第三章 内积空间

3.1 内积与 Gram 矩阵

定义（欧几里德空间）设 V 是实数域 $\mathbb{R}$ 上的线性空间. 映射

$$\tau: V\times V\to\mathbb{R}$$

称为 V 上的一个内积，并记 $\tau(v_1,v_2)=\langle v_1,v_2\rangle$，如果它满足:

1）对称性：对任意 $v_1,v_2\in V$，$\langle v_2,v_1\rangle=\langle v_1,v_2\rangle$；

2）对第二个变元的线性性：对任意 $v_1,v_2,v_3\in V$，$k,l\in\mathbb{R}$，

$$\langle v_1,v_2k+v_3l\rangle=\langle v_1,v_2\rangle k+\langle v_1,v_3\rangle l\text{；}$$

3）正定性：对任意 $v\in V,v\neq 0$，$\langle v,v\rangle>0$.

若在 V 上指定了一个内积，则称 V 关于该内积为一个实内积空间. 有限维的实内积空间也称为欧几里德空间，简称为欧氏空间.

注（双线性）首先注意下面事实：固定v_1，让v变化，则由内积可决定一个映射如下

$$\langle v_1,\cdot\rangle : V \to \mathbb{R}$$
$$v \mapsto \langle v_1, v\rangle.$$

定义中的性质2)就是说该映射$\langle v_1,\cdot\rangle$是从V到$\mathbb{R}$的线性映射，因此称之为对第二个变元是线性的．由性质 1)，2)可推出内积对第一个变元也是线性的．

注（$\mathbb{R}$ 上的标准欧几里德空间 $\mathbb{R}^n$）

首先，$\mathbb{R}^n$ 是 $\mathbb{R}$ 上的标准线性空间. 定义

$$
\begin{aligned}
\langle x, y\rangle &= x^{\mathrm{T}} y \\
&= \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \\
&= x_1 y_1 + x_2 y_2 + \cdots + x_n y_n .
\end{aligned}
$$

容易验证，这样定义的 $\langle\cdot,\cdot\rangle$ 是 $\mathbb{R}^n$ 上的内积，称为 $\mathbb{R}^n$ 上的标准内积，相应地，$\mathbb{R}^n$ 称为标准欧几里德空间.

注（几何空间作为欧几里德空间）考虑几何空间. 集合

$$
\begin{aligned}
V &= \{\text{有向线段}\} \\
&= \left\{\overrightarrow{AB} : A, B\text{取遍几何空间}\right\},
\end{aligned}
$$

关于平行四边形法则或三角形法则定义加法，同向或反向伸缩定义数乘法，构成 $\mathbb{R}$ 上的线性空间. 对任意 $\overrightarrow{OA}, \overrightarrow{OB} \in V$ 定义

$$
\langle \overrightarrow{OA}, \overrightarrow{OB} \rangle = |\overrightarrow{OA}| \cdot |\overrightarrow{OB}| \cdot \cos \angle AOB ,
$$

称为向量 $\overrightarrow{OA}, \overrightarrow{OB}$ 的点乘，或数量积. 这里，符号 $|\cdot|$ 表示有向线段的长度. 可以验证，$\langle \cdot, \cdot \rangle$ 是 V 上的内积.

只验证性质 2)．验证过程并非显然(其实，在中学数学课本或大学高等数学中向量代数与空间解析几何章节中就有)．可基于实数 $|\overrightarrow{OB}|\cdot\cos\angle AOB$ 的如下的几何意义：以 O 为原点，以 $\overrightarrow{OA}$ 方向为正方向，使得 $\overrightarrow{OA}$ 所在的直线成为一数轴，则 $|\overrightarrow{OB}|\cdot\cos\angle AOB$ 就是有向线段 $\overrightarrow{OB}$ 在 $\overrightarrow{OA}$ 方向上的投影矢量(有向线段)在该数轴上的坐标.

