MATH10003.01 高等线性代数 Fall 2025

Lecture 9 内积空间 — 2025.11.11

教授：邵美悦

Scribe: 路人甲

1 大纲

1. 内积空间
2. 正交基
3. 内积空间上的线性映射

2 内积空间

我们知道，Euclidean 空间的核心概念是长度和夹角。在第二讲和第三讲中，我们讨论了向量空间中的范数的概念。范数 $\|\cdot\|$ 将每个向量映射到一个非负实数，代表它的长度。它使得我们能够在抽象的向量空间中度量距离。现在，我们来看看第二个问题：如何定义向量之间的夹角？

在 $\mathbb{R}^n$ 中，两个向量 $\boldsymbol{u}$ 和 $\boldsymbol{v}$ 之间的夹角 θ 可以通过它们的内积来定义：

$$\cos\theta = \frac{\langle \boldsymbol{u}, \boldsymbol{v}\rangle}{\|\boldsymbol{u}\|_2\|\boldsymbol{v}\|_2} = \frac{\langle \boldsymbol{u}, \boldsymbol{v}\rangle}{\sqrt{\langle \boldsymbol{u}, \boldsymbol{u}\rangle}\sqrt{\langle \boldsymbol{v}, \boldsymbol{v}\rangle}}.$$

内积 $\langle\cdot,\cdot\rangle$ 这个单一的代数运算，同时封装了“长度”“夹角”等几何信息。可以说，只要我们把内积定义好，长度和角度的度量问题就迎刃而解了。

2.1 内积的定义

既然内积在 $\mathbb{R}^n$ 中如此重要，我们能否将其推广到任意的向量空间中呢？当然可以！我们不关心向量是坐标、多项式还是更抽象的函数，我们只关心我们希望这个运算具备哪些最核心的性质。观察 $\mathbb{R}^n$ 中内积的性质，我们抽象出以下定义：

定义 2.1 (内积空间)**.** 设 V 是数域 $\mathbb{F}$ ($\mathbb{F} = \mathbb{R}$ *or* $\mathbb{C}$) 上的向量空间。二元函数 $\langle\cdot,\cdot\rangle : V \times V \to \mathbb{F}$ 若满足以下性质，则称为 V 上的一个**内积**，并称 $(V, \langle\cdot,\cdot\rangle)$ 为一个**内积空间**：

对 $\forall \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in V$, $\forall c \in \mathbb{F}$，有

(1) 非负性： $\langle \boldsymbol{x}, \boldsymbol{x} \rangle \geq 0$.

(1a) 正定性： $\langle \boldsymbol{x}, \boldsymbol{x} \rangle = 0 \Leftrightarrow \boldsymbol{x} = \boldsymbol{0}$.

(2) 可加性： $\langle \boldsymbol{x} + \boldsymbol{y}, \boldsymbol{z} \rangle = \langle \boldsymbol{x}, \boldsymbol{z} \rangle + \langle \boldsymbol{y}, \boldsymbol{z} \rangle$.

(3) 齐次性： $\langle cx, y \rangle = c\langle x, y \rangle$.

(4) Hermitian 对称性： $\langle x, y \rangle = \overline{\langle y, x \rangle}$.

根据定义，我们可以验证 $\mathbb{C}^n$ 上的函数 $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x}^\star \boldsymbol{y}$ 是一个内积。邵老师在数值算法课上有讲过一个更一般的内积：$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{y}^\star \boldsymbol{A} \boldsymbol{x}$. 容易验证，这个函数是内积的充要条件为矩阵 $\boldsymbol{A}$ 是 Hermite 正定阵。

根据内积的定义，我们可以推导出以下性质。设有 $a, b, c, d \in \mathbb{F}$, $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{w} \in \mathbb{F}^n$，则有

(1) $\langle \boldsymbol{x}, c\boldsymbol{y} \rangle = \bar{c}\langle \boldsymbol{x}, \boldsymbol{y} \rangle$.

(2) $\langle \boldsymbol{x}, \boldsymbol{y} + \boldsymbol{z} \rangle = \langle \boldsymbol{x}, \boldsymbol{y} \rangle + \langle \boldsymbol{x}, \boldsymbol{z} \rangle$.

(3) $\langle a\boldsymbol{x} + b\boldsymbol{y}, c\boldsymbol{w} + d\boldsymbol{z} = a\bar{c}\langle \boldsymbol{x}, \boldsymbol{w} \rangle + a\bar{d}\langle \boldsymbol{x}, \boldsymbol{z} \rangle + b\bar{c}\langle \boldsymbol{y}, \boldsymbol{w} \rangle + b\bar{d}\langle \boldsymbol{y}, \boldsymbol{z} \rangle$.

(4) $\langle \boldsymbol{x}, \langle \boldsymbol{x}, \boldsymbol{y} \rangle \boldsymbol{y} \rangle = |\langle \boldsymbol{x}, \boldsymbol{y} \rangle|^2$.

(5) $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0 \ (\forall \boldsymbol{y} \in V)\rangle \Leftrightarrow \boldsymbol{x} = \boldsymbol{0}$.

2.2 范数与 Cauchy-Schwarz 不等式

在定义了内积之后，我们再回过头来看看范数，准确来说，是 2-范数。我们可以通过内积来定义向量的 2-范数：

$$\|\boldsymbol{x}\|_2 = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}.$$

接下来，我们就来验证 2-范数满足范数的三条性质：

(1) 正定性：$\|\boldsymbol{x}\|_2 \geq 0$，当且仅当 $\boldsymbol{x} = \boldsymbol{0}$ 时等号成立。

(2) 齐次性：$\|c\boldsymbol{x}\|_2 = \sqrt{\langle c\boldsymbol{x}, c\boldsymbol{x} \rangle} = \sqrt{c\bar{c}\langle \boldsymbol{x}, \boldsymbol{x} \rangle} = |c|\|\boldsymbol{x}\|_2$.

(3) 三角不等式：$\|\boldsymbol{x} + \boldsymbol{y}\|_2 \leq \|\boldsymbol{x}\|_2 + \|\boldsymbol{y}\|_2$.

要证明三角不等式，我们只需要证明

$$\|\boldsymbol{x} + \boldsymbol{y}\|_2^2 = \langle \boldsymbol{x} + \boldsymbol{y}, \boldsymbol{x} + \boldsymbol{y} \rangle \leq (\|\boldsymbol{x}\|_2 + \|\boldsymbol{y}\|_2)^2.$$

两边展开，我们只需要证明

$$2\mathrm{Re}(\langle \boldsymbol{x}, \boldsymbol{y} \rangle) \leq 2\|\boldsymbol{x}\|_2\|\boldsymbol{y}\|_2.$$

我们将左边放大，只需证

$$|\langle \boldsymbol{x}, \boldsymbol{y}\rangle| \leq \|\boldsymbol{x}\|_2\|\boldsymbol{y}\|_2.$$

这就是著名的 Cauchy-Schwarz 不等式。

定理 2.2 (Cauchy-Schwarz 不等式)**.** *设 $\langle\cdot,\cdot\rangle$ 是数域 $\mathbb{F}$ ($\mathbb{F}=\mathbb{R}$ or $\mathbb{C}$) 上的向量空间 V 上的内积，则对任意 $\boldsymbol{x},\boldsymbol{y}\in V$，有*

$$|\langle \boldsymbol{x}, \boldsymbol{y}\rangle|^2 \leq \langle \boldsymbol{x}, \boldsymbol{x}\rangle\langle \boldsymbol{y}, \boldsymbol{y}\rangle = \|\boldsymbol{x}\|_2^2\|\boldsymbol{y}\|_2^2.$$

当且仅当 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 线性相关时等号成立。

我们将给出两种不同的证法。

证明. 若 $\boldsymbol{x}=\boldsymbol{0}$ 或 $\boldsymbol{y}=\boldsymbol{0}$，不等式显然。接下来我们假设 $\boldsymbol{x}\neq\boldsymbol{0}$ 且 $\boldsymbol{y}\neq\boldsymbol{0}$.

