

Lecture 5 线性映射 — 2025.09.30

教授：邵美悦

Scribe: 两人间

1 大纲

1. 线性映射
2. Sherman-Morrison-Woodbury 公式

2 线性映射

线性映射的定义 我们常常遇到从一个向量空间到另一个向量空间的变换。例如，将某个二维向量逆时针旋转 90° ，或者将某个三维向量投影到 xy 平面上。现在我们聚焦一类特殊的变换：线性映射。

定义 2.1 (线性映射). 设 V 和 W 是 \mathbb{K} 上的线性空间，设有映射 $L: V \rightarrow W$ ，若它满足以下两条性质：

1. 保持加法：对任意 $u, v \in V$ ，有 $L(u + v) = L(u) + L(v)$ ；
2. 保持数乘：对任意 $v \in V$ 和 $\alpha \in \mathbb{K}$ ，有 $L(\alpha v) = \alpha L(v)$ ；

则称 L 是从 V 到 W 的线性映射（或线性变换）。

特殊地，若 $V = W$ ，则称 L 是 V 上的线性算子。若 $W = \mathbb{K}$ ，则称 L 是 V 上的线性泛函。

例 2.2. 求导数和求积分是 \mathbb{R} 上的特定函数空间里的线性算子。因为它们都保持加法和数乘。

现在我们来考察一类特殊的求积分线性映射。

例 2.3. 设有函数 $f: \mathbb{R} \rightarrow \mathbb{R}$ ，定义映射 $T: f \rightarrow g$ 为

$$T[f](t) = \int_0^1 k(s, t) f(s) \, ds = g(t).$$

根据定积分的性质，容易验证 T 保持加法和数乘，因此 T 是从所有实值函数到所有实值函数的线性映射。

例 2.4 (Dirac δ 函数). 我们已知

$$L[f] = \int_{-\infty}^{+\infty} k(s)f(s) \, ds$$

是从所有实值函数到实数的线性泛函。是否所有的线性泛函都可以表示成上述形式呢？不然。考虑如下线性泛函：

$$L[f] = f(0),$$

即取函数在 0 处的值。由于 Riemann 积分对于单个点的变动是不敏感的，因此我们没办法通过积分的形式，将某个点的值提取出来。但如果我们想要强行将 $L[f] = f(0)$ 写成积分的形式，且为之奈何？物理学家引入了 Dirac δ 函数的概念，定义

$$\int_{-\infty}^{+\infty} \delta(s)f(s) \, ds = f(0).$$

在这里， $\delta(s)$ 并不是一个真正的函数，而是一个广义函数。事实上，Dirac δ 函数可以看作是一个极限过程：我们想要通过积分提取出 0 处的函数值，那我们就必须让非 0 点处 δ 函数的值为零——否则积分会受到非 0 点处函数值的影响。为了保持 $\int_{-\infty}^{+\infty} \delta(s) \, ds = 1$ ，我们只能让 $\delta(0)$ 的值趋向于无穷大。这就是我们在数学分析中非常熟悉的“单点突起”类型的积分。从另一个角度， δ 函数也可以看成是函数列的极限：

$$\delta(s) = \lim_{n \rightarrow \infty} \delta_n(s), \quad \text{其中 } \delta_n(s) = \begin{cases} n, & |s| < \frac{1}{2n}, \\ 0, & |s| \geq \frac{1}{2n}. \end{cases}$$

如果我们让 $f(x) \equiv 1$ ，则有

$$\int_{-\infty}^{+\infty} \delta(s) \, ds = 1.$$

在微积分中，我们也可以利用积分算子是线性算子，来简化计算。

例 2.5. 计算积分

$$u = \int \frac{\sin x}{3 \sin x + 5 \cos x} \, dx.$$

解. 令

$$v = \int \frac{\cos x}{3 \sin x + 5 \cos x} \, dx.$$

则有

$$3u + 5v = x + C_1$$

$$-5u + 3v = \ln(3 \sin x + 5 \cos x) + C_2.$$

联立解出 u 即可。

这种做法在本质上利用了积分算子是线性算子的性质。通过构造辅助变量，我们将原本复杂的积分问题转化为线性方程组的求解问题，从而简化了计算过程。

我们已经定义了什么是线性映射，它们是作用在向量上的“操作”。一个很自然的问题是：这些“操作”本身能否构成一个集合？这个集合有没有什么有趣的代数结构？

考虑所有从向量空间 V 到 W 的线性映射，我们将其记作 $\mathcal{L}(V, W)$ 。对于两个线性映射 $T, S \in \mathcal{L}(V, W)$ ，我们可以很自然地定义它们的和 $T + S$ ：