注（线性空间$\mathbb{R}^2$上的非标准内积的例）定义

$$\langle x, y\rangle = x_1y_1 + \frac{1}{2}x_1y_2 + \frac{1}{2}x_2y_1 + x_2y_2.$$

验证它是内积:

1） 对称性:

2） x 固定，作为 y 的函数是线性的:

$$\begin{aligned}\langle x, y\rangle &= x_1y_1 + \frac{1}{2}x_1y_2 + \frac{1}{2}x_2y_1 + x_2y_2 \\ &= \left(x_1 + \frac{1}{2}x_2\right)y_1 + \left(\frac{1}{2}x_1 + x_2\right)y_2\end{aligned}$$

3） 正定性:

$$\begin{aligned}\langle x, x\rangle &= x_1x_1 + x_1x_2 + x_2x_2 \\ &= (x_1 + \tfrac{1}{2}x_2)^2 + \tfrac{3}{4}(x_2)^2.\end{aligned}$$

若 $x \neq 0$，则 $x_1 + \frac{1}{2}x_2$ 和 x_2 两数不全为零，从而 $\langle x,x\rangle > 0$

注（平方可积函数空间$\mathcal{L}^2([a,b],\mathbb{R}^n)$）

$\mathcal{L}^2([a,b],\mathbb{R}^n)=\{f \mid f\in\mathcal{F}([a,b],\mathbb{R}^n),\text{且}\int_a^b (f(t))^{\mathrm{T}} f(t)\mathrm{d}t<+\infty\}$，

即以定义在区间$[a,b]$上取之于$\mathbb{R}^n$的平方可积函数为元素的集合.

记

$$f(t)=[f_1(t),\cdots,f_n(t)]^{\mathrm{T}},$$

则

$$\int_a^b (f(t))^{\mathrm{T}} f(t)\mathrm{d}t=\int_a^b [(f_1(t))^2+\cdots+(f_n(t))^2]\mathrm{d}t.$$

首先证明，它是一个线性子空间(不等式)．其次，对定义

$$\langle f,g\rangle=\int_a^b (f(t))^{\mathrm{T}} g(t)\mathrm{d}t=\int_a^b [f_1(t)g_1(t)+\cdots+f_n(t)g_n(t)]\mathrm{d}t.$$

可以证明，这是线性空间$\mathcal{L}^2([a,b],\mathbb{R}^n)$上的一个内积.

注 （概率论中的内积空间 $\mathcal{L}^2(\Omega, \mathcal{B}, P(\cdot))$ ） 概率空间 $(\Omega, \mathcal{B}, P(\cdot))$ 上的有二阶矩的随机变量的集合． $X : \Omega \to \mathbb{R}$ 是随机变量；

$$E(X^2) = \int_{\Omega} (X(\omega))^2 P(d\omega) < +\infty .$$

由概率论知：

（1） 若 $X \in \mathcal{L}^2(\Omega, \mathcal{B}, P(\cdot))$， 则 X 的数学期望 $E(X)$ 和方差 $D(X)$ 都存在；

（2） 若 $X, Y \in \mathcal{L}^2(\Omega, \mathcal{B}, P(\cdot))$， 则 XY 的数学期望 $E(XY)$ 存在.

首先， 可以证明 $\mathcal{L}^2(\Omega, \mathcal{B}, P(\cdot))$ 是线性空间； 其次， 定义

$$\langle X, Y \rangle = E(XY) ,$$

可以证明， 这是线性空间 $\mathcal{L}^2(\Omega, \mathcal{B}, P(\cdot))$ 上的一个内积.

定义（酉空间）设V是复数域$\mathbb{C}$上的线性空间．映射

$$\tau: V\times V\to\mathbb{C}$$

称为V上的一个内积，并记$\tau(v_1,v_2)=\langle v_1,v_2\rangle$，如果它满足：

1）共轭对称性：对任意$v_1,v_2\in V$，$\overline{\langle v_2,v_1\rangle}=\langle v_1,v_2\rangle$；

2）对第二个变元的线性性：对任意$v_1,v_2,v_3\in V$，$k,l\in\mathbb{C}$，

$$\langle v_1, v_2k+v_3l\rangle=\langle v_1,v_2\rangle k+\langle v_1,v_3\rangle l;$$

3）正定性：对任意$v\in V, v\neq 0$，$\langle v,v\rangle$为实数，且$\langle v,v\rangle>0$．

指定了内积的复线性空间称为复内积空间．有限维的复内积空间也称为酉空间

命题 (非负性) 对任意$v \in V$，$\langle v, v \rangle$为实数，且$\langle v, v \rangle \geq 0$. 等号成立，即$\langle v, v \rangle = 0$，当且仅当$v = 0$.

