1 矩阵的秩

Definition 1 (矩阵的秩)**.** $\mathbf{A} \in \mathbb{R}^{m\times n}, \mathsf{r}(\mathbf{A})$ 是 $\mathbf{A}$ 的秩，下列定义等价：

1. $\mathsf{r}(\mathbf{A})$ 是 $\mathbf{A}$ 的列向量组的秩（极大线性无关组的向量个数）。
2. $\mathsf{r}(\mathbf{A})$ 是 $\mathbf{A}$ 的列空间的维数。
3. $\mathsf{r}(\mathbf{A})$ 是 $\mathbf{A}$ 的行向量组的秩，也是行空间的维数。
4. $\mathsf{r}(\mathbf{A})$ 是 $\mathbf{A}$ 的非零子式的最高阶数。
5. $\mathsf{r}(\mathbf{A}) = n - \mathsf{dim}(\mathsf{null}(\mathbf{A}))$。

Theorem 1 (秩的性质)**.** *设* $\mathbf{A} \in \mathbb{R}^{m\times n}, \mathbf{B} \in \mathbb{R}^{n\times p}$*, 则有下列性质：*

1. 初等变换不改变矩阵的秩，因此与可逆矩阵相乘不改变矩阵的秩。

2. $\mathsf{r}(\mathbf{A}) \leq \min(m, n)$。

3. $\mathsf{r}(\mathbf{A}) = \mathsf{r}(\mathbf{A}^\top)$。

4. $\mathsf{r}(\mathbf{AB}) \leq \min(\mathsf{r}(\mathbf{A}), \mathsf{r}(\mathbf{B}))$。

5. $\mathsf{r}(\mathbf{AB}) \geq \mathsf{r}(\mathbf{A}) + \mathsf{r}(\mathbf{B}) - n$。

6. $\mathsf{r}(\mathbf{AA}^\top) = \mathsf{r}(\mathbf{A}^\top\mathbf{A}) = \mathsf{r}(\mathbf{A})$.

7.（看列空间） $\mathsf{r}\left(\begin{bmatrix}\mathbf{A} & \mathbf{AB}\end{bmatrix}\right) = \mathsf{r}(\mathbf{A})$。

8.（看行空间） $\mathsf{r}\left(\begin{bmatrix}\mathbf{A} \\ \mathbf{BA}\end{bmatrix}\right) = \mathsf{r}(\mathbf{A})$。

9. $\mathsf{r}\left(\begin{bmatrix}\mathbf{A} & 0 \\ 0 & \mathbf{B}\end{bmatrix}\right) = \mathsf{r}(\mathbf{A}) + \mathsf{r}(\mathbf{B})$。

10. $\mathbf{A}$ *是方阵时，设* $\mathbf{A}^*$ *是* $\mathbf{A}$ *的伴随矩阵（即* $\mathbf{AA}^* = |\mathbf{A}|\mathbf{E}$*），则有*

$$\mathsf{r}(\mathbf{A}^*) = \begin{cases} n & 若\mathsf{r}(\mathbf{A}) = n \\ 1 & 若\mathsf{r}(\mathbf{A}) = n-1 \\ 0 & 若\mathsf{r}(\mathbf{A}) < n-1 \end{cases}$$

一些证明如下：

Definition 2 (零空间)**.** 设 $\mathbf{A} \in \mathbb{R}^{m\times n}$, 矩阵 $\mathbf{A}$ 的零空间定义为 $\mathsf{null}(\mathbf{A}) = \{\mathbf{x} \in \mathbb{R}^n | \mathbf{Ax} = \mathbf{0}\}$. 其构成一个 $\mathbb{R}$ 上的线性空间。

Definition 3 (列空间)**.** 设 $\mathbf{A} \in \mathbb{R}^{m\times n}$, 矩阵 $\mathbf{A}$ 的列空间即列向量张成的线性空间，也即 $\mathbf{A}$ 作为 $\mathbb{R}^n \to \mathbb{R}^m$ 的线性变换时的值域空间，定义为 $\mathsf{R}(\mathbf{A}) = \{\mathbf{x} \in \mathbb{R}^m | \mathbf{x} = \mathbf{Ay}, \mathbf{y} \in \mathbb{R}^n\}$. 其构成一个 $\mathbb{R}$ 上的线性空间。

