7.4 奇异值分解

5.3 节和 7.1 节的对角化定理在许多应用中均很重要，然而，如我们所知，不是所有矩阵都有分解式 $A=PDP^{-1}$，且 D 是对角的. 但分解 $A=QDP^{-1}$ 对任意 $m\times n$ 矩阵 A 都有可能！这类特殊分解称为奇异值分解，是线性代数应用中最有用的矩阵分解.

奇异值分解基于一般的矩阵对角化性质，可以被长方形矩阵模仿：一个对称矩阵 A 的特征值的绝对值表示度量 A 拉长或压缩一个向量（特征向量）的程度. 如果 $A\boldsymbol{x}=\lambda\boldsymbol{x}$，且 $\|\boldsymbol{x}\|=1$，那么

$$\|A\boldsymbol{x}\|=\|\lambda\boldsymbol{x}\|=|\lambda|\|\boldsymbol{x}\|=|\lambda| \tag{1}$$

如果 λ_1 是具有最大数值的特征值，那么对应的单位特征向量 $\boldsymbol{v}_1$ 确定一个由 A 的拉长影响最大的方向，也就是说，（1）式表示当 $\boldsymbol{x}=\boldsymbol{v}_1$ 时，$A\boldsymbol{x}$ 的长度最大化，且 $\|A\boldsymbol{v}_1\|=|\lambda_1|$. 这个对 $\boldsymbol{v}_1$ 和 $|\lambda_1|$ 的描述对于长方形的矩阵来说也是类似的，这将导致奇异值分解.

例 1 如果 $A=\begin{bmatrix}4&11&14\\8&7&-2\end{bmatrix}$，那么线性变换 $\boldsymbol{x}\mapsto A\boldsymbol{x}$ 将 $\mathbb{R}^3$ 中的单位球 $\{\boldsymbol{x}:\|\boldsymbol{x}\|=1\}$ 映上到 $\mathbb{R}^2$ 中的椭圆，见图 7-13. 找出使得长度 $\|A\boldsymbol{x}\|$ 最大的一个单位向量 $\boldsymbol{x}$，且计算这个最大长度.

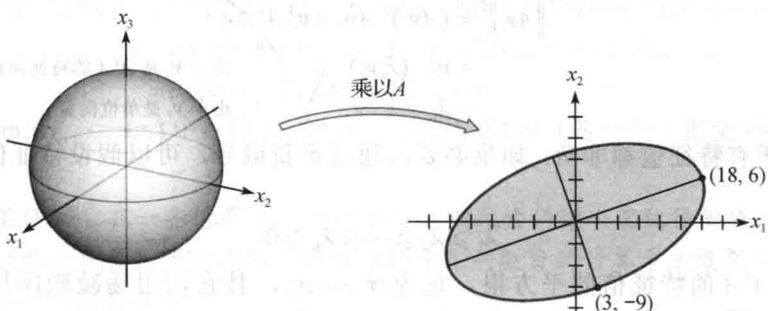

图 7-13 从 $\mathbb{R}^3$ 到 $\mathbb{R}^2$ 的一个变换

解 使 $\|A\boldsymbol{x}\|^2$ 的值最大化的 $\boldsymbol{x}$ 同样使 $\|A\boldsymbol{x}\|$ 的值最大化，但 $\|A\boldsymbol{x}\|^2$ 更容易计算. 注意到

$$\|A\boldsymbol{x}\|^2=(A\boldsymbol{x})^{\mathrm{T}}(A\boldsymbol{x})=\boldsymbol{x}^{\mathrm{T}}A^{\mathrm{T}}A\boldsymbol{x}=\boldsymbol{x}^{\mathrm{T}}(A^{\mathrm{T}}A)\boldsymbol{x}$$

由于 $(A^{\mathrm{T}}A)^{\mathrm{T}}=A^{\mathrm{T}}A^{\mathrm{TT}}=A^{\mathrm{T}}A$，故 $A^{\mathrm{T}}A$ 也是一个对称矩阵. 所以，现在的问题转化为在条件 $\|\boldsymbol{x}\|=1$ 的限制下二次型 $\boldsymbol{x}^{\mathrm{T}}(A^{\mathrm{T}}A)\boldsymbol{x}$ 的最大化. 利用 7.3 节的定理 6，最大值就是矩阵 $A^{\mathrm{T}}A$ 的最大特征值 λ_1，同样，最大值可以在 $A^{\mathrm{T}}A$ 的特征值 λ_1 对应的单位特征向量处获得.