$\langle x,y\rangle\in\mathbb{R}$ 时，这部分的证明是 **trivial** 的。我们考虑函数

$$\begin{aligned} f(\lambda) &= \|\lambda\boldsymbol{x}+\boldsymbol{y}\|_2^2 \\ &= \langle \lambda\boldsymbol{x}+\boldsymbol{y}, \lambda\boldsymbol{x}+\boldsymbol{y}\rangle \\ &= \lambda^2\langle \boldsymbol{x},\boldsymbol{x}\rangle + 2\lambda\langle \boldsymbol{x},\boldsymbol{y}\rangle + \langle \boldsymbol{y},\boldsymbol{y}\rangle \\ &\geq 0, \quad \forall\lambda\in\mathbb{R}. \end{aligned}$$

因为 $f(\lambda)$ 是关于 λ 的二次函数，所以它的判别式小于等于零，即

$$\langle \boldsymbol{x},\boldsymbol{y}\rangle^2 \leq \langle \boldsymbol{x},\boldsymbol{x}\rangle\langle \boldsymbol{y},\boldsymbol{y}\rangle.$$

当且仅当 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 线性相关时等号成立。

当 $\langle x,y\rangle\in\mathbb{C}$ 时，$f(\lambda)$ 的表达式就带有共轭了，不再是一个简单的二次函数。因此，我们需要换一个方案。我们同样考虑

$$\begin{aligned} f(\lambda) &= \|\lambda\boldsymbol{x}+\boldsymbol{y}\|_2^2 \\ &= |\lambda|^2\langle \boldsymbol{x},\boldsymbol{x}\rangle + \lambda\langle \boldsymbol{x},\boldsymbol{y}\rangle + \bar{\lambda}\langle \boldsymbol{y},\boldsymbol{x}\rangle + \langle \boldsymbol{y},\boldsymbol{y}\rangle \\ &= \begin{bmatrix}\lambda \\ 1\end{bmatrix}^\star \begin{bmatrix}\langle \boldsymbol{x},\boldsymbol{x}\rangle & \langle \boldsymbol{y},\boldsymbol{x}\rangle \\ \langle \boldsymbol{x},\boldsymbol{y}\rangle & \langle \boldsymbol{y},\boldsymbol{y}\rangle\end{bmatrix} \begin{bmatrix}\lambda \\ 1\end{bmatrix} \\ &\geq 0, \quad \forall\lambda\in\mathbb{C}. \end{aligned}$$

其中，矩阵

$$\boldsymbol{M} = \begin{bmatrix}\langle \boldsymbol{x},\boldsymbol{x}\rangle & \langle \boldsymbol{x},\boldsymbol{y}\rangle \\ \langle \boldsymbol{y},\boldsymbol{x}\rangle & \langle \boldsymbol{y},\boldsymbol{y}\rangle\end{bmatrix}$$

是一个 Hermite 矩阵。我们可以看到，$\boldsymbol{M}$ 的两个一阶主子式非负，而它的行列式正是 Cauchy 不等式。我们只需证明 $\det\boldsymbol{M}\geq 0$. 等价地，我们可以通过证明 $\boldsymbol{M}$ 半正定来证明 Cauchy-Schwarz

不等式。由于原式的向量组合并没有覆盖 $\mathbb{C}^2$ 上的所有向量，并不足以说明 $\boldsymbol{M}$ 半正定，我们需要重新构造。

考虑

$$\begin{bmatrix}\alpha\\ \beta\end{bmatrix}^\star \boldsymbol{M} \begin{bmatrix}\alpha\\ \beta\end{bmatrix} = (\boldsymbol{x}\alpha+\boldsymbol{y}\beta)^\star(\boldsymbol{x}\alpha+\boldsymbol{y}\beta) = \|\boldsymbol{x}\alpha+\boldsymbol{y}\beta\|_2^2 \geq 0, \quad \forall\alpha,\beta\in\mathbb{C}.$$

因此，$\boldsymbol{M}$ 是半正定的，从而 Cauchy-Schwarz 不等式得证。当 $\boldsymbol{M}$ 有零特征值时，$\det\boldsymbol{M}=0$，因此不等式取得等号。此时 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 线性相关。 □

第二种证明方法充分利用了几何直观。在 Euclidean 空间中，两个向量的内积等于一个向量的投影向量与另一个向量的内积。顺着这样的思路，我们也可以得到 Cauchy-Schwarz 不等式。

证明. 不妨假设 $\boldsymbol{x}\neq\boldsymbol{0}$, $\boldsymbol{y}\neq\boldsymbol{0}$. 假如我们能够实现分解

$$\boldsymbol{y}=\boldsymbol{y}_1+\boldsymbol{y}_2, \quad \text{其中 } \boldsymbol{y}_1 \parallel \boldsymbol{x}, \boldsymbol{y}_2\perp\boldsymbol{x}.$$

则有

$$\begin{aligned}
|\langle\boldsymbol{x},\boldsymbol{y}\rangle| &= |\langle\boldsymbol{x},\boldsymbol{y}_1+\boldsymbol{y}_2\rangle| \\
&= |\langle\boldsymbol{x},\boldsymbol{y}_1\rangle+\langle\boldsymbol{x},\boldsymbol{y}_2\rangle| \\
&= |\langle\boldsymbol{x},\boldsymbol{y}_1\rangle| \\
&= \left|\|\boldsymbol{x}\|_2\langle\frac{\boldsymbol{x}}{\|\boldsymbol{x}\|_2},\boldsymbol{y}_1\rangle\right| \\
&= \|\boldsymbol{x}\|_2\|\boldsymbol{y}_1\|_2
\end{aligned}$$

下面我们证明 $\|\boldsymbol{y}_1\|_2\leq\|\boldsymbol{y}\|_2$.

$$\|\boldsymbol{y}\|_2^2=\langle\boldsymbol{y}_1+\boldsymbol{y}_2,\boldsymbol{y}_1+\boldsymbol{y}_2\rangle=\|\boldsymbol{y}_1\|_2^2+\|\boldsymbol{y}_2\|_2^2\geq\|\boldsymbol{y}_1\|_2^2.$$

因此，我们有

$$|\langle\boldsymbol{x},\boldsymbol{y}\rangle|\leq\|\boldsymbol{x}\|_2\|\boldsymbol{y}\|_2.$$

等号成立当且仅当 $\boldsymbol{y}_2=\boldsymbol{0}$，即 $\boldsymbol{y}$ 与 $\boldsymbol{x}$ 线性相关。

下面我们说明分解 $\boldsymbol{y}=\boldsymbol{y}_1+\boldsymbol{y}_2$ 的存在性。我们所要做的，就是构造一个投影向量 $\boldsymbol{y}_1$，满足 $y_1\parallel\boldsymbol{x}$，且 $\langle\boldsymbol{x},\boldsymbol{y}-\boldsymbol{y}_1\rangle=0$.