命题 (复内积对第一个变元是共轭线性的)

$$\langle v_1 k + v_2 l, v_3 \rangle = \bar{k} \langle v_1, v_3 \rangle + \bar{l} \langle v_2, v_3 \rangle .$$

命题 (零向量与其他向量的内积)

对任意$v \in V$，$\langle v, 0 \rangle = \langle 0, v \rangle = 0$.

注（$\mathbb{C}$ 上的标准酉空间 $\mathbb{C}^n$）定义

$$
\begin{aligned}
\langle x, y\rangle &= \bar{x}^{\mathrm{T}} y \\
&= \begin{bmatrix} \bar{x}_1 & \bar{x}_2 & \cdots & \bar{x}_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \\
&= \bar{x}_1 y_1 + \bar{x}_2 y_2 + \cdots + \bar{x}_n y_n .
\end{aligned}
$$

容易验证，这样定义的 $\langle\cdot,\cdot\rangle$ 是 $\mathbb{C}^n$ 上的内积，称为 $\mathbb{C}^n$ 上的标准内积，相应地，$\mathbb{C}^n$ 称为标准酉空间.

定义（复矩阵的共轭转置）设 $A \in \mathbb{C}^{m\times n}$. 定义记号

$$A^{\mathrm{H}} := \left(\overline{A}\right)^{\mathrm{T}} = \left[b_{ij}\right]_{n\times m},$$

这里，$b_{ij} = \overline{a}_{ji}$ 表示复数 a_{ij} 的共轭.

于是，$\mathbb{C}^{n}$ 上的标准内积可写成矩阵乘法

$$\langle x, y\rangle = x^{\mathrm{H}} y .$$

注（抽象向量内积的抽象矩阵运算表达）将两个向量的内积运算写成两个抽象 1×1 矩阵之间的形式乘法运算如下

$$\langle \alpha, \beta\rangle = \alpha^{\mathrm{H}} \beta .$$

这种记号是基于将抽象向量类比成列数组的想法.（若是类比成行数组，则有一套对偶的记号.）

以下我们只对复内积空间来叙述.

命题（线性组合的内积的矩阵表示）

$$\left\langle \sum_{i=1}^{s} \alpha_i k_i, \quad \sum_{j=1}^{t} \beta_j l_j \right\rangle = \begin{bmatrix} \bar{k}_1 & \cdots & \bar{k}_s \end{bmatrix} \left[\langle \alpha_i, \beta_i \rangle\right]_{s\times t} \begin{bmatrix} l_1 \\ \vdots \\ l_t \end{bmatrix}$$

$$= \overline{\begin{bmatrix} k_1 \\ \vdots \\ k_s \end{bmatrix}}^{\mathrm{T}} \left[\langle \alpha_i, \beta_i \rangle\right]_{s\times t} \begin{bmatrix} l_1 \\ \vdots \\ l_t \end{bmatrix}$$

因此，只要知道两向量组之间，诸对向量的内积，那么，由这两向量组各自任意作线性组合做出的一对向量，其内积可由上述诸对向量的内积及线性组合的系数来计算. 两向量组之间所有配对的内积拼成矩阵，命名为：

定义（两向量组的交互 Gram 矩阵）

设 $\alpha_1,\alpha_2,\ldots,\alpha_s$ 和 $\beta_1,\beta_2,\ldots,\beta_t$ 是复内积空间的两个向量组. 矩阵

$$G(\alpha_1,\alpha_2,\ldots,\alpha_s;\beta_1,\ldots,\beta_t):=\left[\langle\alpha_i,\beta_j\rangle\right]_{s\times t}\in\mathbb{C}^{s\times t}$$

称为 $\alpha_1,\alpha_2,\ldots,\alpha_s$ 和 $\beta_1,\beta_2,\ldots,\beta_t$ 的交互 Gram 矩阵. 简记为 $G(\{\alpha_i\};\{\beta_j\})$.