Theorem 2 (线性代数基本定理（Rank-nullity theorem）**.** 设 $\mathbf{A} \in \mathbb{R}^{m\times n}$, 则有：

$$\mathsf{dim}(\mathsf{null}(\mathbf{A})) + \mathsf{dim}(\mathsf{R}(\mathbf{A})) = n.$$

证明. 设 $\mathsf{dim}_{\mathrm{A}}$ 是 $\mathsf{null}(\mathbf{A})$ 的维数，我们取 $\mathsf{null}(\mathbf{A})$ 的一组基 $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_{\mathsf{dim}_{\mathrm{A}}}$，将其补充为 $\mathbb{R}^n$ 的一组基 $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_{\mathsf{dim}_{\mathrm{A}}}, \mathbf{v}_{\mathsf{dim}_{\mathrm{A}}+1}, \cdots, \mathbf{v}_n$。

则 $\mathbf{A}\mathbf{v}_{\mathsf{dim}_{\mathrm{A}}+1}, \mathbf{A}\mathbf{v}_{\mathsf{dim}_{\mathrm{A}}+2}, \cdots, \mathbf{A}\mathbf{v}_n$ 线性无关，且 $\mathbf{A}\mathbf{v}_{\mathsf{dim}_{\mathrm{A}}+1}, \mathbf{A}\mathbf{v}_{\mathsf{dim}_{\mathrm{A}}+2}, \cdots, \mathbf{A}\mathbf{v}_n$ 可以生成 $\mathsf{R}(\mathbf{A})$。故 $\mathsf{dim}(\mathsf{R}(\mathbf{A})) = n - \mathsf{dim}(\mathsf{null}(\mathbf{A}))$, 故结论成立。 □

Corollary 1. 设 $\mathbf{A} \in \mathbb{R}^{m\times n}, \mathbf{B} \in \mathbb{R}^{n\times p}$, 有 $\mathsf{r}(\mathbf{AB}) \geq \mathsf{r}(\mathbf{A}) + \mathsf{r}(\mathbf{B}) - n$.

证明. 设 $\mathsf{dim}_{\mathrm{B}}$ 是 $\mathsf{null}(\mathbf{B})$ 的维数，$\mathsf{dim}_{\mathrm{A}}$ 是 $\mathsf{null}(\mathbf{A})$ 的维数，$\mathsf{dim}_{\mathrm{AB}}$ 是 $\mathsf{null}(\mathbf{AB})$ 的维数。

设 $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_{\mathsf{dim}_{\mathrm{B}}} \in \mathbb{R}^p$ 是 $\mathsf{null}(\mathbf{B})$ 的一个极大线性无关组，则容易验证 $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_{\mathsf{dim}_{\mathrm{B}}} \in \mathsf{null}(\mathbf{AB})$。

我们可以将 $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_{\mathsf{dim}_{\mathrm{B}}} \in \mathsf{null}(\mathbf{AB})$ 补充为 $\mathsf{null}(\mathbf{AB})$ 中的一个极大线性无关组，设 $\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_{\mathsf{dim}_{\mathrm{AB}}-\mathsf{dim}_{\mathrm{B}}}$ 是补充的向量。则其满足：

$\mathrm{B}\mathbf{w}_1, \mathrm{B}\mathbf{w}_2, \cdots, \mathrm{B}\mathbf{w}_{\mathsf{dim}_{\mathrm{AB}}-\mathsf{dim}_{\mathrm{B}}}$ 线性无关，且 $\mathrm{B}\mathbf{w}_1, \mathrm{B}\mathbf{w}_2, \cdots, \mathrm{B}\mathbf{w}_{\mathsf{dim}_{\mathrm{AB}}-\mathsf{dim}_{\mathrm{B}}} \in \mathsf{null}(\mathbf{A})$。

故 $\mathsf{dim}_{\mathrm{A}} \geq \mathsf{dim}_{\mathrm{AB}} - \mathsf{dim}_{\mathrm{B}}$，即 $n - \mathsf{dim}_{\mathrm{A}} \leq n - \mathsf{dim}_{\mathrm{AB}} + \mathsf{dim}_{\mathrm{B}}$，即 $\mathsf{r}(\mathbf{A}) + \mathsf{r}(\mathbf{B}) - n \leq \mathsf{r}(\mathbf{AB})$。

□

Corollary 2. 设 $\mathbf{A} \in \mathbb{R}^{m\times n}$, 有 $\mathsf{r}(\mathbf{A}^\top\mathbf{A}) = \mathsf{r}(\mathbf{A})$.