在 Euclidean 空间中，$\boldsymbol{y}$ 在 $\boldsymbol{x}$ 方向上的投影为

$$\boldsymbol{y}_1=\frac{\boldsymbol{x}\cdot\langle\boldsymbol{y},\boldsymbol{x}\rangle}{\|\boldsymbol{x}\|_2^2}.$$

那么，对于一般的内积空间，我们仿照 Euclidean 空间的定义，定义

$$\boldsymbol{y}_1=\frac{\boldsymbol{x}\langle\boldsymbol{y},\boldsymbol{x}\rangle}{\langle\boldsymbol{x},\boldsymbol{x}\rangle}.$$

现在我们已经有 $\boldsymbol{y}_1 \parallel \boldsymbol{x}$，接下来我们只需要验证

$$\langle \boldsymbol{x}, \boldsymbol{y} - \boldsymbol{y}_1 \rangle = 0.$$

计算

$$\begin{aligned}
\langle \boldsymbol{x}, \boldsymbol{y} - \boldsymbol{y}_1 \rangle &= \langle \boldsymbol{x}, \boldsymbol{y} \rangle - \langle \boldsymbol{x}, \boldsymbol{y}_1 \rangle \\
&= \langle \boldsymbol{x}, \boldsymbol{y} \rangle - \left\langle \boldsymbol{x}, \frac{\boldsymbol{x}\langle \boldsymbol{y}, \boldsymbol{x} \rangle}{\|x\|_2^2} \right\rangle \\
&= \langle \boldsymbol{x}, \boldsymbol{y} \rangle - \frac{1}{\|x\|_2^2} \overline{\langle \boldsymbol{x}, \boldsymbol{y} \rangle} \langle \boldsymbol{x}, \boldsymbol{x} \rangle \\
&= \langle \boldsymbol{x}, \boldsymbol{y} \rangle - \langle \boldsymbol{x}, \boldsymbol{y} \rangle \\
&= 0.
\end{aligned}$$

因此分解 $\boldsymbol{y} = \boldsymbol{y}_1 + \boldsymbol{y}_2$ 存在，Cauchy-Schwarz 不等式得证。 □

2.3 常见的内积恒等式

余弦定理 我们之前提到过，夹角可以通过内积来定义。现在我们正式给出其定义。

定义 2.3 (夹角). *设 $\mathbf{0} \neq \boldsymbol{x}, \boldsymbol{y} \in V$. 我们将子空间 $\mathrm{span}\{\boldsymbol{x}\}$ 与 $\mathrm{span}\{\boldsymbol{y}\}$ 之间的夹角定义为*

$$\cos\theta = \frac{|\langle \boldsymbol{x}, \boldsymbol{y} \rangle|}{\|\boldsymbol{x}\|_2 \|\boldsymbol{y}\|_2}, \quad 其中\ \theta \in [0, \frac{\pi}{2}].$$

Cauchy-Schwarz 不等式保证了 $\cos\theta \in [0,1]$，因此夹角是良定义的。

平面几何中有一个著名的定理叫做余弦定理。我们可以将其推广到一般的内积空间中。

定理 2.4 (余弦定理). *设 $\mathbf{0} \neq \boldsymbol{x}, \boldsymbol{y} \in V$，且 $\langle \boldsymbol{x}, \boldsymbol{y} \rangle \in \mathbb{R}, \boldsymbol{x} \neq \boldsymbol{y}$，则有*

$$\|\boldsymbol{x} - \boldsymbol{y}\|_2^2 = \|\boldsymbol{x}\|_2^2 + \|\boldsymbol{y}\|_2^2 - 2\|\boldsymbol{x}\|_2 \|\boldsymbol{y}\|_2 \cos\theta.$$

证明. 显然。将左边展开即可。 □

平行四边形恒等式 之前我们已经看到，内积 $\langle \cdot, \cdot \rangle$ 可以诱导出一个范数 $\|v\| = \sqrt{\langle v, v \rangle}$. 那么，反过来，是否所有范数都能由内积诱导出来呢？答案是否定的！那么，我们该如何判断一个范数是否是由内积诱导出来的呢？我们需要用到平行四边形恒等式。

让我们先回到我们最熟悉的几何世界——Euclidean 空间。在 Euclidean 空间中，两个不共线的向量可以看作平行四边形的两条邻边。我们知道，平行四边形的四条边和两条对角线满足一个简单而优美的关系： **两条对角线的平方和等于等于四条边的平方和**。用数学语言来表达，就是

$$\|\boldsymbol{x} + \boldsymbol{y}\|_2^2 + \|\boldsymbol{x} - \boldsymbol{y}\|_2^2 = 2\|\boldsymbol{x}\|_2^2 + 2\|\boldsymbol{y}\|_2^2.$$

那么，这个等式在一般的内积空间中是否也成立呢？

考虑

$$\begin{aligned}\|\boldsymbol{x}+\boldsymbol{y}\|_2^2+\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 &= \langle \boldsymbol{x}+\boldsymbol{y}, \boldsymbol{x}+\boldsymbol{y}\rangle + \langle \boldsymbol{x}-\boldsymbol{y}, \boldsymbol{x}-\boldsymbol{y}\rangle \\ &= 2\|\boldsymbol{x}\|_2^2+2\|\boldsymbol{y}\|_2^2+\langle \boldsymbol{x},\boldsymbol{y}\rangle+\langle \boldsymbol{y},\boldsymbol{x}\rangle-\langle \boldsymbol{x},\boldsymbol{y}\rangle-\langle \boldsymbol{y},\boldsymbol{x}\rangle \\ &= 2\|\boldsymbol{x}\|_2^2+2\|\boldsymbol{y}\|_2^2.\end{aligned}$$

因此，平行四边形恒等式在任意内积空间中都成立。于是，我们证明了：内积诱导的范数必然满足平行四边形恒等式。

反过来，满足平行四边形恒等式的范数是否一定是由内积诱导的呢？我们说，**是的**。范数是内积诱导的充要条件是它满足平行四边形恒等式。那么，如何证明充分性呢？我们先卖个关子。

极化恒等式 平行四边形恒等式告诉我们一个范数何时“可能”来自内积。那么，如果某个范数满足这个条件，我们该如何从范数反解出它所对应的内积呢？答案是极化恒等式。它告诉我们：内积完全由它诱导的范数决定。

我们有实数和复数两种情况的极化恒等式。

(1) 对于 $\langle \boldsymbol{x},\boldsymbol{y}\rangle\in\mathbb{R}$，根据平行四边形恒等式，我们有

$$\|\boldsymbol{x}+\boldsymbol{y}\|_2^2+\|\boldsymbol{x}-\boldsymbol{y}\|_2^2=2\|\boldsymbol{x}\|_2^2+2\|\boldsymbol{y}\|_2^2.$$

写到这里，我们似乎还没看出来内积要怎么定义。我们不妨换个思路，先看看由内积诱导的范数可以怎样表示内积，然后我们仿照它，去定义一套内积。可以观察到

$$\begin{aligned}\|\boldsymbol{x}+\boldsymbol{y}\|_2^2&=\|\boldsymbol{x}\|_2^2+\|\boldsymbol{y}\|_2^2+\langle \boldsymbol{x},\boldsymbol{y}\rangle+\langle \boldsymbol{y},\boldsymbol{x}\rangle\\ \|\boldsymbol{x}-\boldsymbol{y}\|_2^2&=\|\boldsymbol{x}\|_2^2+\|\boldsymbol{y}\|_2^2-\langle \boldsymbol{x},\boldsymbol{y}\rangle-\langle \boldsymbol{y},\boldsymbol{x}\rangle\end{aligned}$$

对上面两个式子进行线性组合，我们可以得到

$$\langle \boldsymbol{x},\boldsymbol{y}\rangle=\frac{1}{4}\left(\|\boldsymbol{x}+\boldsymbol{y}\|_2^2-\|\boldsymbol{x}-\boldsymbol{y}\|_2^2\right).$$

于是，对于满足平行四边形恒等式的范数，我们依葫芦画瓢，也定义

$$\langle \boldsymbol{x},\boldsymbol{y}\rangle=\frac{1}{4}\left(\|\boldsymbol{x}+\boldsymbol{y}\|_2^2-\|\boldsymbol{x}-\boldsymbol{y}\|_2^2\right).$$

现在我们只需要验证这个定义的内积满足内积的四条公理即可。

(a) 非负性和正定性显然成立。

(b) 可加性：

$$\langle \boldsymbol{x}+\boldsymbol{y},\boldsymbol{z}\rangle=\frac{1}{4}\left(\|\boldsymbol{x}+\boldsymbol{y}+\boldsymbol{z}\|_2^2-\|\boldsymbol{x}+\boldsymbol{y}-\boldsymbol{z}\|_2^2\right)$$