对本例中的矩阵 A，

$$A^{\mathrm{T}}A=\begin{bmatrix}4&8\\11&7\\14&-2\end{bmatrix}\begin{bmatrix}4&11&14\\8&7&-2\end{bmatrix}=\begin{bmatrix}80&100&40\\100&170&140\\40&140&200\end{bmatrix}$$

矩阵 $A^{\mathrm{T}}A$ 的特征值是 $\lambda_1=360$，$\lambda_2=90$ 和 $\lambda_3=0$，对应的单位特征向量分别是

$$\boldsymbol{v}_1=\begin{bmatrix}1/3\\2/3\\2/3\end{bmatrix},\boldsymbol{v}_2=\begin{bmatrix}-2/3\\-1/3\\2/3\end{bmatrix},\boldsymbol{v}_3=\begin{bmatrix}2/3\\-2/3\\1/3\end{bmatrix}$$

$\|A\boldsymbol{x}\|^2$ 的最大值是 360，且在 $\boldsymbol{x}$ 为单位向量 $\boldsymbol{v}_1$ 处获得. 向量 $A\boldsymbol{v}_1$ 是图 7-13 中椭圆上离原点最远的点，即

$$A\boldsymbol{v}_1=\begin{bmatrix}4&11&14\\8&7&-2\end{bmatrix}\begin{bmatrix}1/3\\2/3\\2/3\end{bmatrix}=\begin{bmatrix}18\\6\end{bmatrix}$$

对 $\|\boldsymbol{x}\|=1$，$\|A\boldsymbol{x}\|$ 的最大值是 $\|A\boldsymbol{v}_1\|=\sqrt{360}=6\sqrt{10}$. ■

例 1 说明，A 对 $\mathbb{R}^3$ 中单位球面的影响与二次型 $\boldsymbol{x}^{\mathrm{T}}(A^{\mathrm{T}}A)\boldsymbol{x}$ 有关. 事实上，变换 $\boldsymbol{x}\to A\boldsymbol{x}$ 的全部几何特性都可用二次型来说明，下面我们将说明这个问题.

一个 $m\times n$ 矩阵的奇异值

令 A 是 $m\times n$ 矩阵，那么 $A^{\mathrm{T}}A$ 是对称矩阵且可以正交对角化. 令 $\{\boldsymbol{v}_1,\cdots,\boldsymbol{v}_n\}$ 是 $\mathbb{R}^n$ 的单位正交基且构成 $A^{\mathrm{T}}A$ 的特征向量，$\lambda_1,\cdots,\lambda_n$ 是 $A^{\mathrm{T}}A$ 对应的特征值，那么对 $1\leqslant i\leqslant n$，

$$\begin{aligned}\|A\boldsymbol{v}_i\|^2&=(A\boldsymbol{v}_i)^{\mathrm{T}}Av_i=\boldsymbol{v}_i^{\mathrm{T}}A^{\mathrm{T}}A\boldsymbol{v}_i\\&=\boldsymbol{v}_i^{\mathrm{T}}\ (\lambda_i\boldsymbol{v}_i)&&\text{由于 }\boldsymbol{v}_i\text{ 是 }A^{\mathrm{T}}A\text{ 的特征向量}\\&=\lambda_i&&\text{由于 }\boldsymbol{v}_i\text{ 是单位向量}\end{aligned}\tag{2}$$

所以，$A^{\mathrm{T}}A$ 的所有特征值都非负. 如果必要，通过重新编号，可以假设特征值的重新排列满足

$$\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_n\geqslant 0$$

A 的**奇异值**是 $A^{\mathrm{T}}A$ 的特征值的平方根，记为 $\sigma_1,\cdots,\sigma_n$，且它们用递减顺序排列，也就是对 $1\leqslant i\leqslant n$，$\sigma_i=\sqrt{\lambda_i}$. 由（2）可知，$A$ 的奇异值是向量 $A\boldsymbol{v}_1,\cdots,A\boldsymbol{v}_n$ 的长度.

例 2　令 A 是例 1 中的矩阵，由于 $A^{\mathrm{T}}A$ 的特征值是 360，90 和 0，故 A 的奇异值是

$$\sigma_1=\sqrt{360}=6\sqrt{10},\ \sigma_2=\sqrt{90}=3\sqrt{10},\ \sigma_3=0$$

从例 1 可知，A 的第一个奇异值是 $\|A\boldsymbol{x}\|$ 在所有单位向量处的最大值，且最大值可以在单位向量 $\boldsymbol{v}_1$ 处获得. 7.3 节的定理 7 表明，A 的第二个奇异值是 $\|A\boldsymbol{x}\|$ 在所有与 $\boldsymbol{v}_1$ 正交的单位向量上的最大值，且这个最大值可以在第二个单位特征向量 $\boldsymbol{v}_2$ 处获得（习题 22）. 对例 1 中的 $\boldsymbol{v}_2$，

$$A\boldsymbol{v}_2=\begin{bmatrix}4&11&14\\8&7&-2\end{bmatrix}\begin{bmatrix}-2/3\\-1/3\\2/3\end{bmatrix}=\begin{bmatrix}3\\-9\end{bmatrix}$$

这个点在图 7-13 中椭圆的短轴上，像 $A\boldsymbol{v}_1$ 在长轴上一样（见图 7-14）. 矩阵 A 的前两个奇异值是椭圆长半轴和短半轴的长度.

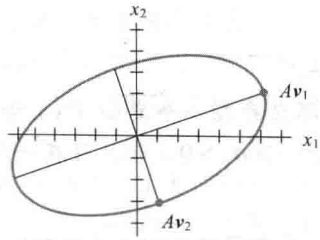

图 7-14

事实上，像下面定理所说，图 7-14 中的 $A\boldsymbol{v}_1$ 和 $A\boldsymbol{v}_2$ 正交不是偶然.

