二、矩阵论（2021.11.21）

1. 行列式、矩阵的基本性质、向量的线性关系

性质 1.1：

行列互换，行列式的值不变，即 $D^T = D$

性质 1.2：
对换行列式中两行的位置，行列式变号。
推论 1.1：
若一个行列式中有两行的对应元素相同，则这个行列式为零。
性质 1.3：
把行列式某一行的所有元素同时乘以数 c，等于用数 c 乘以这个行列式。
推论 1.2：
行列式的某一行有公因子时，可以把这个公因子提到行列式符号外面。
推论 1.3：
如果行列式某两行的对应元素成比例，则这个行列式为零。
性质 1.4：
如果行列式某行的各元素都是两元素的和，则这个行列式等于两个行列式之和。
性质 1.5：
如果把行列式某一行的各元素同乘以数 k 再加到另一行的对应元素上去，则行列式的值不变。

定义 1.1：余子式与代数余子式

在 n 阶行列式 $D = \left|a_{ij}\right|_n$ 中，划去元素 a_{ij} 所在的行和列，剩下的 $(n-1)^2$ 个元素按原来的相对位置构成的 n-1 阶行列式称为行列式 D 中元素 a_{ij} 的余子式，记作 M_{ij}。而称 $(-1)^{i+j} M_{ij}$ 为 a_{ij} 的代数余子式，记为 A_{ij}。

性质 1.6：n 阶行列式 $D = \left|a_{ij}\right|_n$ 等于它的任一行的所有元素与它们各自对应的代数余子式的乘积之和。

即：$D = a_{i1}A_{i1} + a_{i2}A_{i2} + \ldots + a_{in}A_{in} (i = 1, 2 \ldots n)$，这个分解称为 D 按第 i 行的展开式。这样一个 n 阶行列式的计算就可以转换为 n 个 n-1 阶行列式的计算。

推论 1.4：记 n 阶行列式为 $D = \left|a_{ij}\right|_n$，则 $a_{i1}A_{j1} + a_{i2}A_{j2} + \ldots + a_{in}A_{jn} = \begin{cases} D & i = j \\ 0 & i \neq j \end{cases}$

简要说明一下为什么当 $i \neq j$ 时，上面这个式子等于 0。我们可以举一个实例。如

$$D=\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

假设 $i=1, j=2$。那么 $a_{11}A_{21}+a_{12}A_{22}+a_{13}A_{23}$ 这个式子可以看作是行列式 $\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{11} & a_{12} & a_{13} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$ 按照第二行展开的结果。而显然，$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{11} & a_{12} & a_{13} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$ 值为 0。

性质 1.7：设 A、B 是两个 n 阶方阵，$C=AB$，则 $|C|=|A|\cdot|B|$

性质 1.8：设 A 为方阵，k,l 为正整数，则 $A^kA^l=A^{k+l}$，$(A^k)^l=A^{kl}$

性质 1.9：设 A 为 n 阶矩阵，m 是正整数，k 是数，则 $|A^m|=|A|^m$，$|kA|=k^n|A|$

证明：第二个等式，把矩阵 A 乘以数 k，相当于把 A 中的所有元素都同时乘以数 k。而对于 n 阶行列式而言，行列式乘以 k，相当于对其中的一行乘以 k。这样，对矩阵乘以 k 相当于对行列式乘了 n 次 k。

性质 1.10：$(AB)^T=B^TA^T$

性质 1.11：A 为对称矩阵的充要条件是 $A^T=A$，A 为反对称矩阵的充要条件是 $A^T=-A$

性质 1.12 如果矩阵 A 可逆，则 $\left(A^{-1}\right)^T=\left(A^T\right)^{-1}$

证明：$A^T\left(A^{-1}\right)^T=\left(A^{-1}A\right)^T=E$，所以 $\left(A^T\right)^{-1}=\left(A^{-1}\right)^T$

性质 1.13 $|E+tA|=1+t\lambda$，$A\in\mathbb{R}^{n\times n}, rank\left(A\right)=1$，$\lambda$ 为 A 的对角元素之和。