$$(T + S)(v) := T(v) + S(v), \quad \forall v \in V.$$

同样，对于一个标量 $c \in \mathbb{K}$ 和一个线性映射 $T \in \mathcal{L}(V, W)$ ，我们可以定义它们的数乘 cT ：

$$(cT)(v) := c \cdot T(v), \quad \forall v \in V.$$

不难验证，这样新定义的 $T + S$ 和 cT 仍然是 $V \rightarrow W$ 的线性映射。

令人惊奇的是，集合 $\mathcal{L}(V, W)$ 在我们刚刚定义的加法和数乘运算下，它本身也构成了一个新的向量空间！（它的零向量是零映射，即把所有向量都映到 0_W 的映射）。这是一个深刻的发现：我们从研究向量空间 V 和 W 出发，通过考察它们之间的线性映射，得到了一个新的向量空间 $\mathcal{L}(V, W)$ 。

在所有可能的目标空间 W 中，最简单、最基础的是什么？就是我们这个理论体系的基石——底层的数域 \mathbb{K} 本身（因为 \mathbb{K} 本身也是一个一维向量空间）。当 $W = \mathbb{K}$ 时，这些特殊的线性映射（即线性泛函）构成的向量空间 $\mathcal{L}(V, \mathbb{K})$ 就显得尤为重要。

定义 2.6 (对偶空间). 给定向量空间 V ，其上的所有线性泛函构成的向量空间称为 V 的**对偶空间** (*dual space*)，记作 V^* 。即

$$V^* := \mathcal{L}(V, \mathbb{K}).$$

对偶空间 V^* 完整地刻画了如何从向量空间 V 中“测量”或“提取”出数值信息。 V 中的元素是向量，而 V^* 中的元素是作用于向量并产生数值的“测量尺”。它们一个是被“测量”的，一个是用来“测量”的，形成了一种“对偶”关系。

我们已经知道，我们可以使用范数来衡量一个向量的“长度”或者“大小”。一个自然的问题是：我们能否也用某种方式来衡量一个线性映射的“大小”或者“强度”呢？

对于一个向量 $v \in V$ ，我们期望衡量线性映射 $T : V \rightarrow W$ 的“大小”，可以考虑这个映射对 v 的大小的改变。具体来说，我们可以定义一个线性映射的范数为

定义 2.7 (算子范数 (operator norm)). 对于线性映射 $T : V \rightarrow W$ ，其算子范数定义为

$$\|T\| := \sup_{\|v\|_V \neq 0} \frac{\|T(v)\|_W}{\|v\|_V} = \sup_{\|v\|_V = 1} \|T(v)\|_W.$$

算子范数 $\|T\|$ 衡量了线性映射 T 在单位球面上的最大伸缩作用。举个例子， \mathbb{R}^2 上的旋转算子 R_θ 只是旋转向量，但并不改变向量的长度，因此 $\|R_\theta\| = 1$ 。而投影算子 P 将向量投影到某个子空间

上，通常会缩短向量的长度；如果运气好些，向量正好落于这个子空间上，就不会改变长度。因此 $\|P\| = 1$ 。

我们在第三讲中定义了矩阵的算子范数。矩阵的算子范数定义是

$$\|A\|_{\alpha,\beta} = \sup_{x \neq 0} \frac{\|Ax\|_{\beta}}{\|x\|_{\alpha}}.$$

可以发现，矩阵的算子范数和线性映射的算子范数有着完全相同的定义。这并不奇怪，因为矩阵本质上就是一种线性映射。

在定义了算子范数之后，我们自然会关心一个映射的有界性。我们已经知道，从 V 到 W 的线性映射全体构成一个向量空间 $\mathcal{L}(V, W)$ ，而我们可以进一步验证：从 V 到 W 的所有有界线性映射构成的集合 $\mathcal{B}(V, W)$ 也是一个向量空间，亦即 $\mathcal{B}(V, W)$ 是 $\mathcal{L}(V, W)$ 的子空间。有关有界线性算子我们就说到这里，更多内容可以参考泛函分析相关教材。

线性映射的表示矩阵 我们从一个实际问题出发：一个线性映射 $T: L \rightarrow V$ 是一个抽象的映射规则。对于任意一个向量 $v \in V$ ，如果我们想要用计算机或者纸笔来具体计算 $T(v)$ 的结果，该怎么办呢？直接操作抽象的向量 v 和映射 T 是不现实的。但如果我们能将所有东西都坐标化，转换成我们熟悉的数字和数组，计算的可行性就大大提升了。