注（矩阵乘积作为两向量组的交互 Gram 矩阵）

作为标准酉空间 $\mathbb{C}^n$ 上向量组，矩阵 $A\in\mathbb{C}^{n\times s}$ 的列向量组(简称为列向量组 A)和 $B\in\mathbb{C}^{n\times t}$ 的列向量组(简称为列向量组 B)之间的交互 Gram 矩阵为

$$G=A^{\mathrm{H}}B.$$

反之，给定任何矩阵乘积关系

$$C=AB$$

定义（向量组的 Gram 矩阵，内积空间的度量矩阵）
设 $\alpha_1,\alpha_2,\ldots,\alpha_s$ 是复内积空间的一个向量组. 矩阵

$$G(\alpha_1,\alpha_2,\ldots,\alpha_s):=\left[\langle\alpha_i,\alpha_j\rangle\right]_{s\times s}\in\mathbb{C}^{s\times s}$$

称为向量组 $\alpha_1,\alpha_2,\ldots,\alpha_s$ 的 Gram 矩阵. 简记为 $G(\{\alpha_i\})$. 基向量组的 Gram 矩阵称为该基的度量矩阵.

注（抽象向量内积符号的抽象矩阵运算表达）

$$G(\{\alpha_i\};\{\beta_j\})=\begin{bmatrix}\alpha_1 & \cdots & \alpha_s\end{bmatrix}^{\mathrm{H}}\begin{bmatrix}\beta_1 & \cdots & \beta_t\end{bmatrix}.$$

行体制时对偶.

定理（抽象内积通过坐标系和度量矩阵表现为具体内积）

设 $\alpha_1,\alpha_2,\dots,\alpha_n$ 是复内积空间 V 的一个基，

$$G = G(\alpha_1,\dots,\alpha_n) \in \mathbb{C}^{n\times n}$$

为其度量矩阵．若向量 $\alpha,\beta \in V$ 在该基下的坐标分别为 $x, y \in \mathbb{C}^n$，则

$$\langle \alpha, \beta \rangle = \overline{x}^{\mathrm{T}} G y \text{.}$$

特别地，若 G 为单位矩阵 I_n，则 V 上的内积通过坐标化为 $\mathbb{C}^n$ 上的标准内积.

注（任意抽象内积都可以通过适当选择坐标系表现为 $\mathbb{C}^n$ 上的标准内积）

给定内积空间，不同的基底一般会有不同的度量矩阵. 可以证明，总可以选择一组基底，使得其度量矩阵是单位矩阵. 这样的基底就是标准正交基. 标准正交基的存在性可由 Gram-Schmidt 正交化方法或者 Hermit 矩阵的性质来证明.

3.2 向量的长度和夹角

定义（向量的长度）设α是复内积空间的一个向量，记

$$\|\alpha\| := \sqrt{\langle \alpha, \alpha \rangle},$$

称为向量α的长度．向量α和β之间的距离定义为

$$d(\alpha, \beta) := \| \alpha - \beta \|.$$

命题（向量长度的性质）

1） 正性：对任意向量α，$\|\alpha\| \geq 0$．等号成立，当且仅当$\alpha = 0$.

2） 正齐性：对任意向量α及任意复数k，$\|\alpha k\| = \|\alpha\| \cdot |k|$.

3） 三角不等式：对任意向量α, β，$\|\alpha + \beta\| \leq \|\alpha\| + \|\beta\|$.

4） Cauchy-Schwartz 不等式：对任意向量α, β，$|\langle \alpha, \beta \rangle| \leq \|\alpha\| \cdot \|\beta\|$.

5） 平行四边形公式：对任意向量α, β，

$$\|\alpha + \beta\|^2 + \|\alpha - \beta\|^2 = 2\|\alpha\|^2 + 2\|\beta\|^2.$$

由 Cauchy-Schwartz 不等式，对实内积我们有

$$-1 \le \frac{\langle \alpha, \beta \rangle}{\|\alpha\| \cdot \|\beta\|} \le 1.$$

因此，我们可以有下面的定义

定义（实内积空间两向量之间的夹角）
设 α, β 是实内积空间的两个非零向量．定义 α, β 之间的夹角为

$$\angle_{\alpha}^{\beta} := \arccos \frac{\langle \alpha, \beta \rangle}{\|\alpha\| \cdot \|\beta\|}.$$

定义（复内积空间两向量正交）

设 α,β 是复(实)内积空间的两个非零向量．若 $\langle\alpha,\beta\rangle=0$，则称 α,β 相互正交，记为 $\alpha\perp\beta$．在实内积空间的情形，α,β 相互正交，也就是 $\angle_{\alpha}^{\beta}=\frac{\pi}{2}$．

显然，零向量与任何向量正交．反之，与任何向量都正交的向量必是零向量．

命题（勾股定理）

设 α,β 是实内积空间的两个非零向量．下列各条件等价：

（1） $\alpha\perp\beta$；（2） $\angle_{\alpha}^{\beta}=\dfrac{\pi}{2}$；（3） $\langle\alpha,\beta\rangle=0$；