令 $\boldsymbol{u}=\boldsymbol{x}+\boldsymbol{z},\boldsymbol{v}=\boldsymbol{y}+\boldsymbol{z}$，由平行四边形恒等式有

$$\|\boldsymbol{u}+\boldsymbol{v}\|_2^2+\|\boldsymbol{u}-\boldsymbol{v}\|_2^2=2(\|\boldsymbol{u}\|_2^2+\|\boldsymbol{v}\|_2^2).$$

亦即

$$\|\boldsymbol{x}+\boldsymbol{y}+2\boldsymbol{z}\|_2^2+\|\boldsymbol{x}-\boldsymbol{y}\|_2^2=2(\|\boldsymbol{x}+\boldsymbol{z}\|_2^2+\|\boldsymbol{y}+\boldsymbol{z}\|_2^2).$$

类似地，令 $\boldsymbol{u}=\boldsymbol{x}-\boldsymbol{z},\boldsymbol{v}=\boldsymbol{y}-\boldsymbol{z}$，有

$$\|\boldsymbol{x}+\boldsymbol{y}-2\boldsymbol{z}\|_2^2+\|\boldsymbol{x}-\boldsymbol{y}\|_2^2=2(\|\boldsymbol{x}-\boldsymbol{z}\|_2^2+\|\boldsymbol{y}-\boldsymbol{z}\|_2^2).$$

两式相减，有

$$\begin{aligned}\|\boldsymbol{x}+\boldsymbol{y}+2\boldsymbol{z}\|_2^2-\|\boldsymbol{x}+\boldsymbol{y}-2\boldsymbol{z}\|_2^2&=2(\|\boldsymbol{x}+\boldsymbol{z}\|_2^2-\|\boldsymbol{y}+\boldsymbol{z}\|_2^2)+2(\|\boldsymbol{x}-\boldsymbol{z}\|_2^2-\|\boldsymbol{y}-\boldsymbol{z}\|_2^2)\\&=2(\|\boldsymbol{x}+\boldsymbol{z}\|_2^2-\|\boldsymbol{x}-\boldsymbol{z}\|_2^2+\|\boldsymbol{y}+\boldsymbol{z}\|_2^2-\|\boldsymbol{y}-\boldsymbol{z}\|_2^2).\end{aligned}$$

对比我们要证的结论

$$\|\boldsymbol{x}+\boldsymbol{y}+\boldsymbol{z}\|_2^2-\|\boldsymbol{x}+\boldsymbol{y}-\boldsymbol{z}\|_2^2=\|\boldsymbol{x}+\boldsymbol{z}\|_2^2-\|\boldsymbol{x}-\boldsymbol{z}\|_2^2+\|\boldsymbol{y}+\boldsymbol{z}\|_2^2-\|\boldsymbol{y}-\boldsymbol{z}\|_2^2,$$

我们只需证明

$$\|\boldsymbol{x}+\boldsymbol{y}+2\boldsymbol{z}\|_2^2-\|\boldsymbol{x}+\boldsymbol{y}-2\boldsymbol{z}\|_2^2=2(\|\boldsymbol{x}+\boldsymbol{y}+\boldsymbol{z}\|_2^2-\|\boldsymbol{x}+\boldsymbol{y}-\boldsymbol{z}\|_2^2).$$

注意到

$$\begin{aligned}\|(\boldsymbol{x}+\boldsymbol{y}+\boldsymbol{z})+\boldsymbol{z}\|_2^2+\|(\boldsymbol{x}+\boldsymbol{y}+\boldsymbol{z})-\boldsymbol{z}\|_2^2&=2(\|\boldsymbol{x}+\boldsymbol{y}+\boldsymbol{z}\|_2^2+\|\boldsymbol{z}\|_2^2)\\\|(\boldsymbol{x}+\boldsymbol{y}-\boldsymbol{z})+\boldsymbol{z}\|_2^2+\|(\boldsymbol{x}+\boldsymbol{y}-\boldsymbol{z})-\boldsymbol{z}\|_2^2&=2(\|\boldsymbol{x}+\boldsymbol{y}-\boldsymbol{z}\|_2^2+\|\boldsymbol{z}\|_2^2),\end{aligned}$$

两式相减，有

$$\|\boldsymbol{x}+\boldsymbol{y}+2\boldsymbol{z}\|_2^2-\|\boldsymbol{x}+\boldsymbol{y}-2\boldsymbol{z}\|_2^2=2(\|\boldsymbol{x}+\boldsymbol{y}+\boldsymbol{z}\|_2^2-\|\boldsymbol{x}+\boldsymbol{y}-\boldsymbol{z}\|_2^2).$$

因此，可加性得证。

(c) 齐次性：显然。

(d) Hermitian 对称性：由于 $\langle\cdot,\cdot\rangle\in\mathbb{R}$，Hermitian 对称性退化为对称性，结论是显然的。

(2) 那么，对于 $\langle\boldsymbol{x},\boldsymbol{y}\rangle\in\mathbb{C}$ 的情况，我们又能得到什么恒等式呢？可以观察到

$$\begin{aligned}\|\boldsymbol{x}+\boldsymbol{y}\|_2^2&=\|\boldsymbol{x}\|_2^2+\|\boldsymbol{y}\|_2^2+\langle\boldsymbol{x},\boldsymbol{y}\rangle+\langle\boldsymbol{y},\boldsymbol{x}\rangle\\\|\boldsymbol{x}-\boldsymbol{y}\|_2^2&=\|\boldsymbol{x}\|_2^2+\|\boldsymbol{y}\|_2^2-\langle\boldsymbol{x},\boldsymbol{y}\rangle-\langle\boldsymbol{y},\boldsymbol{x}\rangle\\\|\boldsymbol{x}+\mathrm{i}\boldsymbol{y}\|_2^2&=\|\boldsymbol{x}\|_2^2+\|\boldsymbol{y}\|_2^2+\mathrm{i}\langle\boldsymbol{x},\boldsymbol{y}\rangle-\mathrm{i}\langle\boldsymbol{y},\boldsymbol{x}\rangle\\\|\boldsymbol{x}-\mathrm{i}\boldsymbol{y}\|_2^2&=\|\boldsymbol{x}\|_2^2+\|\boldsymbol{y}\|_2^2-\mathrm{i}\langle\boldsymbol{x},\boldsymbol{y}\rangle+\mathrm{i}\langle\boldsymbol{y},\boldsymbol{x}\rangle.\end{aligned}$$

对上面四个式子进行线性组合，我们可以得到

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \frac{1}{4}\left(\|\boldsymbol{x}+\boldsymbol{y}\|_2^2 - \|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + \mathrm{i}\|\boldsymbol{x}-\mathrm{i}\boldsymbol{y}\|_2^2 - \mathrm{i}\|\boldsymbol{x}+\mathrm{i}\boldsymbol{y}\|_2^2\right).$$

于是，对于满足平行四边形恒等式的范数，我们依葫芦画瓢，也定义

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \frac{1}{4}\left(\|\boldsymbol{x}+\boldsymbol{y}\|_2^2 - \|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + \mathrm{i}\|\boldsymbol{x}-\mathrm{i}\boldsymbol{y}\|_2^2 - \mathrm{i}\|\boldsymbol{x}+\mathrm{i}\boldsymbol{y}\|_2^2\right).$$

按理来说，这里我要验证这个定义的内积满足内积的四条公理。不过，我比较懒，就只证一个 Hermitian 对称性，**其余部分留给读者作为练习**。

(d) Hermitian 对称性：

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \frac{1}{4}(\|\boldsymbol{x}+\boldsymbol{y}\|_2^2 - \|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + \mathrm{i}\|\boldsymbol{x}-\mathrm{i}\boldsymbol{y}\|_2^2 - \mathrm{i}\|\boldsymbol{x}+\mathrm{i}\boldsymbol{y}\|_2^2)$$
$$\langle \boldsymbol{y}, \boldsymbol{x} \rangle = \frac{1}{4}(\|\boldsymbol{y}+\boldsymbol{x}\|_2^2 - \|\boldsymbol{y}-\boldsymbol{x}\|_2^2 + \mathrm{i}\|\boldsymbol{y}-\mathrm{i}\boldsymbol{x}\|_2^2 - \mathrm{i}\|\boldsymbol{y}+\mathrm{i}\boldsymbol{x}\|_2^2)$$