如果我们找出 V 的一组基，即 $V = \text{span}\{e_1, \dots, e_n\}$ ，那么对于任意 $v \in V$ ，我们可以将 v 表示为 $v = \sum_{i=1}^n x_i e_i$ 。这样，我们就可以将 v 与其坐标向量 $(x_1, \dots, x_n)^T$ 一一对应。

另一方面，如果我们设存在一个从 V 到 W 的线性映射 L ，并且 W 也有一组基 $\{f_1, \dots, f_m\}$ ，那么，对于 V 中任意一个向量 v ，我们都可以将 $L(v)$ 表示为 W 的基的线性组合。现在我们的问题是：如何从输入坐标 $(x_1, \dots, x_n)^T$ 得到输出的坐标 $(y_1, \dots, y_m)^T$ ？要回答这个问题，我们首先需要解决一个更基础的问题：如何将 $L(e_i)$ 表示为 W 的基的线性组合？不妨设

$$L(e_i) = \sum_{j=1}^m f_j a_{j,i}.$$

如果我们记 $E = [e_1, \dots, e_n], F = [f_1, \dots, f_m]$ ，那么上式可以写成

$$L(E) = FA,$$

其中 $A = (a_{j,i})_{m \times n}$ 是一个 $m \times n$ 的矩阵，被称为 L 关于基 E 和 F 的表示矩阵。进一步地，如果我们设矩阵 L 是线性映射 L 在标准基下的表示矩阵，那么我们有

$$LE = FA.$$

现在，我们回到最初的问题：如何从输入坐标 $(x_1, \dots, x_n)^T$ 得到输出的坐标 $(y_1, \dots, y_m)^T$ ？这里我们需要利用一个重要的关系：对于任意 $v \in V$ ，我们有

$$L(v) = L(e_1 x_1 + \dots + e_n x_n) = L(e_1) x_1 + \dots + L(e_n) x_n.$$

如果我们将 v 的坐标向量记作 $x = (x_1, \dots, x_n)^T$, 则上式可以写成

$$Lv = LEx = FAx.$$

因此, $L(v)$ 在基 F 下的坐标向量为 $y = Ax$. 这就回答了我们最初的问题: 通过线性映射 L , 我们可以将输入向量 v 的坐标 x 通过矩阵 A 转换为输出向量 $L(v)$ 的坐标 y .

例 2.8. 我们记 D 为求导算子。在三阶多项式空间 $P_3(\mathbb{R})$ 上, 若以 $\{1, x, x^2, x^3\}$ 为一组基, 则有

$$D[1, x, x^2, x^3] = [1, x, x^2, x^3] \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

由表示矩阵的特征值为零, 我们知道这个求导算子是有界的。

现在我们研究线性映射的复合运算。

定理 2.9. 复合映射的表示矩阵等于各个映射的表示矩阵的乘积。

证明. 我们不妨设

$$V \xrightarrow[A_1]{L_1} W \xrightarrow[A_2]{L_2} U.$$

不妨设 V 的基拼成的矩阵为 E , W 的基拼成的矩阵为 F , U 的基拼成的矩阵为 G , 则有

$$L_1E = FA_1, \quad L_2F = GA_2.$$

因此

$$L_2L_1E = L_2FA_1 = L_2FA_1 = GA_2A_1.$$

这就证明了复合映射的表示矩阵等于各个映射的表示矩阵的乘积。这也从另一个角度说明了矩阵乘法的合理性: 矩阵乘法正是为了表示线性映射的复合运算而设计的。□

从线性映射出发, 我们可以重新理解一些线性代数的基本概念。

例如, 对于一个恒等映射 $I: V \rightarrow V$, 它在标准正交基下的表示矩阵为单位阵。如果我们选取两组不同的基 E 和 F , 则有

$$IE = FA,$$

其中 A 是恒等映射关于基 E 和 F 的表示矩阵, 亦即从基 E 到基 F 的过渡矩阵。

对于 V 中任意一个向量 v , 设其在基 E 下的坐标为 x , 即 $v = Ex$, 则有

$$v = Ex = IEx = FAx = F(Ax) := Fy.$$

根据上面这个公式，我们可以得到同一个向量 v 在不同基下的坐标之间的转换关系：

$$y = Ax,$$

其中 x 是 v 在基 E 下的坐标， y 是 v 在基 F 下的坐标， A 是从基 E 到基 F 的过渡矩阵。

如果 V 和 W 是同一个一维向量空间，我们取同一组基 x . 设线性映射 $L: V \rightarrow W$ 在标准基下的表示矩阵为 A ，从基 x 到基 x 的表示矩阵为 λ ，则有

$$Hx = x\lambda.$$

从几何的角度而言，上式表示了线性映射 H 在基 x 下的作用方式：它将基向量 x 映射到其自身的一个标量倍 λx ；从代数的角度而言，上式给出了矩阵 H 的特征值 λ 和特征向量 x 之间的关系。