定理 9 假若 $\{\boldsymbol{v}_1,\cdots,\boldsymbol{v}_n\}$ 是包含 $A^{\mathrm{T}}A$ 的特征向量的 $\mathbb{R}^n$ 上的单位正交基，重新整理使得对应的 $A^{\mathrm{T}}A$ 的特征值满足 $\lambda_1 \geqslant \cdots \geqslant \lambda_n$. 假若 A 有 r 个非零奇异值，那么 $\{A\boldsymbol{v}_1,\cdots,A\boldsymbol{v}_r\}$ 是 $\mathrm{Col}A$ 的一个正交基，且 $\mathrm{rank}\, A = r$.

证 由于当 $i \neq j$ 时，$\boldsymbol{v}_i$ 和 $\lambda_j \boldsymbol{v}_j$ 正交，所以

$$(A\boldsymbol{v}_i)^{\mathrm{T}}(A\boldsymbol{v}_j) = \boldsymbol{v}_i^{\mathrm{T}} A^{\mathrm{T}} A \boldsymbol{v}_j = \boldsymbol{v}_i^{\mathrm{T}}(\lambda_j \boldsymbol{v}_j) = 0$$

从而 $\{A\boldsymbol{v}_1,\cdots,A\boldsymbol{v}_n\}$ 是一个正交基. 更进一步，由于向量 $A\boldsymbol{v}_1,\cdots,A\boldsymbol{v}_n$ 的长度是 A 的奇异值，且因为有 r 个非零奇异值，因此 $A\boldsymbol{v}_i \neq \boldsymbol{0}$ 的充分必要条件是 $1 \leqslant i \leqslant r$. 所以 $A\boldsymbol{v}_1,\cdots,A\boldsymbol{v}_r$ 是线性无关向量，且属于 $\mathrm{Col}A$. 最后，对任意属于 $\mathrm{Col}\, A$ 的 $\boldsymbol{y}$，比如 $\boldsymbol{y} = A\boldsymbol{x}$，我们可以写出 $\boldsymbol{x} = c_1\boldsymbol{v}_1 + \cdots + c_n\boldsymbol{v}_n$，且

$$\begin{aligned} \boldsymbol{y} = A\boldsymbol{x} &= c_1 A\boldsymbol{v}_1 + \cdots + c_r A\boldsymbol{v}_r + c_{r+1}A\boldsymbol{v}_{r+1} + \cdots + c_n A\boldsymbol{v}_n \\ &= c_1 A\boldsymbol{v}_1 + \cdots + c_r A\boldsymbol{v}_r + 0 + \cdots + 0 \end{aligned}$$

这样，$\boldsymbol{y}$ 在 $\mathrm{Span}\{A\boldsymbol{v}_1,\cdots,A\boldsymbol{v}_r\}$ 中，这说明 $\{A\boldsymbol{v}_1,\cdots,A\boldsymbol{v}_r\}$ 是 $\mathrm{Col}\, A$ 的一个（正交）基. 因此 $\mathrm{rank}\, A = \dim \mathrm{Col}\, A = r$. ■

数值计算的注解 在一些情形下，A 的秩对 A 中元素的微小变化很敏感. 如果用计算机对矩阵 A 作行化简，那么很明显矩阵 A 的主元列数目的计算方法效果不好，舍入误差常导致满秩的阶梯矩阵.

实际上，估计大矩阵 A 的秩时，最可靠的方法是计算非零奇异值的个数. 在这种情形下，特别小的非零奇异值在实际计算中常假定为零，矩阵 A 的有效秩是剩余非零奇异值的数目.[㊀]

奇异值分解

矩阵 A 的分解涉及一个 $m \times n$ “对角”矩阵 Σ，其形式是

$$\Sigma = \begin{array}{c} \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} \begin{array}{l} \\ \leftarrow m-r \text{ 行} \end{array} \\ \uparrow\ n-r \text{ 行} \end{array} \tag{3}$$

其中，D 是一个 $r \times r$ 对角矩阵，且 r 不超过 m 和 n 中较小的那个（如果 r 等于 m 或 n，或都相等，

㊀ 一般地，秩的估计问题并不简单，对这个问题的详细讨论见 Philip E. Gill, Walter Murray, and Margaret H. Wright, *Numerical Linear Algebra and Optimization*, vol.1 (Redwood City, CA: Addison-Wesley, 1991), 5.8 节.

则 M 中不会出现零矩阵）.

定理 10 （奇异值分解）

设 A 是秩为 r 的 $m\times n$ 矩阵，那么存在一个类似（3）中的 $m\times n$ 矩阵Σ，其中 D 的对角线元素是 A 的前 r 个奇异值，$\sigma_1\geqslant\sigma_2\geqslant\cdots\geqslant\sigma_r>0$，并且存在一个 $m\times m$ 正交矩阵 U 和一个 $n\times n$ 正交矩阵 V 使得 $A=U\Sigma V^{\mathrm{T}}$.

任何分解 $A=U\Sigma V^{\mathrm{T}}$ 称为 A 的一个**奇异值分解**（或 SVD），其中 U 和 V 是正交矩阵，Σ形如（3）式，D 具有正的对角线元素. 矩阵 U 和 V 不是由 A 惟一确定的，但Σ的对角线元素必须是 A 的奇异值，见习题 19. 这样的一个分解中 U 的列称为 A 的**左奇异向量**，而 V 的列称为 A 的**右奇异向量**.