我们只需证明

$$\|\boldsymbol{x}+\mathrm{i}\boldsymbol{y}\|_2^2 - \|\boldsymbol{x}-\mathrm{i}\boldsymbol{y}\|_2^2 = \|\boldsymbol{y}-\mathrm{i}\boldsymbol{x}\|_2^2 - \|\boldsymbol{y}+\mathrm{i}\boldsymbol{x}\|_2^2.$$

只需证

$$\|\boldsymbol{x}+\mathrm{i}\boldsymbol{y}\|_2^2 + \|\boldsymbol{y}+\mathrm{i}\boldsymbol{x}\|_2^2 = \|\boldsymbol{x}-\mathrm{i}\boldsymbol{y}\|_2^2 + \|\boldsymbol{y}-\mathrm{i}\boldsymbol{x}\|_2^2.$$

由范数的齐次性，我们有

$$\|\boldsymbol{y}+\mathrm{i}\boldsymbol{x}\|_2^2 = \|\boldsymbol{x}-\mathrm{i}\boldsymbol{y}\|_2^2, \quad \|\boldsymbol{y}-\mathrm{i}\boldsymbol{x}\|_2^2 = \|\boldsymbol{x}+\mathrm{i}\boldsymbol{y}\|_2^2.$$

因此原式成立。

回到我们之前遗留的问题：满足平行四边形恒等式的范数是否一定是由内积诱导的呢？我们现在已经能够给出答案了。对于满足平行四边形恒等式的范数，我们可以通过极化恒等式构造一个内积。由于我们是从范数出发构造内积的，因此这个内积必然诱导出原来的范数。至此，我们完成了证明。因此，我们可以利用平行四边形恒等式来判断一个范数是否是由内积诱导出来的。

2.4 正交补空间

定义 2.5 (正交补空间). *若 $\mathcal{W}$ 是内积空间 $\mathcal{V}$ 的子空间，则*

$$\mathcal{W}^{\perp} = \{\boldsymbol{x} \in \mathcal{V} : \langle x, w \rangle = 0, \forall y \in \mathcal{W}\},$$

也是 $\mathcal{V}$ 的子空间，称为 $\mathcal{W}$ 在 $\mathcal{V}$ 中的正交补空间。

有关正交补空间，有一个非常重要的定理。

定理 2.6. 设 $\boldsymbol{A}$ 是 $\mathbb{F}$ 上的 $m \times n$ 矩阵，则有

$$\operatorname{Ker}(\boldsymbol{A}) = \operatorname{Range}(\boldsymbol{A}^{\star})^{\perp}.$$

这个结论虽然很显然，但它非常重要。它告诉我们，矩阵的零空间是其共轭转置矩阵的列空间的正交补。

3 正交基

3.1 正交基与 QR 分解

我们知道，有限维线性空间是有基的。接下来，我们将目光聚焦到有限维的内积空间上，看看在特定基下，内积的表达式会变成什么样。

引理 3.1. 设 $\mathcal{V}$ 是数域 $\mathbb{F}$ 上的有限维内积空间。若 $\mathcal{V}$ 上的向量组 $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\}$ 满足正交归一性，即

$$\langle \boldsymbol{e}_i, \boldsymbol{e}_j \rangle = \delta_{i,j} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases},$$

则这个向量组线性无关。

证明. 设有

$$\boldsymbol{e}_1 a_1 + \boldsymbol{e}_2 a_2 + \cdots + \boldsymbol{e}_n a_n = \boldsymbol{0}.$$

两边与 $\boldsymbol{e}_j$ 取内积，有

$$\langle \boldsymbol{e}_1 a_1 + \boldsymbol{e}_2 a_2 + \cdots + \boldsymbol{e}_n a_n, \boldsymbol{e}_j \rangle = \langle \boldsymbol{0}, \boldsymbol{e}_j \rangle = 0.$$

而左边要等于 a_j，这说明 $a_j = 0$. 因为 j 是任意的，所以所有系数均为零，向量组线性无关。 □

定义 3.2 (正交基)**.** 设 $\mathcal{V}$ 是数域 $\mathbb{F}$ 上的有限维内积空间。若 $\mathcal{V}$ 上的一组基满足正交归一性，则称其为 $\mathcal{V}$ 的一组正交基。

接下来我们来看看内积在正交基与非正交基下的表达式。我们先求出在非正交基下的内积表达式。

考虑基 $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\}$，它们拼在一起构成矩阵 $\boldsymbol{E} = [\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n]$，则任意向量 $\boldsymbol{x}, \boldsymbol{y}$ 可以表示为

$$\begin{aligned} \boldsymbol{x} &= \boldsymbol{E}\boldsymbol{\alpha} \\ \boldsymbol{y} &= \boldsymbol{E}\boldsymbol{\beta}, \end{aligned}$$

其中 α, β 是坐标向量。现在，我们对 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 取内积，有

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{E}\boldsymbol{\alpha}, \boldsymbol{E}\boldsymbol{\beta} \rangle = \boldsymbol{\alpha}^{\star} \langle \boldsymbol{E}, \boldsymbol{E} \rangle \boldsymbol{\beta},$$

其中

$$\langle \boldsymbol{E}, \boldsymbol{E} \rangle = \begin{bmatrix} \langle \boldsymbol{e}_1, \boldsymbol{e}_1 \rangle & \langle \boldsymbol{e}_2, \boldsymbol{e}_1 \rangle & \cdots & \langle \boldsymbol{e}_n, \boldsymbol{e}_1 \rangle \\ \langle \boldsymbol{e}_1, \boldsymbol{e}_2 \rangle & \langle \boldsymbol{e}_2, \boldsymbol{e}_2 \rangle & \cdots & \langle \boldsymbol{e}_n, \boldsymbol{e}_2 \rangle \\ \vdots & \vdots & & \vdots \\ \langle \boldsymbol{e}_1, \boldsymbol{e}_n \rangle & \langle \boldsymbol{e}_2, \boldsymbol{e}_n \rangle & \cdots & \langle \boldsymbol{e}_n, \boldsymbol{e}_n \rangle \end{bmatrix},$$

显然根据内积的定义，它一定会是一个 Hermite 正定矩阵。我们把这个矩阵叫做 Gram 矩阵，也叫做度量矩阵，它的 (i,j) 元素为 $\langle \boldsymbol{e}_j, \boldsymbol{e}_i \rangle$.