（4） $\|\alpha\|^2+\|\beta\|^2=\|\alpha+\beta\|^2$；（5） $\|\alpha\|^2+\|\beta\|^2=\|\alpha-\beta\|^2$．

在复内积空间情形，有

（1）$\Leftrightarrow$（3）$\Rightarrow$（4） $\Leftrightarrow$ （5）．

注（复内积空间勾股定理）

在复内积空间的情形，易见

$$\begin{aligned}\|\alpha+\beta\|^2 &= \|\alpha\|^2+\|\beta\|^2+\langle\alpha,\beta\rangle+\langle\beta,\alpha\rangle\\ &= \|\alpha\|^2+\|\beta\|^2+2\operatorname{Re}\langle\alpha,\beta\rangle.\end{aligned}$$

于是，$\|\alpha\|^2+\|\beta\|^2=\|\alpha+\beta\|^2$ 等价于 $\operatorname{Re}\langle\alpha,\beta\rangle=0$．只需举例说明有 $\operatorname{Re}\langle\alpha,\beta\rangle=0$，但 $\langle\alpha,\beta\rangle\neq 0$ 的情形．考虑标准复内积空间 $\mathbb{C}^2$．取

$$\alpha=\begin{bmatrix}0\\1\end{bmatrix},\quad \beta=\begin{bmatrix}1\\0\end{bmatrix}+\begin{bmatrix}1\\1\end{bmatrix}i.$$

则

$$\langle\alpha,\beta\rangle=\overline{\alpha}^{\mathrm{T}}\beta=i\neq 0,$$

但

$$\operatorname{Re}\langle\alpha,\beta\rangle=0.$$

3.3 标准正交基与酉矩阵

定义 (正交组，标准正交组，标准正交基) 给定复内积空间V .

(1) V 中若干个非零向量构成的向量组称为正交向量组，简称正交组若组中任意两个向量正交.

(2) 正交向量组称为标准正交组，若其中每个向量的长度为 1.

(3) 标准正交组若还是V 的一个基，则称为标准正交基.

定理 (正交组的 Gram 矩阵)

(1) 一个向量组为正交向量组的充要条件是其 Gram 矩阵为非奇异对角矩阵.

(2) 一个向量组为标准正交组的充要条件是其 Gram 矩阵为单位矩阵.

推论 (正交组是线性无关组)

此推论亦可直接证明

内积空间中标准正交组的好处体现在

定理（沿正交组展开时坐标的解耦性）

设$\alpha_1,\alpha_2,\ldots,\alpha_s$是复内积空间$V$的一个正交组， $\beta\in V$． 若有

$$\beta=\alpha_1k_1+\alpha_2k_2+\cdots+\alpha_sk_s$$
$$=\begin{bmatrix}\alpha_1 & \alpha_2 & \cdots & \alpha_s\end{bmatrix}\begin{bmatrix}k_1\\k_2\\\vdots\\k_s\end{bmatrix}$$

则

$$k_i=\frac{\langle\alpha_i,\beta\rangle}{\langle\alpha_i,\alpha_i\rangle},\ i=1,2,\ldots,s.$$

特别地， 若$\alpha_1,\alpha_2,\ldots,\alpha_s$进而是标准正交组， 则

$$k_i=\langle\alpha_i,\beta\rangle,\ i=1,2,\ldots,s.$$

Gram-Schmidt 正交化方法

下面给出一种将内积空间V 中线性无关的向量组$\beta_1,\beta_2,\ldots,\beta_s$递归地改造成正交组$\alpha_1,\alpha_2,\ldots,\alpha_s$的方法，称为 Gram-Schmidt 正交化方法.

$$\alpha_1=\beta_1,$$
$$\alpha_j=\beta_j-\left(\alpha_1 k_{1,j}+\alpha_2 k_{2,j}+\cdots+\alpha_{j-1}k_{j-1,j}\right),\ j=2,3,\ldots,s.$$

这里，所有系数$k_{ij}, i<j$，按如下方式递归地计算

$$k_{ij}=\frac{\langle \alpha_i,\beta_j\rangle}{\langle \alpha_i,\alpha_i\rangle}, i<j.$$