接下来我们来看另外一个问题。我们知道，每一个矩阵都有一个相抵标准型。那么，这个标准型在线性映射的框架下，有什么意义呢？同样地，我们不妨设线性映射 $T: V \rightarrow W$ 在标准基下的表示矩阵为 L ，线性空间 V 的基拼成的矩阵为 E ，线性空间 W 的基拼成的矩阵为 F ， T 关于基 E 和 F 的表示矩阵为 A ，则有

$$LE = FA.$$

而 A 的相抵标准型为 D ，则存在可逆矩阵 P, Q ，使得

$$P^{-1}AQ = D, \Leftrightarrow A = PDQ^{-1}.$$

因此

$$LE = FPDQ^{-1},$$

亦即

$$L(EQ) = (FP)D.$$

这意味着，我们可以做到，在 V 和 W 中选取合适的基，使得线性映射 T 的表示矩阵为其相抵标准型 D 。

现在我们来看看相似矩阵在线性映射的框架下有什么意义。设有恒等映射 $I: V \rightarrow V$ ， V 有两组基 E 和 F ， I 关于基 E 和 F 的表示矩阵为 P ，则有

$$IE = FP.$$

现在假设我们有一个线性映射 $L: V \rightarrow V$ ，它在基 E 下的表示矩阵为 A ，即

$$LE = EA.$$

则有

$$LF = LEP^{-1} = EAP^{-1} = FPAP^{-1} = F(PAP^{-1}).$$

这就说明了，线性映射 L 在基 F 下的表示矩阵为 PAP^{-1} 。因此，同一个线性映射在不同基下的表示矩阵是相似矩阵。

线性映射的像与核 我们知道，线性映射 $T: V \rightarrow W$ 会将 V 中的每一个向量都变为 W 中的一个向量。那么，所有这些产出的向量 $T(v)$ 会填满整个目标空间 W 吗？还是说它们只构成了 W 的一个子集？这个“产出”的集合长什么样子呢？

这个问题引导我们去定义所有可能输出的集合。这个集合就叫做 T 的像或值域。

定义 2.10 (线性映射的像). 设有线性映射 $T: V \rightarrow W$ ，则 T 的像定义为

$$\text{Range}(T) := \{T(v) \in W : \forall v \in V\} \subset W.$$

从直观上来说，“像”就好像是线性映射 T 在目标空间 W 中的落脚点。这个像不仅仅是一个普通的集合，它本身也是一个向量空间。这使得像空间也拥有了代数结构。像的“大小”我们也称之为秩 (rank)。

定义 2.11 (线性映射的秩). 设有线性映射 $T: V \rightarrow W$ ，则 T 的秩定义为

$$\text{rank}(T) := \dim(\text{Range}(T)).$$

秩衡量了线性映射“扩展”信息的能力。秩越大，说明线性映射能够覆盖目标空间 W 的维度就越多，信息传递得也就越充分。

如果 $\text{Range}(T) = W$ ，则称 T 是满射 (surjective)。它的像覆盖了整个目标空间。

如果 $\text{Range}(T)$ 只是 W 的一个真子空间，说明 T 在某种程度上“压缩”了空间，丢失了一部分信息。

接下来，我们来考虑丢失的这部分信息。我们知道，线性映射将 0_V 映射到 0_W ；那么，有没有可能其他的非零向量也被映射到 0_W 呢？如果有，哪些向量被 T “湮灭”成了零向量呢？这个问题引导我们去定义所有被映射为零向量的输入向量的集合。这个集合就叫做 T 的核。

定义 2.12 (线性映射的核). 设有线性映射 $T: V \rightarrow W$ ，则 T 的核定义为

$$\text{Ker}(T) := \{v \in V : T(v) = 0_W\} \subset V.$$

“核”这个概念就是在线性映射 T 的视角下，“看起来像零”的所有向量的集合。“核”也不是一个杂乱的集合，它同样是一个向量空间。核空间的大小我们称之为零度 (nullity)。