假如 $\boldsymbol{E}$ 是正交基拼成的矩阵，那么 Gram 矩阵就是单位阵 $\boldsymbol{I}$，此时内积的表达式就大大简化了：

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{\alpha}^\star \boldsymbol{I} \boldsymbol{\beta} = \boldsymbol{\alpha}^\star \boldsymbol{\beta}.$$

那么假如 $\boldsymbol{E}$ 不是正交基拼成的矩阵，我们能不能通过一定的坐标变换，把它变成正交基呢？

由于 $\langle \boldsymbol{E}, \boldsymbol{E} \rangle$ 是正定阵，我们可以对它做 Cholesky 分解，得到

$$\langle \boldsymbol{E}, \boldsymbol{E} \rangle = \boldsymbol{R}^\star \boldsymbol{R},$$

其中 $\boldsymbol{R}$ 是上三角矩阵。那么，内积就变成了

$$\begin{aligned} \boldsymbol{\alpha}^\star \langle \boldsymbol{E}, \boldsymbol{E} \rangle \boldsymbol{\beta} &= \boldsymbol{\alpha}^\star \boldsymbol{R}^\star \boldsymbol{R} \boldsymbol{b} eta \\ &= (\boldsymbol{R}\boldsymbol{\alpha})^\star (\boldsymbol{R}\boldsymbol{\beta}) \end{aligned}$$

于是，我们定义新的坐标向量

$$\tilde{\boldsymbol{\alpha}} = \boldsymbol{R}\boldsymbol{\alpha},$$
$$\tilde{\boldsymbol{\beta}} = \boldsymbol{R}\boldsymbol{\beta}.$$

则内积变成了

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \tilde{\boldsymbol{\alpha}}^\star \tilde{\boldsymbol{\beta}}.$$

这说明，在新的坐标系下，基 $\{\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n\}$ 是正交归一的，其中

$$\begin{bmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 & \cdots & \boldsymbol{q}_n \end{bmatrix} = \boldsymbol{E}\boldsymbol{R}^{-1}.$$

将 $\boldsymbol{R}$ 写到新基的右边，就有

$$\boldsymbol{E} = \begin{bmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 & \cdots & \boldsymbol{q}_n \end{bmatrix} \boldsymbol{R} = \boldsymbol{Q}\boldsymbol{R},$$

其中 $\boldsymbol{Q}$ 是正交归一基拼成的矩阵。这个分解叫做 QR 分解。QR 分解的意义就在于：它告诉我们，任意基都可以正交化，从而得到简洁的内积表达式。这个分解的过程其实就是 Gram-Schmidt 正交化的过程。注意到

$$[\boldsymbol{e}_1, \cdots, \boldsymbol{e}_n] = [\boldsymbol{q}_1, \cdots, \boldsymbol{q}_n] \begin{bmatrix} r_{1,1} & r_{1,2} & \cdots & r_{1,n} \\ & r_{2,2} & \cdots & r_{2,n} \\ & & \ddots & \vdots \\ & & & r_{n,n} \end{bmatrix}$$

我们写成列向量的形式：

$$\begin{aligned}
\boldsymbol{e}_1 &= \boldsymbol{q}_1 r_{1,1}, \\
\boldsymbol{e}_2 &= \boldsymbol{q}_1 r_{1,2} + \boldsymbol{q}_2 r_{2,2}, \\
&\vdots \\
\boldsymbol{e}_n &= \boldsymbol{q}_1 r_{1,n} + \boldsymbol{q}_2 r_{2,n} + \cdots + \boldsymbol{q}_n r_{n,n}.
\end{aligned}$$

上面我们是用 $\boldsymbol{q}_i$ 来表示 $\boldsymbol{e}_i$；反过来，假如我们用 $\boldsymbol{e}_i$ 表示 $\boldsymbol{q}_i$，我们有

$$\begin{aligned}
& r_{1,1} = \|\boldsymbol{e}_1\|_2, \quad \boldsymbol{q}_1 = \frac{\boldsymbol{e}_1}{r_{1,1}}, \\
& \hat{\boldsymbol{q}}_2 = \boldsymbol{e}_2 - \langle \boldsymbol{q}_1, \boldsymbol{e}_2 \rangle \boldsymbol{q}_1, \quad r_{2,2} = \|\hat{\boldsymbol{q}}_2\|_2, \quad \boldsymbol{q}_2 = \frac{\hat{\boldsymbol{q}}_2}{r_{2,2}}, \\
& \quad \vdots \\
& \hat{\boldsymbol{q}}_n = \boldsymbol{e}_n - \sum_{j=1}^{n-1} \langle \boldsymbol{q}_j, \boldsymbol{e}_n \rangle \boldsymbol{q}_j, \quad r_{n,n} = \|\hat{\boldsymbol{q}}_n\|_2, \quad \boldsymbol{q}_n = \frac{\hat{\boldsymbol{q}}_n}{r_{n,n}}.
\end{aligned}$$

这就是 Classical Gram-Schmidt 正交化 (CGS)。还有一种 Modified Gram-Schmidt (MGS)，邵老师在数值算法课上讲过了，在这里就不再赘述。

定理 3.3 (QR 分解)**.** *若 $\boldsymbol{A} \in \mathbb{C}^{m\times n}$ 列满秩，则存在酉矩阵 $\boldsymbol{Q} \in \mathbb{C}^{m\times m}$ 和对角元为正实数的上三角矩阵 $\boldsymbol{R} \in \mathbb{C}^{m\times n}$，使得 $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{R}$. 并且当 $\boldsymbol{A}$ 是实矩阵时，$\boldsymbol{Q}$ 和 $\boldsymbol{R}$ 都可以取实矩阵。*

若 $\boldsymbol{A} \in \mathbb{C}^{m\times n}$ 列满秩，则存在带有标准正交列的矩阵 $\boldsymbol{Q} \in \mathbb{C}^{m\times n}$ 和对角元为正实数的上三角矩阵 $\boldsymbol{R} \in \mathbb{C}^{n\times n}$，使得 $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{R}$，并且该分解是唯一的。当 $\boldsymbol{A}$ 是实矩阵时，$\boldsymbol{Q}$ 和 $\boldsymbol{R}$ 都是实矩阵。

证明. 存在性我们在上文中已经利用内积的性质证明过了。下面我们证明唯一性。

由于 $\boldsymbol{Q} = \boldsymbol{E}\boldsymbol{R}^{-1}$，我们只需证明 $\boldsymbol{R}$ 唯一。而 $\boldsymbol{R}$ 是来自 $\langle \boldsymbol{E}, \boldsymbol{E} \rangle$ 的 Cholesky 分解。由 Cholesky 分解的唯一性可知，$\boldsymbol{R}$ 唯一。因此，QR 分解唯一。

若 $\boldsymbol{Q} \in \mathbb{C}^{m\times m}$，则 $\boldsymbol{R}$ 的第 $n+1$ 行到第 m 行均为零，因此我们可以将 $\boldsymbol{Q}$ 精简为它的前 n 列，将 $\boldsymbol{R}$ 精简为它的前 n 行，所有结论仍然是成立的。 □

那么，对于亏秩的矩阵 $\boldsymbol{A}$ 呢？它的 QR 分解是否存在且唯一呢？如果我们还要求对角元为正实数的话，恐怕是连存在都无法存在的。不过，假如我们把这个条件放宽，对角元可以是非负实数，那么 QR 分解仍然是存在的。

对于亏秩矩阵的 QR 分解，我们将给出三种方法对上面的结论进行证明：Gram-Schmidt 正交化、Householder 变换、扰动法。

GS 我们不妨设 $\mathrm{rank}(\boldsymbol{A}) = r$. 如果我们按 Gram-Schmidt 正交化的过程去做，我们会遇到以下几种情况：

1. 第一列是零：此时我们可以向后找一个非零列与其交换，然后继续正交化。
2. 中间某一列与前面的列线性相关：向后找一个与前面的列线性无关的向量，与其交换位置，然后继续正交化。如果找不到，则说明剩下的列都与前面的列线性相关，正交基已经找完，我们只需配出 $\boldsymbol{R}$ 剩下的系数即可。

通过列选主元的方式，我们可以得到亏秩矩阵的 QR 分解 $\boldsymbol{A} = \boldsymbol{Q}R$，其中 $\boldsymbol{Q} \in \mathbb{C}^{m\times r}$ 列标准正交，而 $\boldsymbol{R} = [\boldsymbol{R}_1, \boldsymbol{R}_2] \in \mathbb{C}^{r\times n}$，其中 $\boldsymbol{R}_1 \in \mathbb{C}^{r\times r}$ 是对角元为正实数的上三角矩阵，而 $\boldsymbol{R}_2 \in \mathbb{C}^{r\times(n-r)}$.