证明. 事实上，$\mathsf{null}(\mathbf{A}^\top\mathbf{A}) = \mathsf{null}(\mathbf{A})$。一方面，$\mathsf{null}(\mathbf{A}) \subset \mathsf{null}(\mathbf{A}^\top\mathbf{A})$ 是显然的。另一方面，若 $\mathbf{x} \in \mathsf{null}(\mathbf{A}^\top\mathbf{A})$，则 $\mathbf{A}^\top\mathbf{A}\mathbf{x} = \mathbf{0}$，即 $\mathbf{x}^\top\mathbf{A}^\top\mathbf{A}\mathbf{x} = 0$，即 $\mathbf{A}\mathbf{x} = \mathbf{0}$，即 $\mathbf{x} \in \mathsf{null}(\mathbf{A})$。故 $\mathsf{null}(\mathbf{A}^\top\mathbf{A}) = \mathsf{null}(\mathbf{A})$，故 $\mathsf{r}(\mathbf{A}^\top\mathbf{A}) = \mathsf{r}(\mathbf{A})$。

□

Theorem 3. $\mathbf{A}$ 是 n 阶方阵时，设 $\mathbf{A}^*$ 是 $\mathbf{A}$ 的伴随矩阵（即 $\mathbf{A}\mathbf{A}^* = |\mathbf{A}|\mathbf{E}$），则有

$$\mathsf{r}(\mathbf{A}^*) = \begin{cases} n & \text{若}\mathsf{r}(\mathbf{A}) = n \\ 1 & \text{若}\mathsf{r}(\mathbf{A}) = n-1 \\ 0 & \text{若}\mathsf{r}(\mathbf{A}) < n-1 \end{cases}$$

证明. 当 $\mathsf{r}(\mathbf{A}) = n$ 时，$\mathbf{A}$ 可逆，故 $\mathbf{A}^* = |\mathbf{A}|\mathbf{A}^{-1}$ 可逆，故 $\mathsf{r}(\mathbf{A}^*) = n$。

当 $\mathsf{r}(\mathbf{A}) = n-1$ 时，一方面，$\mathbf{A}$ 有 $n-1$ 阶非零主子式，故 $\mathbf{A}^*$ 是非零矩阵，故 $\mathsf{r}(\mathbf{A}^*) \geq 1$。另一方面，由 $\mathbf{A}\mathbf{A}^* = |\mathbf{A}|\mathbf{E}$，我们有

$$0 = \mathsf{r}(\mathbf{0}) = \mathsf{r}(|\mathbf{A}|\mathbf{E}) = \mathsf{r}(\mathbf{A}\mathbf{A}^*) \geq \mathsf{r}(\mathbf{A}) + \mathsf{r}(\mathbf{A}^*) - n,$$

于是 $\mathsf{r}(\mathbf{A}^*) \leq 1$，故 $\mathsf{r}(\mathbf{A}^*) = 1$。

当 $\mathsf{r}(\mathbf{A}) < n-1$ 时，$\mathbf{A}$ 的 $n-1$ 阶主子式均未 0，故 $\mathbf{A}^* = \mathbf{0}$，故 $\mathsf{r}(\mathbf{A}^*) = 0$。 □

2 秩一矩阵，对角化，奇异值分解，Householder 变换

Definition 4 (秩一矩阵 (rank one matrix)). 设 $\boldsymbol{\alpha} \in \mathbb{R}^{m\times 1}$, $\boldsymbol{\beta} \in \mathbb{R}^{n\times 1}$ 均为非零向量，矩阵 $\boldsymbol{\alpha}\boldsymbol{\beta}^\top \in \mathbb{R}^{m\times n}$ 称为秩一矩阵。

Theorem 4 (秩一矩阵的性质). *设 $\boldsymbol{\alpha} \in \mathbb{R}^{m\times 1}$, $\boldsymbol{\beta} \in \mathbb{R}^{n\times 1}$ 均为非零向量，秩一矩阵 $\boldsymbol{\alpha}\boldsymbol{\beta}^\top$ 有下列性质：*