定义 2.13 (线性映射的零度). 设有线性映射 $T: V \rightarrow W$ ，则 T 的零度定义为

$$\text{null}(T) := \dim(\text{Ker}(T)).$$

它衡量了线性映射“压缩”信息的能力。零度越大，说明线性映射湮灭的信息就越多，信息丢失得也就越严重。

如果 $\text{Ker}(T) = \{0_V\}$ ，即核空间只有零向量，这说明 T 没有丢失任何信息。不同的输入必然得到不同的输出（可用反证法证明），因此 T 必然是一个单射（injective）。

反之，如果 $\text{Ker}(T)$ 包含非零向量，说明 T 将多个不同的输入向量“混淆”在一起，映射到了同一个输出向量。这是因为如果 $v_1 - v_2 \in \text{Ker}(T)$ ，那么必定会有 $T(v_1) = T(v_2)$ 。这就造成了信息的丢失。 T 也不再是单射。

线性映射的像与核是两个非常重要的概念。像回答了“我们得到了什么”这个问题，描述了映射的输出范围和维度；核回答了“我们失去了什么”这个问题，揭示了映射中信息的丢失和压缩。这两个概念是互补的，共同深刻地刻画了一个线性映射的本质。它们之间的关系最终由强大的秩-零度定理揭示，揭示了输入空间的维度、信息损失的维度和有效输出的维度之间的守恒关系。

定理 2.14 (秩-零度定理). 设有线性映射 $T: V \rightarrow W$ ，则有

$$\text{rank}(T) + \text{null}(T) = \dim(V).$$

证明. 设 $\dim V = n, \dim W = m$. 我们取 V 的一组基 $\{v_1, \dots, v_n\}$ ，并设 $\{v_1, \dots, v_{n-r}\}$ 是 $\text{Ker}(T)$ 的一组基。

对 V 中任意一个向量 v ，我们设

$$v = \sum_{i=1}^n v_i x_i.$$

则有

$$T(v) = \sum_{i=1}^n T(v_i) x_i = \sum_{i=n-r+1}^n T(v_i) x_i.$$

由于线性映射保持线性无关性，因此 $\{T(v_{n-r+1}), \dots, T(v_n)\}$ 是 $\text{Range}(T)$ 的一组基。因此 $\text{rank}(T) = r$ 。另一方面，由于 $\{v_1, \dots, v_{n-r}\}$ 是 $\text{Ker}(T)$ 的一组基，因此 $\text{null}(T) = n - r$ 。综上所述，我们有

$$\text{rank}(T) + \text{null}(T) = r + (n - r) = n = \dim(V).$$

□

满秩分解 我们知道，一个 $m \times n$ 的矩阵 A 代表了一个从 \mathbb{R}^n 到 \mathbb{R}^m 的线性映射。我们还知道

$$\text{rank}(A) \leq \min\{m, n\}.$$

那么，如果矩阵的秩 r 远小于 n 和 m ，这意味着什么？这意味着，虽然输入和输出空间的维度很高，但这个映射的“有效信息”或者“核心作用”实际上只发生在一个低维的 r 维空间里。整个变换过程存在一个“瓶颈”，限制了信息的传递和变化。我们能否把这个过程明确地写出来呢？即：

1. 一个**编码/压缩**的过程：将高维输入向量 $x \in \mathbb{R}^n$ 压缩到一个低维的 r 维中间表示。

2. 一个**解码/重构**的过程：从这个低维的中间表示，重构出最终的高维输出向量 $Ax \in \mathbb{R}^m$.

“压缩”过程是从 \mathbb{R}^n 到 \mathbb{R}^r 的映射，我们希望它能够保留所有来自行空间的信息。最理想的编码器是一个 $n \times r$ 的矩阵 R ，它应该是行满秩的，也就是说，它没有冗余的行，最大程度地保留信息。

“重构”过程是从 \mathbb{R}^r 到 \mathbb{R}^m 的映射，我们希望它能够构建出完整的像空间。最理想的解码器是一个 $m \times r$ 的矩阵 C ，它应当是列满秩的，也就是说，它的列向量本身就是像空间的一组基，没有冗余的维度。

我们的目标就是找到这样的 C 和 R ，使得原映射 A 可以分解为这两个映射的复合：

$$A_{m \times n} = C_{m \times r} R_{r \times n},$$

其中 $\text{rank}(C) = \text{rank}(R) = \text{rank}(A) = r$. 这个分解就叫做满秩分解。那么，我们该怎样构造这个分解呢？