Householder 以防有人没上过数值算法，这里简单介绍一下 Householder 变换。在前面的学习中我们使用 Gram-Schmidt 正交化来计算 QR 分解。这种方法直观易懂，但数值稳定性较差。当矩阵的列向量接近线性相关时，Gram-Schmidt 方法会产生较大的舍入误差。为了解决这个问题，我们可以使用 Householder 变换来计算 QR 分解。Householder 变换的核心想法是一次性将多个分量归零。

让我们从几何中寻找灵感。假设我们有一个向量 $\boldsymbol{x} \in \mathbb{R}^3$，我们希望将其关于某个镜面进行反射，得到另一个向量 $\boldsymbol{y} \in \mathbb{R}^3$，它们具有相同的长度，但方向不同。一个关键的观察是：$\boldsymbol{x}, \boldsymbol{y}$ 在镜面上的投影是相同的，在垂直镜面的方向上，它们的分量大小相等方向相反。因此，给定一个平面，我们需要构造一个线性变换，它满足：

1. 沿着平面的方向，向量保持不变；
2. 垂直于平面的方向，向量的分量取反。

这种变换就叫做 Householder 变换。

现在，假如我们需要一个 Householder 变换，将向量 $\boldsymbol{x}$ 反射到向量 $\boldsymbol{y}$，我们该如何构造这个变换呢？可以注意到：$\boldsymbol{y} - \boldsymbol{x}$ 是垂直于反射面的向量。如果我们将其归一化，设

$$\boldsymbol{w} = \frac{\boldsymbol{y} - \boldsymbol{x}}{\|\boldsymbol{y} - \boldsymbol{x}\|_2},$$

垂直于反射面方向的向量构成一个子空间，这个子空间的投影矩阵为 $\boldsymbol{w}\boldsymbol{w}^{\mathrm{T}}$，于是反射面的投影矩阵就是 $\boldsymbol{I} - \boldsymbol{w}\boldsymbol{w}^{\mathrm{T}}$.

根据 Householder 变换的定义，沿着反射面的方向，向量保持不变，而垂直于反射面的方向，向量的分量取反。因此，Householder 变换可以表示为

$$\boldsymbol{H} = (\boldsymbol{I} - \boldsymbol{w}\boldsymbol{w}^{\mathrm{T}}) - \boldsymbol{w}\boldsymbol{w}^{\mathrm{T}} = \boldsymbol{I} - 2\boldsymbol{w}\boldsymbol{w}^{\mathrm{T}}.$$

那么，$\boldsymbol{H}$ 有着怎样的性质呢？

(1) 对称性：显然。

(2) 正交性：一个向量关于某个平面反射两次，等于没有反射，因此 $\boldsymbol{H}^2 = I$，亦即 $\boldsymbol{H}\boldsymbol{H}^{\mathrm{T}} = \boldsymbol{H}^{\mathrm{T}}\boldsymbol{H} = \boldsymbol{I}$.

有了 Householder 变换的定义，我们就可以用它来计算矩阵的 QR 分解了：对于列满秩的矩阵 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$，

1. 将 $\boldsymbol{A}[1:m,1]$ 反射为 $\|\boldsymbol{A}[1:m,1]\|_2 \cdot \boldsymbol{e}_1 \in \mathbb{R}^m$；
2. 将 $\boldsymbol{A}[2:m,2]$ 反射为 $\|\boldsymbol{A}[2:m,2]\|_2 \cdot \boldsymbol{e}_1 \in \mathbb{R}^{m-1}$；
3. 依此类推，直到将 $\boldsymbol{A}[n:m,n]$ 反射为 $\|\boldsymbol{A}[n:m,n]\|_2 \cdot \boldsymbol{e}_1 \in \mathbb{R}^{m-n+1}$.

于是，我们可以得到

$$\boldsymbol{H}_n \cdots \boldsymbol{H}_2\boldsymbol{H}_1\boldsymbol{A} = \boldsymbol{R},$$

其中 $\boldsymbol{R}$ 是上三角矩阵。我们把所有的 Householder 变换都逆过去，并设 $\boldsymbol{Q} = \boldsymbol{H}_1 \cdots \boldsymbol{H}_n$，有

$$\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{R}.$$

对于亏秩的矩阵，反射进行到某一步时，可能会发现某一列已经全部归零了，这时我们可以进行列交换，继续反射，直到换无可换，最终得到一个行阶梯型的矩阵 $\boldsymbol{R} \in \mathbb{R}^{r\times n}$，其中 r 是矩阵 $\boldsymbol{A}$ 的秩。

扰动法 大体的思路是：我们可以构造一个列满秩矩阵序列 $\{\boldsymbol{A}_k\}_{k\in\mathbb{N}}$，使得 $\boldsymbol{A}_k \to \boldsymbol{A}$. 对于 $\boldsymbol{A}_k$，我们有 QR 分解 $\boldsymbol{A}_k = \boldsymbol{Q}_k\boldsymbol{R}_k$，而 $\boldsymbol{Q}_k$ 可以收敛到某个酉矩阵 $\boldsymbol{Q}$，$\boldsymbol{R}_k$ 可以收敛到某个上三角矩阵 $\boldsymbol{R}$. 于是，我们就得到了 $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{R}$.

现在我们来完善一下细节。第一个问题是：我们该如何构造列满秩矩阵序列 $\{\boldsymbol{A}_k\}_{k\in\mathbb{N}}$？一个简单的方法是对 $\boldsymbol{A}$ 的主对角元加上扰动：$\boldsymbol{A}_k = \boldsymbol{A} + \theta_k[\boldsymbol{I}, \boldsymbol{0}]^{\mathrm{T}}$, $\theta_k \to 0$. 对于每一个 $\boldsymbol{A}_k$，我们都可以计算出它的 QR 分解 $\boldsymbol{A}_k = \boldsymbol{Q}_k\boldsymbol{R}_k$.

在第三讲中，我们有提到过，酉矩阵序列必定有收敛子列。这是因为 $\|\boldsymbol{Q}_k\|_2 = 1$，根据 Bolzano-Weierstrass 定理，酉矩阵序列必定有收敛子列。

对于 $\boldsymbol{R}_k$，由 $\|\boldsymbol{A}_k\|_2 = \|\boldsymbol{R}_k\|_2$ 可知，$\boldsymbol{R}_k$ 也是有界的，因此也有收敛子列。我们取出公共收敛子标，记为 $\{k_j\}_{j\in\mathbb{N}}$，则有

$$\boldsymbol{Q}_{k_j} \to \boldsymbol{Q}, \quad \boldsymbol{R}_{k_j} \to \boldsymbol{R}, \boldsymbol{A}_{k_j} \to \boldsymbol{A}.$$

于是，我们有

$$\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{R}.$$

3.2 案例分析

对于 n 阶多项式空间 $\mathcal{P}_n(\mathbb{F})$，我们可以取出一组基 $\{1, x, \ldots, x^n\}$. 在这个空间中，内积的定义为

$$\langle f(x), g(x)\rangle = \int_0^1 f(x)\overline{g}(x)\ \mathrm{d}x.$$

考虑基与基之间的内积，我们有

$$\langle x^i, x^j\rangle = \int_0^1 x^{i+j}\ \mathrm{d}x = \frac{1}{i+j+1}.$$

因此，这组基的 Gram 矩阵为

$$\boldsymbol{H} = \begin{bmatrix} 1/1 & 1/2 & 1/3 & \cdots & 1/(n+1) \\ 1/2 & 1/3 & 1/4 & \cdots & 1/(n+2) \\ 1/3 & 1/4 & 1/5 & \cdots & 1/(n+3) \\ \vdots & \vdots & \vdots & & \vdots \\ 1/(n+1) & 1/(n+2) & 1/(n+3) & \cdots & 1/(2n+1) \end{bmatrix}.$$