证 假设λ_i和$\boldsymbol{v}_i$ 如定理 9，使得$\{A\boldsymbol{v}_1,\cdots,A\boldsymbol{v}_r\}$是 Col A 的正交基. 将每一个 $A\boldsymbol{v}_i$ 单位化得到一个单位正交基$\{\boldsymbol{u}_1,\cdots,\boldsymbol{u}_r\}$，其中

$$\boldsymbol{u}_i=\frac{1}{\|A\boldsymbol{v}_i\|}A\boldsymbol{v}_i=\frac{1}{\sigma_i}A\boldsymbol{v}_i$$

而且

$$A\boldsymbol{v}_i=\sigma_i\boldsymbol{u}_i\quad（1\leqslant i\leqslant r） \tag{4}$$

现在将$\{\boldsymbol{u}_1,\cdots,\boldsymbol{u}_r\}$扩充为 $\mathbb{R}^m$ 的单位正交基$\{\boldsymbol{u}_1,\cdots,\boldsymbol{u}_m\}$，并且取

$$U=[\boldsymbol{u}_1\quad\boldsymbol{u}_2\quad\cdots\quad\boldsymbol{u}_m]\quad 和\quad V=[\boldsymbol{v}_1\quad\boldsymbol{v}_2\quad\cdots\quad\boldsymbol{v}_n]$$

由构造可知，U和V 是正交矩阵，同样由（4）

$$AV=[A\boldsymbol{v}_1\ \cdots\ A\boldsymbol{v}_r\ \ \boldsymbol{0}\ \cdots\ \boldsymbol{0}]=[\sigma_1\boldsymbol{u}_1\ \cdots\ \sigma_r\boldsymbol{u}_r\ \ \boldsymbol{0}\ \cdots\ \boldsymbol{0}]$$

设 D 是对角线元素为 $\sigma_1,\cdots,\sigma_r$ 的对角矩阵，Σ是上面（3）中的形式，那么

$$\begin{aligned}U\Sigma&=[\boldsymbol{u}_1\ \boldsymbol{u}_2\cdots\boldsymbol{u}_m]\left[\begin{array}{cccc|c}\sigma_1 & & & 0 & \\ & \sigma_2 & & & 0\\ & & \ddots & & \\ 0 & & & \sigma_r & \\ \hline & & 0 & & 0\end{array}\right]\\&=[\sigma_1\boldsymbol{u}_1\cdots\sigma_r\boldsymbol{u}_r\ \boldsymbol{0}\cdots\boldsymbol{0}]\\&=AV\end{aligned}$$

由于 V 是一个正交矩阵，因此 $U\Sigma V^{\mathrm{T}}=AVV^{\mathrm{T}}=A$. ■

下面的两个例子侧重于讨论奇异值分解的内部结构. 一个有效而数值稳定的分解算法会使用另外一种方法. 参见本节末尾处的“数值计算的注解”.

例 3 使用例 1 和例 2 中的结果求 $A=\begin{bmatrix}4 & 11 & 14\\ 8 & 7 & -2\end{bmatrix}$的一个奇异值分解.

解 一个奇异值分解可分三步进行.

第一步. 将矩阵 $A^{\mathrm{T}}A$ 正交对角化. 即求矩阵 $A^{\mathrm{T}}A$ 的特征值及其对应的特征向量的单位正交集. 如果 A 只有两列，则其计算可以手算完成. 对规模较大的矩阵通常需要使用矩阵程序. ㊀

㊀ 参阅软件和几何计算器命令的学习指南. 例如，MATLAB 用一个命令 eig 就可以求出特征值和特征向量.

而对此处的矩阵 A，在例 1 中已经给出了 $A^{\mathrm{T}}A$ 的特征值及其相应的特征向量.

第二步. 算出 V 和Σ. 将 $A^{\mathrm{T}}A$ 的特征值按降序排列. 在例 1 中,特征值已经按降序排列:360，90，0. 它们对应的单位特征向量 $\boldsymbol{v}_1,\boldsymbol{v}_2,\boldsymbol{v}_3$ 是 A 的右奇异向量. 使用例 1 的结果，构造

$$V=[\boldsymbol{v}_1\ \ \boldsymbol{v}_2\ \ \boldsymbol{v}_3]=\begin{bmatrix}1/3 & -2/3 & 2/3\\ 2/3 & -1/3 & -2/3\\ 2/3 & 2/3 & 1/3\end{bmatrix}$$

特征值的平方根就是奇异值：$\sigma_1=\sqrt{360}=6\sqrt{10},\ \sigma_2=\sqrt{90}=3\sqrt{10},\ \sigma_3=0$. 其中的非零奇异值是矩阵 D 的对角线元素. 矩阵Σ与矩阵 A 的行列数相同，以矩阵 D 为其左上角，其他元素为 0.

$$D=\begin{bmatrix}6\sqrt{10} & 0\\ 0 & 3\sqrt{10}\end{bmatrix},\ \Sigma=[D\ \ 0]=\begin{bmatrix}6\sqrt{10} & 0 & 0\\ 0 & 3\sqrt{10} & 0\end{bmatrix}$$