我们知道， A 的主元列构成了其像空间的一组基。我们直接把这些主元列从原矩阵 A 中提取出来，组成一个新的矩阵 C . 根据定义，这个矩阵 C 就是 $m \times r$ 的，并且是列满秩的。

我们还知道，矩阵 A 的行最简阶梯型的非零行构成了其行空间的一组基。我们直接把这些非零行抽取出来，组成一个新的矩阵 R . 根据定义，这个矩阵 R 就是 $r \times n$ 的，并且是行满秩的。

现在我们有 C 和 R ，为什么一定有 $A = CR$ 呢？这是因为， A 的任何一列都可以表示为 C 的列向量的线性组合，而组合的系数恰好就是 R 的对应列。这个关系可以精确地写成

$$A = CR.$$

这就完成了满秩分解。我们最终的结论就是：任何秩为 r 的矩阵 A ，都可以被看作是先通过一个行满秩矩阵 R 将向量从 \mathbb{R}^n 映射到 \mathbb{R}^r （保留核心信息），再通过一个列满秩矩阵 C 将向量从 \mathbb{R}^r 映射到 \mathbb{R}^m （重构最终图象）的过程。从几何的角度看，满秩分解 $A = CR$ 精确地解释了任何线性映射 T_A 的内在结构：它本质上是一个投影（到行空间），然后是一个同构（从行空间到列空间），最后是一个嵌入（将列空间放入目标空间）。 R 负责投影， C 负责同构和嵌入。

Schur 补 下面我们基于线性映射的观点，引入与秩相关的另一个重要话题：Schur 补。当然，这种引入并不是最自然的引入方式，或多或少有些强行解释的意味。但谁让本讲的内容是线性映射呢？

我们考虑线性映射 $T: V \rightarrow W$. 我们知道， V 和 W 都有其子空间。我们不妨将 V 和 W 分别分解为两个子空间的直和：

$$V = V_1 \oplus V_2, \quad W = W_1 \oplus W_2.$$

那么, T 会如何作用于这些子空间呢? 如果我们设 T 在标准基下的表示矩阵为 A , 则 A 就可以被分解为四个块:

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},$$

其中 A_{11} 是从 V_1 到 W_1 的映射, A_{12} 是从 V_2 到 W_1 的映射, A_{21} 是从 V_1 到 W_2 的映射, A_{22} 是从 V_2 到 W_2 的映射。

但是, 从 V_2 到 W_2 的映射可能受到 V_1 的影响。我们能否通过一定方式, 消除这种影响, 使得 V_2 到 W_2 的映射只依赖于 V_2 本身, 而不受 V_1 的干扰呢? 这就是 Schur 补的核心思想。假设 A_{11} 是可逆的, 我们可以通过以下变换来消除 V_1 对 W_2 的影响:

$$S := A_{22} - A_{21}A_{11}^{-1}A_{12}.$$

这个矩阵 S 就是 A 关于块 A_{11} 的 Schur 补。Schur 补 S 的意义在于, 它描述了在消除 V_1 对 W_2 影响后的, V_2 到 W_2 的“纯净”映射。

当然, 这个解释有些牵强。下面我们给出更加自然的理解方式。

我们知道, 求解线性方程组可以利用 Gauss 消去法。那么, 如果方程 $Ax = b$ 中, A, x, b 都天然地可以被分成块, 我们又应该如何求解呢? 设

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}.$$

我们能不能尝试模拟高斯消去法, 尝试消去一部分变量呢? 就像 Gauss 消去法中, 我们需要选取主元一样, 这里我们假设 A_{11} 是可逆的——它的作用就好像是 Gauss 消去法里的主元。然后, 我们就可以快乐地消去了:

$$A = \begin{bmatrix} I & O \\ A_{21}A_{11}^{-1} & I \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ O & S \end{bmatrix}, \quad S = A_{22} - A_{21}A_{11}^{-1}A_{12}.$$

这样, 我们就将原本的方程组 $Ax = b$ 转化为了两个更简单的方程组:

$$\begin{cases} A_{11}x_1 + A_{12}x_2 = b_1, \\ Sx_2 = b_2 - A_{21}A_{11}^{-1}b_1. \end{cases}$$

我们可以先解第二个方程组, 得到 x_2 , 然后再代入第一个方程组, 解出 x_1 . 这就是 Schur 补的一个重要应用。而在上面, 我们也看到了, 我们可以利用 Schur 补来对矩阵进行 LU 分解, 进而可以利用 LU 分解来高效地求解线性方程组。