这个矩阵叫做 Hilbert 矩阵。在这组基下，内积的表达式为

$$\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{\beta}.$$

由于 Hilbert 矩阵极其病态，因此基 $\{1, x, \ldots, x^n\}$ 并不是一个好的基。我们可以通过正交化把它变成一个好的基。

4 内积空间上的线性映射

到目前为止，我们主要研究了内积空间本身的结构——长度、角度、正交性、正交基等等。现在，我们要将目光转向在这些空间上操作的线性映射。我们知道，一般线性空间上的线性映射是保持线性运算的，那么，当我们在线性映射中要求“保持内积空间的结构”时，我们能得到什么结论呢？

在内积空间中，我们有了额外的几何结构——内积。我们自然会问：什么样的线性映射能够保持住这个几何结构呢？我们先看看 $\mathbb{R}^2$ 或 $\mathbb{R}^3$ 中的线性变换：

- 有些变换会扭曲几何形状——拉伸、剪切等；
- 有些变换则保持几何形状不变——旋转、反射等。

这些保持几何形状不变的线性变换，它们实际上保持了：

1. 长度不变：$\|\boldsymbol{T}(\boldsymbol{x})\| = \|\boldsymbol{x}\|$；

2. 角度不变：$\angle\langle \boldsymbol{T}(\boldsymbol{x}), \boldsymbol{T}(\boldsymbol{y})\rangle = \angle\langle \boldsymbol{x}, \boldsymbol{y}\rangle$；

3. 正交性不变：若 $\langle \boldsymbol{x}, \boldsymbol{y}\rangle = 0$，则 $\langle \boldsymbol{T}(\boldsymbol{x}), \boldsymbol{T}(\boldsymbol{y})\rangle = 0$.

基于这些共性，我们可以抽象出内积空间上的一种特殊线性映射，称为正交变换。

定义 4.1 (正交变换)**.** *设 $\mathcal{V}$ 是线性空间，线性变换 $\boldsymbol{Q} : \mathcal{V} \to \mathcal{V}$ 若能保持内积，即*

$$\langle \boldsymbol{Q}\boldsymbol{x}, \boldsymbol{Q}\boldsymbol{y}\rangle = \langle \boldsymbol{x}, \boldsymbol{y}\rangle, \quad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{V},$$

则称 $\boldsymbol{Q}$ 为 $\mathcal{V}$ 上的正交变换（实内积空间）或酉变换（复内积空间）。

由定义，我们立即有以下结论：

(1) 保范数：令 $\boldsymbol{y} = \boldsymbol{x}$ 即可。

(2) 保正交性：若 $\langle \boldsymbol{x}, \boldsymbol{y}\rangle = 0$，则由定义可知 $\langle \boldsymbol{Q}\boldsymbol{x}, \boldsymbol{Q}\boldsymbol{y}\rangle = 0$.

(3) 有限维内积空间上的正交变换在标准正交基下的表示矩阵为酉矩阵。

证明. 不妨设表示矩阵为 $\boldsymbol{Q}$，由坐标变换公式，新的坐标为 $\boldsymbol{Q}\boldsymbol{x}$. 由于正交变换保内积，有

$$\langle \boldsymbol{Q}\boldsymbol{x}, \boldsymbol{Q}\boldsymbol{y}\rangle = \langle \boldsymbol{x}, \boldsymbol{y}\rangle.$$

由内积的定义，有

$$\boldsymbol{x}^\star \langle \boldsymbol{Q}, \boldsymbol{Q}\rangle \boldsymbol{y} = \boldsymbol{x}^\star \boldsymbol{y}.$$

因此

$$\langle \boldsymbol{Q}, \boldsymbol{Q}\rangle = \boldsymbol{I},$$

这意味着 $\boldsymbol{Q}$ 是酉矩阵。 □

接下来我们介绍几种特殊的正交变换。

(1) 镜像变换：其实就是 Householder 变换。它的表示矩阵为

$$\boldsymbol{H} = \boldsymbol{I} - 2\boldsymbol{w}\boldsymbol{w}^{\mathrm{T}}.$$

Householder 变换可以一次性将多个分量归零，因此在数值计算中非常有用。

(2) 旋转变换：又称为 Givens 变换。我们知道，$\mathbb{R}^2$ 上的旋转可以表示为

$$G(\theta) = \begin{bmatrix} c & -s \\ s & c \end{bmatrix}, \quad c = \cos\theta, s = \sin\theta.$$

利用 Givens 变换，我们可以对 $\mathbb{R}^2$ 中的向量进行旋转，旋转到与某个坐标轴平行，从而将其中一个分量归零。

但在更高维的空间里，我们就不能一次性得到与坐标轴平行的向量了。高维空间的 Givens 变换一次只能旋转其中的两个分量，它的表示矩阵为

$$G(p,q,\theta)=\begin{bmatrix}1 & & & & & & \\ & \ddots & & & & & \\ & & c & & -s & & \\ & & & \ddots & & & \\ & & s & & c & & \\ & & & & & \ddots & \\ & & & & & & 1\end{bmatrix}$$

其中非对角元分别位于 (p,q) 和 (q,p) 位置。经过一次这样的变换，我们可以将向量的第 p 个分量和第 q 个分量旋转，选取一个合适的 θ，就可以将其中一个分量归零。经过一系列的 Givens 变换后，我们就可以将一个向量变成与坐标轴平行的向量，从而实现分量归零的目的。

镜像变换和旋转变换有着一定的内在联系。我们知道，在二维空间中，一个正交矩阵的行列式要么是 1，要么是 -1. 当行列式为 1 时，其表示矩阵必然可以表示为旋转矩阵（读者可以写出来自行验证）；而当行列式为 -1 时，其表示矩阵则可以表示为镜像变换矩阵。那么，假如我有两个镜像变换 $\boldsymbol{Q}_1,\boldsymbol{Q}_2$，它们的复合 $\boldsymbol{Q}_1\boldsymbol{Q}_2$ 行列式为 1，因此它们的复合将会是一个旋转变换。这一点可以利用几何直观地理解：将一个向量沿两个镜子反射，等价于绕某个转轴旋转某个角度。在更高维的空间中，这个结论仍然是成立的，只不过我们就无法用简单的几何图形来验证了。

现在让我们聚焦三维空间的旋转。我们有一个定理：

定理 4.2 (欧拉角定理)**.** *三维空间中的任意旋转都可以视为若干个绕坐标轴的旋转的复合。*

证明. 证明的核心思路是：我们对任意旋转矩阵 $\boldsymbol{R}$ 进行基于 Givens 旋转的 QR 分解，得到

$$\boldsymbol{G}_3\boldsymbol{G}_2\boldsymbol{G}_1\boldsymbol{R}=\boldsymbol{I}.$$

下面，我们来完善一下细节。

首先，为什么最后可以得到单位阵？这时因为 $\boldsymbol{R}$ 是正交矩阵，每一个列向量都正交，这意味着任何一个列向量都没有在其他列向量方向上的分量，这意味着得到的上三角矩阵一定是对角阵；而旋转变换不改变向量的范数，因此最后得到的上三角矩阵必定是单位阵。

其次，我们要怎么转？我们可以分三步走：

1. 将第一列的第二个分量归零，我们需要旋转矩阵 $\boldsymbol{G}(1,2)$. 这个旋转是在 xy 平面上进行的，因此它相当于绕 z 轴旋转；

2. 将第一列的第三个分量归零，我们需要旋转矩阵 $\boldsymbol{G}(1,3)$. 这个旋转是在 xz 平面上进行的，因此它相当于绕 y 轴旋转。此时，第一列已经变成了 $\boldsymbol{e}_1$，而后两列的第一个分量也已经清零了；

3. 将第二列的第三个分量归零，我们需要旋转矩阵 $\boldsymbol{G}(2,3)$. 这个旋转是在 yz 平面上进行的，因此它相当于绕 x 轴旋转。此时，第二列也变成了 $\boldsymbol{e}_2$，而第三列的前两个分量也已经清零了，因此第三列已经变成了 $\boldsymbol{e}_3$，无需再旋转了。

$\square$