第三步. 构造 U. 当矩阵 A 的秩为 r 时，矩阵U 的前 r 列是从 $A\boldsymbol{v}_1\cdots A\boldsymbol{v}_r$ 计算得到的单位向量. 此例中，A 有两个非零奇异值，因此 $\operatorname{rank} A=2$. 根据例 2 之前的内容和方程（2），有 $\|A\boldsymbol{v}_1\|=\sigma_1$，$\|A\boldsymbol{v}_2\|=\sigma_2$，于是

$$\boldsymbol{u}_1=\frac{1}{\sigma_1}A\boldsymbol{v}_1=\frac{1}{6\sqrt{10}}\begin{bmatrix}18\\ 6\end{bmatrix}=\begin{bmatrix}3/\sqrt{10}\\ 1/\sqrt{10}\end{bmatrix}$$

$$\boldsymbol{u}_2=\frac{1}{\sigma_2}A\boldsymbol{v}_2=\frac{1}{3\sqrt{10}}\begin{bmatrix}3\\ -9\end{bmatrix}=\begin{bmatrix}1/\sqrt{10}\\ -3/\sqrt{10}\end{bmatrix}$$

留意到$\{\boldsymbol{u}_1,\boldsymbol{u}_2\}$已是$\mathbb{R}^2$ 的一个基，因此构造U 不需另外的向量，$U=[\boldsymbol{u}_1\ \ \boldsymbol{u}_2]$. A 的奇异值分解是

$$A=\underset{\substack{\uparrow\\ U}}{\begin{bmatrix}3/\sqrt{10} & 1/\sqrt{10}\\ 1/\sqrt{10} & -3/\sqrt{10}\end{bmatrix}}\underset{\substack{\uparrow\\ \Sigma}}{\begin{bmatrix}6\sqrt{10} & 0 & 0\\ 0 & 3\sqrt{10} & 0\end{bmatrix}}\underset{\substack{\uparrow\\ V^{\mathrm{T}}}}{\begin{bmatrix}1/3 & 2/3 & 2/3\\ -2/3 & -1/3 & 2/3\\ 2/3 & -2/3 & 1/3\end{bmatrix}}$$

■

例 4 求 $A=\begin{bmatrix}1 & -1\\ -2 & 2\\ 2 & -2\end{bmatrix}$ 的一个奇异值分解.

解 首先，计算 $A^{\mathrm{T}}A=\begin{bmatrix}9 & -9\\ -9 & 9\end{bmatrix}$. $A^{\mathrm{T}}A$ 的特征值是 18 和 0，相应的单位特征向量是

$$\boldsymbol{v}_1=\begin{bmatrix}1/\sqrt{2}\\ -1/\sqrt{2}\end{bmatrix},\ \boldsymbol{v}_2=\begin{bmatrix}1/\sqrt{2}\\ 1/\sqrt{2}\end{bmatrix}$$

这两个单位向量构成 V 的列：

$$V=[\boldsymbol{v}_1\ \ \boldsymbol{v}_2]=\begin{bmatrix}1/\sqrt{2} & 1/\sqrt{2}\\ -1/\sqrt{2} & 1/\sqrt{2}\end{bmatrix}$$

矩阵的奇异值为 $\sigma_1=\sqrt{18}=3\sqrt{2},\ \sigma_2=0$. 因为只有一个非零的奇异值，所以“矩阵” D可写成单个数值，即 $D=3\sqrt{2}$. 矩阵Σ与矩阵 A 的行列数相同，以矩阵 D 为其左上角：

$$\Sigma = \begin{bmatrix} D & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 3\sqrt{2} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

为构造 U ，首先构造 $A\boldsymbol{v}_1$和$A\boldsymbol{v}_2$：

$$A\boldsymbol{v}_1 = \begin{bmatrix} 2/\sqrt{2} \\ -4/\sqrt{2} \\ 4/\sqrt{2} \end{bmatrix}, \quad A\boldsymbol{v}_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

为检查计算，验证$\|A\boldsymbol{v}_1\| = \sigma_1 = \sqrt{18} = 3\sqrt{2}$. 当然，$A\boldsymbol{v}_2 = 0$，因为$\|A\boldsymbol{v}_2\| = \sigma_2 = 0$. 目前只找到$U$ 的一列是

$$\boldsymbol{u}_1 = \frac{1}{3\sqrt{2}} A\boldsymbol{v}_1 = \begin{bmatrix} 1/3 \\ -2/3 \\ 2/3 \end{bmatrix}$$

U 的其他列是将集合 $\{\boldsymbol{u}_1\}$ 扩充为 $\mathbb{R}^3$ 的单位正交基而得到的. 此时，我们需要两个单位正交向量 $\boldsymbol{u}_2$ 和 $\boldsymbol{u}_3$ 且都与 $\boldsymbol{u}_1$ 正交. （见图 7-15.）每个向量必须满足 $\boldsymbol{u}_1^{\mathrm{T}}\boldsymbol{x} = 0$，这等价于方程 $x_1 - 2x_2 + 2x_3 = 0$，该方程的解集构成的基是

$$\boldsymbol{w}_1 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \boldsymbol{w}_2 = \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}$$