易知，这样计算系数就保证了$\alpha_1,\alpha_2,\ldots,\alpha_s$的正交性

定理 (Gram-Schmidt 正交化方法保持子空间序列)

$$\operatorname{span}\{\alpha_1,\alpha_2,\ldots,\alpha_t\}=\operatorname{span}\{\beta_1,\beta_2,\ldots,\beta_t\},\ t=1,2,\ldots,s.$$

定理（矩阵的正交-三角分解）设 $A\in\mathbb{C}^{n\times r}$ 列满秩. 则存在唯一的一对矩阵 $Q\in\mathbb{C}^{n\times r}$ 和 $R\in\mathbb{C}^{r\times r}$ 满足:

（1） $A=QR$；

（2） $Q^HQ=I_r$， 或者说， Q 的列向量组是标准酉空间 $\mathbb{C}^n$ 中的标准正交组;

（3） R 是对角线元素为正实数的上三角矩阵.

特别地，若 A 为非奇异方阵，则相应的 Q 为酉矩阵.

酉矩阵

本小节研究标准酉空间 $\mathbb{C}^n$， 结果自然也包括了标准欧几里得空间 $\mathbb{R}^n$ 的情形. 回忆向量 $x, y \in \mathbb{C}^n$ 的标准内积

$$\langle x, y\rangle = \bar{x}_1 y_1 + \cdots + \bar{x}_n y_n = x^{\mathrm{H}} y .$$

定义 （酉矩阵作为标准酉空间中的标准正交基的基矩阵）
标准酉空间中的任一标准正交基所拼成的矩阵称为一个复正交矩阵, 常称为酉矩阵. 换言之, 复数域上的一个方阵称为酉矩阵, 如果其列向量组构成标准酉空间的一个标准正交基.

相应地, 有实正交矩阵.

定义（用矩阵的逆来定义酉矩阵）

矩阵 $U \in \mathbb{C}^{n\times n}$ 称为酉矩阵，如果 $U^{-1} = U^{\mathrm{H}}$，即

$$U^{\mathrm{H}}U = I_n .$$

矩阵的求逆运算 $(\cdot)^{-1}$ 相比于矩阵的求共轭转置运算 $(\cdot)^{\mathrm{H}}$，一般说来，复杂程度相去天渊。这就从矩阵运算的角度说明酉矩阵之好，也说明了酉矩阵在计算数学中的重要地位.

定义（酉相似）

设 $A, B \in \mathbb{C}^{n\times n}$. 若有酉矩阵 $U \in \mathbb{C}^{n\times n}$ 使得

$$U^{\mathrm{H}}AU = B ,$$

则称 A 与 B 酉相似.

酉相似的几何意义是 A 作为线性变换在标准正交基 U 下的矩阵表示为 B .

定理 (酉矩阵保持度量)

设$U \in \mathbb{C}^{n\times n}$，则下列等价

(1) U 为酉矩阵，即U 的列向量组是标准酉空间$\mathbb{C}^n$ 中的标准正交基.

(2) U^{H} 为酉矩阵，即U 的行向量组的共轭转置是标准酉空间$\mathbb{C}^n$ 中的标准正交基.

(3) $U^{\mathrm{H}}U = I_n$.

(4) $UU^{\mathrm{H}} = I_n$.

(5) (保内积) 对任意$x, y \in \mathbb{C}^n$,

$$\langle Ux, Uy\rangle = \langle x, y\rangle .$$

(6) (保长度) 对任意$x \in \mathbb{C}^n$,

$$\|Ux\| = \|x\| .$$

例（Givens 旋转，Householder 镜像）

3.4 正交投影与最佳逼近

问题（最佳逼近问题）

设 V 为复内积空间，向量 $\beta \in V$， W 为 V 的一个子空间. 求 $x^* \in W$，使得

$$d(\beta, x^*) = \min\{d(\beta, x) : x \in W\}.$$

即

$$d(\beta, x^*) \le d(\beta, x),\ \text{对任意 } x \in W.$$

或者写成优化问题的通常形式：

$$\min \|x - \beta\|,$$
$$\text{s.t. } x \in W.$$

我们关心解 x^* 的存在性，唯一性，及其具体求法.

许多应用科学问题可抽象为此问题模式.