证明：设 $A=\left[\mathbf{a}_1,\mathbf{a}_2,...,\mathbf{a}_n\right]$（注意：此 A 的任意两列都是线性相关的），则

$$\begin{aligned}
&|E+tA|=|\mathbf{e}_1+t\mathbf{a}_1,\mathbf{e}_2+t\mathbf{a}_2,\mathbf{e}_3+t\mathbf{a}_3,..,\mathbf{e}_n+t\mathbf{a}_n| \\
&=|\mathbf{e}_1,\mathbf{e}_2+t\mathbf{a}_2,\mathbf{e}_3+t\mathbf{a}_3,..,\mathbf{e}_n+t\mathbf{a}_n|+|t\mathbf{a}_1,\mathbf{e}_2+t\mathbf{a}_2,\mathbf{e}_3+t\mathbf{a}_3,...,\mathbf{e}_n+t\mathbf{a}_n| \\
&=|\mathbf{e}_1,\mathbf{e}_2+t\mathbf{a}_2,\mathbf{e}_3+t\mathbf{a}_3,...,\mathbf{e}_n+t\mathbf{a}_n|+|t\mathbf{a}_1,\mathbf{e}_2,\mathbf{e}_3,..,\mathbf{e}_n|(rank(A)=1) \\
&=|\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3+t\mathbf{a}_3,...,\mathbf{e}_n+t\mathbf{a}_n|+|\mathbf{e}_1,t\mathbf{a}_2,\mathbf{e}_3+t\mathbf{a}_3,...,\mathbf{e}_n+t\mathbf{a}_n|+|t\mathbf{a}_1,\mathbf{e}_2,\mathbf{e}_3,..,\mathbf{e}_n| \\
&=|\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3+t\mathbf{a}_3,...,\mathbf{e}_n+t\mathbf{a}_n|+|\mathbf{e}_1,t\mathbf{a}_2,\mathbf{e}_3,...,\mathbf{e}_n|+|t\mathbf{a}_1,\mathbf{e}_2,\mathbf{e}_3,..,\mathbf{e}_n| \\
&\vdots \\
&=|\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3,...,\mathbf{e}_n|+|\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3,...,t\mathbf{a}_n|+...+|\mathbf{e}_1,t\mathbf{a}_2,\mathbf{e}_3,...,\mathbf{e}_n|+|t\mathbf{a}_1,\mathbf{e}_2,\mathbf{e}_3,..,\mathbf{e}_n| \\
&=1+t\mathbf{a}_{nn}+...+t\mathbf{a}_{22}+t\mathbf{a}_{11}=1+t\lambda
\end{aligned}$$

定义 1.2：伴随矩阵

设 $A=\left(a_{ij}\right)_n$ 为 n 阶矩阵，A_{ij} 为$|A|$中元素 a_{ij} 的代数余子式，则称矩阵

$\begin{bmatrix} A_{11} & A_{21} \dots A_{n1} \\ A_{12} & A_{22} \dots A_{n2} \\ \vdots & \\ A_{1n} & A_{2n} \dots A_{nn} \end{bmatrix}$为 A 的伴随矩阵，记为 A^*

由伴随矩阵的定义可知：$AA^*=A^*A=|A|E$

我们来简要说明一下上面的这个结论，用一个具体例子来看。

$A=\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$，则 $A^*=\begin{bmatrix} A_{11} & A_{21} & A_{31} \\ A_{12} & A_{22} & A_{32} \\ A_{13} & A_{23} & A_{33} \end{bmatrix}$

而行列式可以利用代数余子式按行展开来计算，有

$$a_{i1}A_{j1}+a_{i2}A_{j2}+\dots+a_{in}A_{jn}=\begin{cases} |A|, i=j \\ 0, i\neq j \end{cases}$$

这样容易得到 $AA^*=\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}\begin{bmatrix} A_{11} & A_{21} & A_{31} \\ A_{12} & A_{22} & A_{32} \\ A_{13} & A_{23} & A_{33} \end{bmatrix}=\begin{bmatrix} |A| & 0 & 0 \\ 0 & |A| & 0 \\ 0 & 0 & |A| \end{bmatrix}=|A|E=A^*A$

定理 1.1： n 阶矩阵 A 可逆的充要条件为$|A|\neq 0$。如果 A 可逆，则 $A^{-1}=\dfrac{1}{|A|}A^*$

证明：必要性 如果 A 可逆，则 $AA^{-1}=A^{-1}A=E$，两边取行列式得$|AA^{-1}|=|A^{-1}||A|=1$

则$|A|\neq 0$

充分性 若$|A|\neq 0$则 $A(\dfrac{1}{|A|}A^*)=\dfrac{1}{|A|}AA^*=\dfrac{1}{|A|}|A|E=E$，则 A 可逆。同时已经求得 A 的逆矩阵。