（经检验 $\boldsymbol{w}_1$和$\boldsymbol{w}_2$都与$\boldsymbol{u}_1$正交.）应用格拉姆-施密特方法（和标准化）于$\{\boldsymbol{w}_1,\ \boldsymbol{w}_2\}$，可以得到

$$\boldsymbol{u}_2 = \begin{bmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \\ 0 \end{bmatrix}, \ \boldsymbol{u}_3 = \begin{bmatrix} -2/\sqrt{45} \\ 4/\sqrt{45} \\ 5/\sqrt{45} \end{bmatrix}$$

最后，令 $U = [\boldsymbol{u}_1 \quad \boldsymbol{u}_2 \quad \boldsymbol{u}_3]$,以及由上所得的$V^{\mathrm{T}}$和$\Sigma$，有

$$A = \begin{bmatrix} 1 & -1 \\ -2 & 2 \\ 2 & -2 \end{bmatrix} = \begin{bmatrix} 1/3 & 2/\sqrt{5} & -2/\sqrt{45} \\ -2/3 & 1/\sqrt{5} & 4/\sqrt{45} \\ 2/3 & 0 & 5/\sqrt{45} \end{bmatrix} \begin{bmatrix} 3\sqrt{2} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

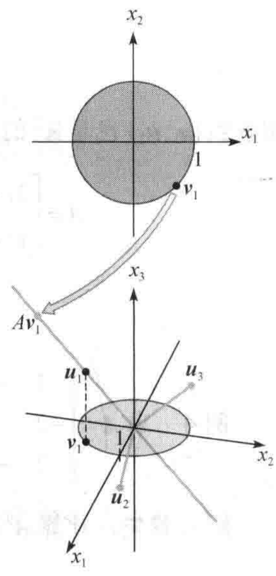

图 7-15 ■

奇异值分解的应用

如上所述，奇异值分解常用于估计矩阵的秩. 下面简单叙述它在数值计算中的其他应用，7.5 节给出一个奇异值分解在图像处理中的应用.

例 5（条件数） 当应用矩阵 A 的奇异值分解时，多数涉及方程 $A\boldsymbol{x} = \boldsymbol{b}$ 的数值计算要尽可能地可靠. 两个正交矩阵 U 和 V 不影响向量的长度和两个向量的夹角（6.2 节定理 7）. 数值计算中的任何不稳定因素都与Σ有关. 如果 A 的奇异值非常大或非常小，则舍入误差几乎不可避

免，此时，知道Σ和V中的元素对分析误差特别有用.

如果A是一个$n\times n$可逆矩阵，那么最大奇异值和最小奇异值的比率σ_1/σ_n给出了矩阵A的**条件数**. 2.3节的习题41~43表明条件数如何影响$A\boldsymbol{x}=\boldsymbol{b}$的解对$A$的元素变化（或误差）的敏感程度（事实上，$A$的“条件数”可用几种方式计算，但这里给出的定义可广泛用于$A\boldsymbol{x}=\boldsymbol{b}$的研究）. ■

例6（基本子空间的基） 给定$m\times n$矩阵A的一个奇异值分解，取$\boldsymbol{u}_1,\cdots,\boldsymbol{u}_m$是左奇异向量，$\boldsymbol{v}_1,\cdots,\boldsymbol{v}_n$是右奇异向量，且$\sigma_1,\cdots,\sigma_n$是奇异值，$r$为$A$的秩. 由定理9，

$$\{\boldsymbol{u}_1,\cdots,\boldsymbol{u}_r\} \tag{5}$$

是$\mathrm{Col}\,A$的一个单位正交基. 见图7-16.

回忆6.1节的定理3，$(\mathrm{Col}\,A)^{\perp}=\mathrm{Nul}\,A^{\mathrm{T}}$，因此

$$\{\boldsymbol{u}_{r+1},\cdots,\boldsymbol{u}_m\} \tag{6}$$

是$\mathrm{Nul}\,A^{\mathrm{T}}$的一个单位正交基.

由于当$1\leqslant i\leqslant n$时$\|A\boldsymbol{v}_i\|=\sigma_i$，且$\sigma_i$是零的充分必要条件是$i>r$，因此向量$\boldsymbol{v}_{r+1},\cdots,\boldsymbol{v}_n$生成一个维数为$n-r$的子空间$\mathrm{Nul}\,A$. 由秩定理可知，$\dim\mathrm{Nul}\,A=n-\mathrm{rank}\,A$，从而说明

$$\{\boldsymbol{v}_{r+1},\cdots,\boldsymbol{v}_n\} \tag{7}$$

是$\mathrm{Nul}\,A$的一个单位正交基，见4.5节的基定理.

从（5）和（6）可知，空间$\mathrm{Nul}\,A^{\mathrm{T}}$的正交补是$\mathrm{Col}\,A$. 将$A$和$A^{\mathrm{T}}$交换，我们有

$$(\mathrm{Nul}\,A)^{\perp}=\mathrm{Col}\,A^{\mathrm{T}}=\mathrm{Row}\,A$$

因此，从（7）得

$$\{\boldsymbol{v}_1,\cdots,\boldsymbol{v}_r\} \tag{8}$$

是$\mathrm{Row}\,A$的一个单位正交基.