- W 为一维子空间的情形.

$$W=\operatorname{span}\{\alpha\},\quad \alpha\neq 0.$$

解法一: 化为函数极值问题(以实内积空为例来说明)

W 中的元素可有如下的参数化表示:

$$W=\{\alpha k : k\in\mathbb{R}\}$$

$$f(k)=(d(\beta,\alpha k))^2=\|\beta-\alpha k\|^2=\langle\beta-\alpha k,\beta-\alpha k\rangle$$

$$=\langle\alpha,\alpha\rangle k^2-2\langle\alpha,\beta\rangle k+\langle\beta,\beta\rangle$$

是关于 k 的二次函数, 可知其有唯一的最小值, 最小值点为

$$k^*=-\frac{-2\langle\alpha,\beta\rangle}{2\langle\alpha,\alpha\rangle}=\frac{\langle\alpha,\beta\rangle}{\langle\alpha,\alpha\rangle}=\langle\alpha,\alpha\rangle^{-1}\langle\alpha,\beta\rangle.$$

于是问题的唯一解为

$$x^*=\alpha\langle\alpha,\alpha\rangle^{-1}\langle\alpha,\beta\rangle$$

用内积的抽象矩阵乘法运算表达为

$$x^*=\alpha\left(\alpha^{\mathrm{T}}\alpha\right)^{-1}\alpha^{\mathrm{T}}\beta$$

解法二：用勾股定理直接定系数

若 $\beta-\alpha k^{*}\perp W$，注意到

$$(\beta-\alpha k)-(\beta-\alpha k^{*})=\alpha(k^{*}-k)\in W,$$

由勾股定理得

$$\|\beta-\alpha k\|^{2}=\|\beta-\alpha k^{*}\|^{2}+\|\alpha(k^{*}-k)\|^{2}\geq\|\beta-\alpha k^{*}\|^{2}$$

因 此， 使 得 $\beta-\alpha k^{*}\perp W$ 的 αk^{*} 为 所 求． 又 $\beta-\alpha k^{*}\perp W$ 等 价 于

$\beta-\alpha k^{*}\perp\alpha$， 于是

$$\langle\alpha,\beta-\alpha k\rangle=0$$

即

$$\langle\alpha,\alpha\rangle k=\langle\alpha,\beta\rangle$$

$$k^{*}=\langle\alpha,\alpha\rangle^{-1}\langle\alpha,\beta\rangle$$

- W 为有限维子空间的情形.

设$\alpha_1,\alpha_2,\ldots,\alpha_m$ 为W 的一组基.

记

$$x=\alpha_1k_1+\alpha_2k_2+\cdots\alpha_mk_m=\begin{bmatrix}\alpha_1 & \cdots & \alpha_m\end{bmatrix}\begin{bmatrix}k_1\\ \vdots\\ k_m\end{bmatrix}$$

$$\beta-x\perp W \iff \alpha_i\perp\beta-x,\ i=1,\ldots,m$$

于是,

$$\langle\alpha_i,\beta-x\rangle=0,i=1,\ldots,m$$

写成矩阵形式,

$$\begin{bmatrix}\langle\alpha_1,\alpha_1\rangle & \cdots & \langle\alpha_1,\alpha_m\rangle\\ \vdots & \vdots & \vdots\\ \langle\alpha_m,\alpha_1\rangle & \cdots & \langle\alpha_m,\alpha_m\rangle\end{bmatrix}\begin{bmatrix}k_1\\ \vdots\\ k_m\end{bmatrix}=\begin{bmatrix}\langle\alpha_1,\beta\rangle\\ \vdots\\ \langle\alpha_m,\beta\rangle\end{bmatrix}$$

$$x^* = \begin{bmatrix} \alpha_1 & \cdots & \alpha_m \end{bmatrix} \begin{bmatrix} k_1 \\ \vdots \\ k_m \end{bmatrix}^*$$

$$= \begin{bmatrix} \alpha_1 & \cdots & \alpha_m \end{bmatrix} G(\alpha_1, \alpha_2, \ldots, \alpha_m)^{-1} \begin{bmatrix} \langle \alpha_1, \beta \rangle \\ \vdots \\ \langle \alpha_m, \beta \rangle \end{bmatrix}$$

$$= \begin{bmatrix} \alpha_1 & \cdots & \alpha_m \end{bmatrix} G(\alpha_1, \alpha_2, \ldots, \alpha_m)^{-1} G(\alpha_1, \alpha_2, \ldots, \alpha_m; \beta).$$