不变子空间 现在我们来讨论线性映射的另一个重要话题: 不变子空间。设有线性映射 $T: V \rightarrow V$, 这个映射可能会非常复杂。它在某个基下的表示矩阵可能是没有任何明显规律的稠密矩阵。我

我们的终极目标是尽可能地简化我们对 T 的理解。何谓“简单”？最简单的矩阵自然是对角阵；如果不行，退而求其次，分块对角阵也不错；再不济，分块上三角阵也行。我们如何找到一个合适的基，使得 T 在这个基下的表示矩阵尽可能地“简单”（最好是分块对角的或分块上三角的）呢？

我们已经知道一个非常特殊的情况：特征向量。如果 v 是 T 的一个特征向量，有 $T(v) = v\lambda$ 。这意味着，由 v 张成的一维子空间 $U = \text{span}(v)$ 有一个非常良好的性质：对于任何在 U 中的向量 u ，它的像 $T(u)$ 仍然在 U 中。

我们可以说，算子 T 把子空间 U “封闭”在自身内部。 T 在这个子空间 U 上的作用就被极大地简化了：仅仅是一个数乘。这个被“封闭”的一维子空间就是我们即将学习的不变子空间的最简单的案例。

既然存在被 T 封闭的直线，那么是否存在被 T 封闭的更高维的子空间呢？想象一下，如果我们能找到 V 的一个子空间 U ，使得 T 将 U 中的每一个向量都映射回 U 内部，那么，我们就可以把注意力暂时从整个复杂的空间 V 转移到这个更小、更简单的子空间 U 上，去研究 T 在 U 上的局部行为。这个局部算子 $T|_U : U \rightarrow U$ 的性质将会更加简单明了。

上面的想法直接引出了不变子空间的正式定义：

定义 2.15 (不变子空间). 设 $T : V \rightarrow V$ 是一个线性算子。如果 V 的一个子空间 U 满足

$$T(u) \in U, \quad \forall u \in U,$$

则称 U 是 T 的一个不变子空间，记作 $T(U) \subseteq U$ 。

不变子空间到底好在哪？它好就好在，它直通我们的最初目标——寻找一个简单的矩阵表示。假设我们找到了一个 k 维的不变子空间 U 。现在我们用一种巧妙的方式来构造 V 的基：

1. 首先，为 U 选取一组基 $\{u_1, \dots, u_k\}$ 。
2. 然后，将这组基扩展为 V 的一组基 $\{u_1, \dots, u_k, v_{k+1}, \dots, v_n\}$ 。

现在，我们就来看看， T 在这组基下的表示矩阵是什么样子的。由于 U 是不变子空间，对于基的前 k 个向量而言， $T(u_j)$ 仍然在 U 中，因此它可以表示为 $\{u_1, \dots, u_k\}$ 的线性组合，而与 $\{v_i\}$ 无关。这就意味着，表示矩阵的前 k 列的下半部分全是零。所以，矩阵 A 必定具有如下的分块上三角形式：

$$A = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix},$$

其中 A_{11} 是 $k \times k$ 的矩阵，表示 T 在不变子空间 U 上的局部作用，亦即 $T|_U$ 在基 $\{u_1, \dots, u_k\}$ 下的表示矩阵。

进一步地，如果整个空间 V 可以被分解为两个不变子空间 U_1, U_2 的直和，即 $V = U_1 \oplus U_2$ ，那么我们得到的表示矩阵将是分块对角的！

$$A = \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix},$$

其中 A_{11} 和 A_{22} 分别是 T 在 U_1 和 U_2 上的局部作用的表示矩阵。这就大大简化了我们对 T 的理解。我们可以分别研究 A_{11} 和 A_{22} ，而不需要面对整个复杂的矩阵 A 。这就是不变子空间的强大之处。寻找不变子空间的过程，本质上就是在为线性算子寻找一个“断裂点”，从而将其矩阵表示分解为更小、更易于处理的块。这是通往算子分解和理解其内在结构（如 Jordan 标准型）的关键一步。

3 Sherman-Morrison-Woodbury 公式

在许多现实问题中，我们经常会遇到这样的情况：我们可能已经花费了巨大的计算资源求得了一个大矩阵 A 的逆 A^{-1} ，但后来我们发现，我们还需要对矩阵 A 进行一个小小的修正，得到一个新的矩阵 A' 。这个修正可能是由于加入了新的数据、改变了某个参数或者修正了模型。我们是否需要从头重新计算它的逆呢？如果我们能利用已经计算好的 A^{-1} ，只对修正部分进行增量更新，那将大大节省计算资源。