可逆矩阵的基本性质

性质 1.14

设 A，B，$A_i(i=1,2,...,m)$ 为 n 阶可逆矩阵，k 为非零常数，则 A^{-1}，kA，AB，$A_1A_2...A_m$ 也都是可逆矩阵，且

（1）$(A^{-1})^{-1}=A$　（2）$(kA)^{-1}=\dfrac{1}{k}A^{-1}$　（3）$(AB)^{-1}=B^{-1}A^{-1},(A_1A_2...A_m)^{-1}=A_m^{-1}...A_2^{-1}A_1^{-1}$

（4）$\left|A^{-1}\right|=\frac{1}{|A|}$　（5）$(A^m)^{-1}=(A^{-1})^m$　（6）$\left|A^*\right|=|A|^{n-1}$

我们来证明一下第 6 个性质。
由伴随矩阵的定义知道，

$AA^*=|A|E$，两边取行列式得到，$\left|AA^*\right|=|A|\left|A^*\right|=\left||A|E\right|=|A|^n$

下面分三种情况讨论。

1) $|A|\neq 0$，即 A 可逆，则有 $\left|A^*\right|=|A|^n/|A|=|A|^{n-1}$

2) 如果 $|A|=0$，且 $A=O$，则 $A^*=O$，则结论显然成立

3) 如果 $|A|=0$，且 $A\neq O$，反设 $\left|A^*\right|\neq 0$，则 A^* 可逆，因而 $A=(AA^*)\left(A^*\right)^{-1}=\left(|A|E\right)\left(A^*\right)^{-1}=|A|\left(A^*\right)^{-1}=O$，与假设 $A\neq O$ 矛盾，所以 $\left|A^*\right|=0=|A|^{n-1}$

定义 1.3　*k* 阶子式

在矩阵 $A=\left(a_{ij}\right)_{m\times n}$ 中任取 k 行和 k 列（$1\le k\le \min\{m,n\}$），位于这 k 行和 k 列的交叉点上的 k^2 个元素，按照它们在矩阵 A 中的相对位置所组成的 k 阶行列式称为矩阵 A 的一个 k 阶子式。

定义 1.4 矩阵的秩

若矩阵 **A** 中存在一个 r 阶子式不为 0，而 **A** 中所有 r+1 阶子式都为 0，则 **A** 的秩为 r

由矩阵的秩的定义容易知道，

性质 1.15 当 **A** 为 n 阶方阵时，$r(\mathbf{A})=n$ 等价于 $|\mathbf{A}|\neq 0$ 等价于 **A** 可逆

定理 1.2 任何矩阵经初等变换后，其秩不变。

推论 1.5 设矩阵 A 为 $m\times n$ 矩阵，P 为 m 阶可逆矩阵，Q 为 n 阶可逆矩阵，则

$r(A)=r(PA)=r(AQ)=r(PAQ)$

矩阵的等价
定义 1.5 如果矩阵 A 经过有限次初等变换后变为矩阵 B，则称矩阵 A 与 B 等价。
由定义可见，
推论 1.6 矩阵 A 与 B 等价的充要条件是存在可逆矩阵 P 与可逆矩阵 Q 使得 $PAQ=B$

等价的矩阵必有相同的秩和相同的标准形。等价关系有如下性质：自反性、对称性和传递性。

定义 1.6 线性表出

对于 n 维向量 $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m, \mathbf{b}$，如果存在数 $k_1, k_2, \ldots, k_m$，

使得 $\mathbf{b} = k_1\mathbf{a}_1 + k_2\mathbf{a}_2 + \ldots + k_m\mathbf{a}_m$，则称 $\mathbf{b}$ 是 $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m$ 的一个线性组合，或称 $\mathbf{b}$ 可由

$\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m$ 线性表出。

定义 1.7 向量组的等价

有两个向量组（I）和（II）。若（I）中的向量都可以被向量组（II）线性表出，则称向量组（I）可以被向量组（II）线性表出。如果向量组（I）和（II）可以互相线性表出，则称向量组（I）与（II）等价。

关于向量组的等价有以下性质：

性质 1.16 向量组的等价具有（1）反身性；（2）对称性；（3）传递性。

定义 1.8 极大无关组

设 $\mathbf{a}_{i1}, \mathbf{a}_{i2}, \ldots, \mathbf{a}_{ir}$ 是向量组 $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m$ 的部分向量组，它满足