用交互 Gram 矩阵的抽象矩阵乘法运算表达

$$x^* = \begin{bmatrix} \alpha_1 & \cdots & \alpha_m \end{bmatrix} \left(\begin{bmatrix} \alpha_1 & \cdots & \alpha_m \end{bmatrix}^{\mathrm{H}} \begin{bmatrix} \alpha_1 & \cdots & \alpha_m \end{bmatrix} \right)^{-1} \begin{bmatrix} \alpha_1 & \cdots & \alpha_m \end{bmatrix}^{\mathrm{H}} \beta$$

进一步，将 W 和它的基矩阵 $\begin{bmatrix}\alpha_1 & \cdots & \alpha_m\end{bmatrix}$ 用同一符号 W 表示，可得更紧凑的表示

$$x^* = W\left(W^{\mathrm{H}}W\right)^{-1}W^{\mathrm{H}}\beta.$$

固定子空间 W，此公式决定了一个从 V 到 W 的映射

$$\mathcal{P}_W : \beta \mapsto x^* = W\left(W^{\mathrm{H}}W\right)^{-1}W^{\mathrm{H}}\beta.$$

可形式地记

$$\mathcal{P}_W(\cdot) = W\left(W^{\mathrm{H}}W\right)^{-1}W^{\mathrm{H}}(\cdot).$$

这是向子空间 W 的正交投影算子.

例（标准酉空间$\mathbb{C}^n$中，向子空间$\mathrm{im}A$的正交投影矩阵）

这里，A是列满秩的. 记矩阵

$$P_A = A\left(A^{\mathrm{H}}A\right)^{-1}A^{\mathrm{H}}.$$

则

$$x^* = P_A\beta.$$

对固定的A，线性映射$\beta \mapsto x^*$是从酉空间$V=\mathbb{C}^n$到子空间$W=\mathrm{im}A$的投影算子，它由投影矩阵P_A决定.

例（最小二乘数据拟合）

对"因变量" y 和 p 个 "自变量" $x_1, x_2, \ldots, x_p$ 进行 n 次观测，得数据如下表:

编号	y	x_0	x_1	x_2	$\ldots$	x_p
1	y_1	1	x_{11}	x_{12}	$\ldots$	x_{1p}
2	y_2	1	x_{21}	x_{22}	$\ldots$	x_{2p}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ldots$	$\vdots$
n	y_n	1	x_{n1}	x_{n2}	$\ldots$	x_{np}

为方便起见，补充常值变量 $x_0 \equiv 1$ 作为"自变量".

要找“经验公式”

$$
\begin{aligned}
y &= a_0 + a_1 x_1 + a_2 x_2 + \cdots + a_p x_p \\
&= 1a_0 + x_1 a_1 + x_2 a_2 + \cdots + x_p a_p \\
&= \begin{bmatrix} 1 & x_1 & x_2 & \cdots & x_p \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix},
\end{aligned}
$$

即求出“参数”$a_0, a_1, \ldots, a_p$，使得这个函数关系能最好地“拟合”那 n 组观测数据.

最小二乘准则的几何解释：设 $V=\mathbb{R}^n$，

$$Y=\begin{bmatrix} y_1\\ y_2\\ \vdots\\ y_n \end{bmatrix},\quad \alpha_0=\begin{bmatrix} 1\\ 1\\ \vdots\\ 1 \end{bmatrix},\alpha_1=\begin{bmatrix} x_{11}\\ x_{21}\\ \vdots\\ x_{n1} \end{bmatrix},\alpha_2=\begin{bmatrix} x_{12}\\ x_{22}\\ \vdots\\ x_{n2} \end{bmatrix},\cdots,\alpha_p=\begin{bmatrix} x_{1p}\\ x_{2p}\\ \vdots\\ x_{np} \end{bmatrix},$$

$$X=\begin{bmatrix} \alpha_0 & \alpha_1 & \alpha_2 & \cdots & \alpha_p \end{bmatrix}\in\mathbb{R}^{n\times(p+1)}\qquad W=\operatorname{im}X$$

$$k=\begin{bmatrix} a_0\\ a_1\\ a_2\\ \vdots\\ a_p \end{bmatrix}.$$

于是，最小二乘估计 $\hat{k}=\left(X^{\mathrm{T}}X\right)^{-1}X^{\mathrm{T}}Y$.