最常见和最重要的一类修正是“低秩修正”，即新的矩阵可以表示为 $A' = A - XB^{-1}Y^*$ ，其中 $X \in \mathbb{R}^{n \times k}, Y \in \mathbb{R}^{n \times k}, B \in \mathbb{R}^{k \times k}$ ，并且通常 $k \ll n$ 。当 $k = 1$ 时，这就是最简单的秩一修正： $A' = A - xy^*$ ，其中 x, y 是列向量。我们的目标就是：已知 A^{-1} ，如何高效地计算 A'^{-1} ？

我们可以先将 A^{-1} 提出来，得到

$$A'^{-1} = (A - XB^{-1}Y^*)^{-1} = (I - A^{-1}XB^{-1}Y^*)^{-1}A^{-1}.$$

现在，我们只需要解决如何计算 $(I - XY^*)^{-1}$ 即可。

首先我们可以先来猜一下答案。由 Neumann 级数展开，我们有

$$\begin{aligned} (I - XY^*)^{-1} &= I + XY^* + XY^*XY^* + \dots \\ &= I + X(I + Y^*X + (Y^*X)^2 + \dots)Y^* \\ &= I + X(I - Y^*X)^{-1}Y^*. \end{aligned}$$

当然，这个推导并不严谨，因为我们并没有证明级数的收敛性。但它给了我们一个启发：或许答案就是 $I + X(I - Y^*X)^{-1}Y^*$ 。接下来，我们就来验证这个结果的正确性。

$$\begin{aligned}
(I - XY^*)(I + X(I - Y^*X)^{-1}Y^*) &= I - XY^* + XY^* + XY^*X(I - Y^*X)^{-1}Y^* \\
&= I + X(Y^*X(I - Y^*X)^{-1} - I)(Y^*) \\
&= I + X((I - Y^*X)(I - Y^*X)^{-1} - I)(Y^*) \\
&= I.
\end{aligned}$$

这就验证了我们的猜测是正确的。我们最终得到了一个非常优雅公式：

$$(I - XY^*)^{-1} = I + X(I - Y^*X)^{-1}Y^*.$$

我们代入之前的式子，得到

$$(A - XB^{-1}Y^*)^{-1} = A^{-1} + A^{-1}X(B - Y^*A^{-1}X)^{-1}Y^*A^{-1}.$$

这就是著名的 Sherman-Morrison-Woodbury 公式。

定理 3.1 (Sherman-Morrison-Woodbury 公式). 设 $A \in \mathbb{R}^{n \times n}$ 是可逆矩阵, $X, Y \in \mathbb{R}^{n \times k}$, $B \in \mathbb{R}^{k \times k}$ 是可逆矩阵, 若 $B - Y^*A^{-1}X$ 可逆, 则有

$$(A - XB^{-1}Y^*)^{-1} = A^{-1} + A^{-1}X(B - Y^*A^{-1}X)^{-1}Y^*A^{-1}.$$

那么, 这个公式的威力何在呢? 公式右侧最关键的部分是计算 $(B - Y^*A^{-1}X)$ 的逆。注意到, 这个矩阵的大小是 $k \times k$, 而 k 通常远小于 n . 因此, 我们只需要花费 $O(k^3)$ 的代价来计算这个矩阵的逆, 其余部分都是矩阵乘法。整个更新计算的复杂度大约是 $O(k^3 + n^2k)$, 这个代价应当是远远低于 $O(n^3)$ 的。这就是 SMW 公式的威力所在。

我们同样可以注意到, 公式中的核心项 $B - Y^*A^{-1}X$ 正是 Schur 补。如我们考虑分块矩阵

$$\begin{bmatrix} A & X \\ Y^* & B \end{bmatrix} = \begin{bmatrix} I & O \\ Y^*A^{-1} & I \end{bmatrix} \begin{bmatrix} A & O \\ O & B - Y^*A^{-1}X \end{bmatrix} \begin{bmatrix} I & X(B - Y^*A^{-1}X)^{-1} \\ O & I \end{bmatrix},$$

如果我们想要求大矩阵的逆, 我们只需要分别求三个矩阵的逆, 然后将它们按相反顺序相乘即可。第一个和第三个矩阵的逆都非常简单, 而中间的矩阵的逆正是

$$\begin{bmatrix} A^{-1} & O \\ O & (B - Y^*A^{-1}X)^{-1} \end{bmatrix}.$$

这就再次说明了 Schur 补在这里的核心作用。