1. $\mathsf{r}(\boldsymbol{\alpha}\boldsymbol{\beta}^\top) = 1$；
2. $(\boldsymbol{\alpha}\boldsymbol{\beta}^\top)^k = (\boldsymbol{\alpha}^\top\boldsymbol{\beta})^{k-1} \cdot \boldsymbol{\alpha}\boldsymbol{\beta}^\top$；
3. $\boldsymbol{\alpha}\boldsymbol{\beta}^\top$ 的列空间（作为线性变换的值域）是 $\mathbb{R}^m$ 的一维子空间，其列向量均为 $\boldsymbol{\alpha}$ 的倍数；
4. $\boldsymbol{\alpha}\boldsymbol{\beta}^\top$ 的行空间是 $\mathbb{R}^n$ 的一维子空间，其行向量均为 $\boldsymbol{\beta}$ 的倍数；
5. $\boldsymbol{\alpha}\boldsymbol{\beta}^\top$ 的核空间是 $\mathbb{R}^n$ 的 $n-1$ 维子空间，也是 $\boldsymbol{\beta}$ 的正交空间，即：
$$\mathsf{Ker}(\boldsymbol{\alpha}\boldsymbol{\beta}^\top) = \{\mathbf{x} \in \mathbb{R}^n \mid \boldsymbol{\alpha}\boldsymbol{\beta}^\top\mathbf{x} = (\boldsymbol{\beta}^\top\mathbf{x})\boldsymbol{\alpha} = \mathbf{0}\} = \{\mathbf{x} \in \mathbb{R}^n \mid \boldsymbol{\beta}^\top\mathbf{x} = 0\} \tag{1}$$
6. 当 $m = n$ 时，$\boldsymbol{\alpha}\boldsymbol{\beta}^\top$ 有两个特征值：$\boldsymbol{\alpha}^\top\boldsymbol{\beta}$ 和 0，特征向量空间分别为 $\boldsymbol{\alpha}$ 张成的空间和 $\boldsymbol{\beta}$ 的正交空间。当 $\boldsymbol{\alpha}^\top\boldsymbol{\beta} = 0$ 时，退化为一个特征值 0，特征向量空间为 $\boldsymbol{\beta}$ 的正交空间；
7. 当 $m = n$ 且 $\boldsymbol{\alpha} = \boldsymbol{\beta}$ 时，$\boldsymbol{\alpha}\boldsymbol{\alpha}^\top$ 是对称矩阵，因此可以对角化。特征值是 $\|\boldsymbol{\alpha}\|_2^2$ 和 0，特征向量空间分别为 $\boldsymbol{\alpha}$ 张成的空间和 $\boldsymbol{\alpha}$ 的正交空间。

证明. 直接验证是简单的。 □

秩一矩阵的一个作用在于给出了矩阵空间的一个基表示。设 $\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_m$ 是 $\mathbb{R}^m$ 的标准正交基,$\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ 是 $\mathbb{R}^n$ 的标准正交基，则 $\mathbf{e}_i\mathbf{v}_j^\top$ 是 $\mathbb{R}^{m\times n}$ 的一组基。任意矩阵 $\mathbf{A} \in \mathbb{R}^{m\times n}$ 可以表示为 $\mathbf{A} = \sum_{i=1}^m \sum_{j=1}^n a_{ij}\mathbf{e}_i\mathbf{v}_j^\top$。

同时，秩一矩阵也揭示了对角化的作用。设 $\mathbf{A} \in \mathbb{R}^{n\times n}$，$\mathbf{A}$ 的特征向量 $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ 线性无关，$\mathbf{A}$ 可以对角化为：

设 $\mathbf{A} \in \mathbb{R}^{n\times n}$，若 $\mathbf{A}$ 的特征向量 $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ 线性无关，则 $\mathbf{A}$ 可以对角化为：

$$\begin{aligned}
\mathbf{A} &= \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^\top \\
&= \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^\top \\ \mathbf{v}_2^\top \\ \vdots \\ \mathbf{v}_n^\top \end{bmatrix} \\
&= \begin{bmatrix} \lambda_1\mathbf{v}_1 & \lambda_2\mathbf{v}_2 & \cdots & \lambda_n\mathbf{v}_n \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^\top \\ \mathbf{v}_2^\top \\ \vdots \\ \mathbf{v}_n^\top \end{bmatrix} \\
&= \sum_{i=1}^n \lambda_i\mathbf{v}_i\mathbf{v}_i^\top
\end{aligned}$$