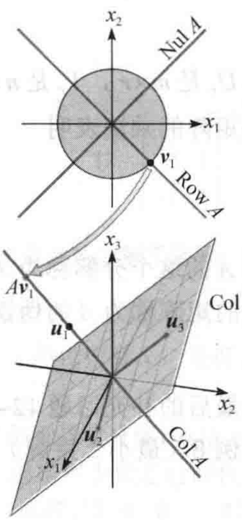

图7-16 例6中的基本子空间

图7-17给出（5）~（8）的总结，但说明的是$\{\sigma_1\boldsymbol{u}_1,\cdots,\sigma_r\boldsymbol{u}_r\}$为$\mathrm{Col}\,A$的正交基，而不是标准基. 注意到对$1\leqslant i\leqslant r$，$A\boldsymbol{v}_i=\sigma_i\boldsymbol{u}_i$. 由$A$确定的四个基本子空间的单位正交基在某些计算中非常有用，特别是在条件优化问题中更是如此.

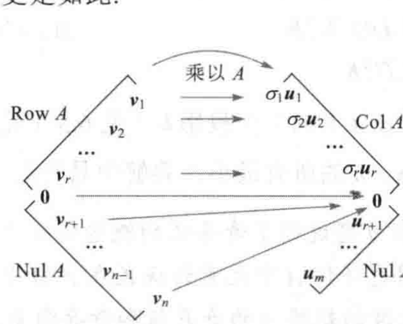

图7-17 四个基本子空间和A的作用 ■

四个基本子空间和奇异值的概念给可逆矩阵定理提供了最后的命题.（回忆为了避免几乎将可逆矩阵定理的命题数目翻倍，定理中已将与 A^{T} 有关的命题删除了.）可逆矩阵定理的其他命题在 2.3 节、2.9 节、3.2 节、4.6 节和 5.2 节中已经给出.

定理（可逆矩阵定理（最后补充））

设 A 为 $n\times n$ 矩阵，那么下述命题中的每一个都与 A 是可逆矩阵的命题等价.

u. $(\mathrm{Col}\,A)^{\perp}=\{\mathbf{0}\}$.

v. $(\mathrm{Nul}\,A)^{\perp}=\mathbb{R}^n$.

w. $\mathrm{Row}\,A=\mathbb{R}^n$.

x. A 有 n 个非零的奇异值.

例 7（奇异值分解的简化和 A 的伪逆） 当Σ包含零元素的行或列时，矩阵 A 具有更简洁的分解. 利用上面建立的符号，取 $r=\mathrm{rank}\,A$，将U和V 矩阵分块为第一块包含 r 列的子矩阵：

$$U=[U_r \quad U_{m-r}], \quad 其中U_r=[\boldsymbol{u}_1 \quad \cdots \quad \boldsymbol{u}_r]$$
$$V=[V_r \quad V_{n-r}], \quad 其中\ V_r=[\boldsymbol{v}_1 \quad \cdots \quad \boldsymbol{v}_r]$$

那么U_r是 $m\times r$，V_r是 $n\times r$.（为简化符号，我们考虑U_{m-r}和V_{n-r}，即使其中一个没有任何列.）分块矩阵的乘法表明

$$A=[U_r \quad U_{m-r}]\begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} V_r^{\mathrm{T}} \\ V_{n-r}^{\mathrm{T}} \end{bmatrix}=U_rDV_r^{\mathrm{T}} \tag{9}$$

矩阵 A 的这个分解称为 A 的**简化的奇异值分解**. 由于 D 的对角线元素非零，因此 D 是可逆矩阵. 下面的矩阵称为 A 的**伪逆**（也称**穆尔-彭罗斯逆**）：

$$A^{+}=V_rD^{-1}U_r^{\mathrm{T}} \tag{10}$$

本章最后的补充习题 12~14 研究了简化的奇异值分解和伪逆的一些性质. ■

例 8（最小二乘解） 给定方程 $A\boldsymbol{x}=\boldsymbol{b}$，利用式（10）中给出的 A 的伪逆，定义

$$\hat{\boldsymbol{x}}=A^{+}\boldsymbol{b}=V_rD^{-1}U_r^{\mathrm{T}}\boldsymbol{b}$$

那么由（9）式中的奇异值分解得到

$$\begin{aligned} A\hat{\boldsymbol{x}} &= (U_rDV_r^{\mathrm{T}})(V_rD^{-1}U_r^{\mathrm{T}}\boldsymbol{b}) \\ &= U_rDD^{-1}U_r^{\mathrm{T}}\boldsymbol{b} && 由于\ V_r^{\mathrm{T}}V_r=I_r \\ &= U_rU_r^{\mathrm{T}}\boldsymbol{b} \end{aligned}$$

从（5）式可知，$U_rU_r^{\mathrm{T}}\boldsymbol{b}$ 是 $\boldsymbol{b}$ 在 $\mathrm{Col}\,A$ 上的正交投影 $\hat{\boldsymbol{b}}$（见 6.3 节定理 10），因此 $\hat{\boldsymbol{x}}$ 是 $A\boldsymbol{x}=\boldsymbol{b}$ 的最小二乘解. 实际上，这个 $\hat{\boldsymbol{x}}$ 在 $A\boldsymbol{x}=\boldsymbol{b}$ 的所有最小二乘解中具有最小长度，见补充习题 14. ■

数值计算的注解 例 1~4 和习题说明了奇异值的概念并给出如何手工计算. 实际上，应该避免 $A^{\mathrm{T}}A$ 的计算，原因是任何 A 中元素的误差在 $A^{\mathrm{T}}A$ 中被平方. 存在快速的迭代方法，可计算精确到很多位数的矩阵 A 的奇异值和奇异向量.

