.$$

So

$$c_1 = \frac{\sum\limits_{i=1}^m x_iy_i - m\bar{x}\bar{y}}{\sum\limits_{i=1}^m (x_i - \bar{x})^2} = \frac{\sum\limits_{i=1}^m (x_i - \bar{x})(y_i - \bar{y})}{\sum\limits_{i=1}^m (x_i - \bar{x})^2} \quad \text{and} \quad c_0 = \frac{\bar{y}\sum\limits_{i=1}^m x_i^2 - \bar{x}\sum\limits_{i=1}^m x_iy_i}{\sum\limits_{i=1}^m (x_i - \bar{x})^2}.$$

■