可见如果我们不做任何努力，我们只能得到一个方阵可以分解为 n^2 个秩一矩阵，但是可对角化的矩阵 $\mathbf{A}$ 可以表示成 n 个对称的秩一矩阵的和。对于一般的矩阵，我们也能写成一些更少的秩一矩阵的和吗？答案是肯定的，这就是奇异值分解。

Theorem 5 (奇异值分解（Single Value Decomposition，SVD）)**.** *设 $\mathbf{A} \in \mathbb{R}^{m\times n}$，则 $\mathbf{A}^\top\mathbf{A}$ 是一个对称半正定矩阵，且 $\mathsf{Ker}(\mathbf{A}^\top\mathbf{A}) = \mathsf{Ker}(\mathbf{A})$，设 $\mathsf{r}(\mathbf{A}^\top\mathbf{A}) = \mathsf{r}(\mathbf{A}) = r$，则 $\mathbf{A}^\top\mathbf{A}$ 的秩为有 r 个非零特征值，设为 $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$，对应的特征向量为 $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_r$，令 $\mathbf{v}_i = \frac{1}{\sqrt{\sigma_i}}\mathbf{A}\mathbf{u}_i, i = 1, 2, \cdots, r$，则 $\mathbf{A}$ 有如下分解:*

$$
\begin{aligned}
\mathbf{A} &= \sum_{i=1}^{r} \sqrt{\sigma_i}\mathbf{u}_i\mathbf{v}_i^\top \\
&= \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_r \end{bmatrix} \begin{bmatrix} \sqrt{\sigma_1} & 0 & \cdots & 0 \\ 0 & \sqrt{\sigma_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\sigma_r} \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^\top \\ \mathbf{v}_2^\top \\ \vdots \\ \mathbf{v}_r^\top \end{bmatrix} \\
&= \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top
\end{aligned}
$$

其中 $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_r] \in \mathbb{R}^{m\times r}$，$\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_r] \in \mathbb{R}^{n\times r}$，
$\mathbf{\Sigma} = \mathsf{diag}(\sqrt{\sigma_1}, \sqrt{\sigma_2}, \cdots, \sqrt{\sigma_r}) \in \mathbb{R}^{r\times r}$。

证明. 只需验证 $\sum_{i=1}^{r} \sqrt{\sigma_i}\mathbf{u}_i\mathbf{v}_i^\top$ 和 $\mathbf{A}$ 在 $\mathbb{R}^{m\times n}$ 中的基 $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_r, \mathbf{u}_{r+1}, \cdots, \mathbf{u}_m$ 上一致即可，这里 $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_r$ 是上述定义的 $\mathbf{A}^\top\mathbf{A}$ 的特征向量，$\mathbf{u}_{r+1}, \cdots, \mathbf{u}_m$ 是为构成 $\mathbb{R}^m$ 的一组基而补充的向量，注意 $\mathbf{u}_{r+1}, \cdots, \mathbf{u}_m$ 实际上构成了 $\mathsf{Ker}(\mathbf{A})$ 的一组基。 □

接下来讨论 Householder 变换。

Theorem 6 (Householder 变换的性质)**.** *设 $\mathbf{x} \in \mathbb{R}^n$ 为单位向量，$\mathbf{H} = \mathbf{E} - 2\mathbf{x}\mathbf{x}^\top$，则 $\mathbf{H}$ 称为 Householder 变换，具有如下性质：*

1. *$\mathbf{H}$ 是对称矩阵，可以对角化，特征值为 1 和 -1，特征向量空间分别为 $\mathbf{x}$ 的正交空间和 $\mathbf{x}$ 张成的空间；*
2. *$\mathbf{H}$ 是正交矩阵，即 $\mathbf{H}^\top\mathbf{H} = \mathbf{H}\mathbf{H}^\top = \mathbf{E}$；*
3. *$\mathbf{H}$ 是幂等矩阵，即 $\mathbf{H}^2 = \mathbf{E}$；*
4. *$\mathbf{H}$ 可以反射 $\mathbf{x}$，即 $\mathbf{H}\mathbf{x} = -\mathbf{x}$。*

证明. 直接验证是简单的。 □