进一步阅读

Horn, Roger A., and Charles R. Johnson, *Matrix Analysis*[⊖], vol. 1(Cambridge: Cambridge University Press, 1985), pp. 414–445.

Long, Cliff, "Visualization of Matrix Singular Value Decomposition." *Mathematics Magazine* **56** (1983), pp. 161–167.

Moler, C. B., and D. Morrison, "Singular Value Analysis of Cryptograms." *Amer. Math. Monthly* **90** (1983), pp. 78–87.

Strang, Gilbert, *Linear Algebra and Its Applications,* 3rd ed. (San Diego: Harcourt Brace Jovanovich, 1988), pp. 442–452.

Watkins, David S., *Fundamentals of Matrix Computations* (New York: Wiley, 1991), pp. 390–398, 409–421.

练习题

1. 给定一个奇异值分解 $A=U\Sigma V^{\mathrm{T}}$，求 A^{T} 的一个奇异值分解. A 和 A^{T} 的奇异值之间有什么联系？
2. 对任意的 $n\times n$ 矩阵 A，利用奇异值分解证明存在一个 $n\times n$ 的正交阵 Q，使得 $A^{\mathrm{T}}A=Q^{\mathrm{T}}(A^{\mathrm{T}}A)Q$.

注：练习题 2 表明对任意的 $n\times n$ 矩阵 A，矩阵 AA^{T} 和 $A^{\mathrm{T}}A$ 的正交相似的.

习题 7.4

求习题 1~4 中矩阵的奇异值分解.

1. $\begin{bmatrix}1 & 0\\ 0 & -3\end{bmatrix}$
2. $\begin{bmatrix}-3 & 0\\ 0 & 0\end{bmatrix}$
3. $\begin{bmatrix}2 & 3\\ 0 & 2\end{bmatrix}$
4. $\begin{bmatrix}3 & 0\\ 8 & 3\end{bmatrix}$

求习题 5~12 中每个矩阵的奇异值分解. （提示：在习题 11 中，U 的一个选择可以是 $\begin{bmatrix}-1/3 & 2/3 & 2/3\\ 2/3 & -1/3 & 2/3\\ 2/3 & 2/3 & -1/3\end{bmatrix}$，习题 12 中 U 的一个列可以是 $\begin{bmatrix}1/\sqrt{6}\\ -2/\sqrt{6}\\ 1/\sqrt{6}\end{bmatrix}$.）

5. $\begin{bmatrix}-2 & 0\\ 0 & 0\end{bmatrix}$
6. $\begin{bmatrix}-3 & 0\\ 0 & -2\end{bmatrix}$
7. $\begin{bmatrix}2 & -1\\ 2 & 2\end{bmatrix}$
8. $\begin{bmatrix}4 & 6\\ 0 & 4\end{bmatrix}$
9. $\begin{bmatrix}3 & -3\\ 0 & 0\\ 1 & 1\end{bmatrix}$
10. $\begin{bmatrix}7 & 1\\ 5 & 5\\ 0 & 0\end{bmatrix}$
11. $\begin{bmatrix}-3 & 1\\ 6 & -2\\ 6 & -2\end{bmatrix}$
12. $\begin{bmatrix}1 & 1\\ 0 & 1\\ -1 & 1\end{bmatrix}$
13. 求 $A=\begin{bmatrix}3 & 2 & 2\\ 2 & 3 & -2\end{bmatrix}$ 的奇异值分解（提示：对 A^{T} 进行分解）.
14. 在习题 7 中，求一个单位向量 $\boldsymbol{x}$，使得 $A\boldsymbol{x}$ 具有最大长度.
15. 假设下面是矩阵 A 的奇异值分解，其中 U 和 V 的元素是四舍五入精确到小数点后两位数字.
$$A=\begin{bmatrix}0.40 & -0.78 & 0.47\\ 0.37 & -0.33 & -0.87\\ -0.84 & -0.52 & -0.16\end{bmatrix}\begin{bmatrix}7.10 & 0 & 0\\ 0 & 3.10 & 0\\ 0 & 0 & 0\end{bmatrix}\times\begin{bmatrix}0.30 & -0.51 & -0.81\\ 0.76 & 0.64 & -0.12\\ 0.58 & -0.58 & 0.58\end{bmatrix}$$
    a. A 的秩是多少？
    b. 利用 A 的这个分解，不用计算，写出 $\mathrm{Col}\,A$ 的一个基和 $\mathrm{Nul}\,A$ 的一个基.（提示：首先写出 V 的列.）
16. 对下列 3×4 矩阵 A 的奇异值分解，重做习题 15.

⊖ 该书中文版已由机械工业出版社引进出版. ——编辑注