（1）$\mathbf{a}_{i1}, \mathbf{a}_{i2}, \ldots, \mathbf{a}_{ir}$ 线性无关

（2）向量组 $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m$ 的每一个向量都可由 $\mathbf{a}_{i1}, \mathbf{a}_{i2}, \ldots, \mathbf{a}_{ir}$ 线性表出。

则称向量组 $\mathbf{a}_{i1}, \mathbf{a}_{i2}, \ldots, \mathbf{a}_{ir}$ 是向量组 $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m$ 的一个极大线性无关组，简称极大无关组。

由极大无关组的定义容易推出以下性质。

性质 1.17 一个向量组与它的极大无关组（若有极大无关组的话）等价。一个向量组的任意两个极大无关组等价。

定义 1.9 向量组的秩

向量组 $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m$ 的极大无关组所含向量的个数称为向量组的秩。

我们知道一个向量组的秩是唯一的，则可知：

性质 1.18 一个向量组若有两个极大无关组，则它们所含向量个数相同。

线性无关的向量组其极大无关组就是它自身，因此有，

性质 1.19 向量组 $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m$ 线性无关的充要条件是 $r\left(\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m\right) = m$。

由向量组等价的传递性可知，任意两个等价的向量组的极大无关组也等价，所以有，

性质 1.20 等价的向量组必有相同的秩。

证明：设向量组 $\mathbf{X}$ 和向量组 $\mathbf{Y}$ 等价。设 $\mathbf{A}$ 是 $\mathbf{X}$ 的一个极大无关组，$\mathbf{B}$ 是 $\mathbf{Y}$ 的一个极大无关组，由等价的传递性，可以知道 $\mathbf{A}$ 也等价与 $\mathbf{B}$。因此，$\mathbf{B}$ 也是 $\mathbf{X}$ 的一个极大无关组，由性质 1.18 知道，极大无关组 $\mathbf{A}$ 与 $\mathbf{B}$ 含有的向量个数相同，因此 $r(\mathbf{X}) = r(\mathbf{Y})$。

定理 1.3 设 $\mathbf{A}$ 是 $m \times n$ 矩阵，则 $\mathbf{A}$ 的列向量组 $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ 的秩等于矩阵 $\mathbf{A}$ 的秩

2. 线性方程组的解

线性方程组可表达为：

$$A\mathbf{x}=\mathbf{b}\text{，其中 } A\in\mathbb{R}^{m\times n}\text{，}\ \mathbf{x}\in\mathbb{R}^{n\times 1}\text{，}\ \mathbf{b}\in\mathbb{R}^{m\times 1} \tag{2-1}$$

显然，这个方程组包含了 m 个方程、n 个未知数。

（2-1）有解的充要条件是：

$$r(A)=r\left[A;\mathbf{b}\right]$$

如何来理解线性方程组有解的条件？

假设 $A=\left[\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,...,\boldsymbol{\alpha}_n\right]$，

$A\mathbf{x}=\mathbf{b}$ 有解 $\Leftrightarrow \exists\mathbf{x}\left(x_1,x_2,...,x_n\right),x_1\boldsymbol{\alpha}_1+x_2\boldsymbol{\alpha}_2+...+x_n\boldsymbol{\alpha}_n=\mathbf{b}$，即 $\mathbf{b}$ 可由向量组 $\left\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,...,\boldsymbol{\alpha}_n\right\}$ 线性表出（定义 1.6）

$\Leftrightarrow \left\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,...,\boldsymbol{\alpha}_n,\mathbf{b}\right\}$ 可由向量组 $\left\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,...,\boldsymbol{\alpha}_n\right\}$ 线性表出。又由于 $\left\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,...,\boldsymbol{\alpha}_n\right\}$ 显然可以由 $\left\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,...,\boldsymbol{\alpha}_n,\mathbf{b}\right\}$ 线性表出

$\Leftrightarrow \left\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,...,\boldsymbol{\alpha}_n,\mathbf{b}\right\}$ 与 $\left\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,...,\boldsymbol{\alpha}_n\right\}$ 等价（定义 1. 7）

$\Leftrightarrow r\left\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,...,\boldsymbol{\alpha}_n,\mathbf{b}\right\}=r\left\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,...,\boldsymbol{\alpha}_n\right\}$（性质 1.20）

$\Leftrightarrow r\left[A,b\right]=r\left(A\right)$（定理 1.3）

（2-1）有唯一解的充要条件是：

$$r(A)=r\left[A;\mathbf{b}\right]=n$$

我们证明一下充分性。

假设 $r(A)=r\left[A;\mathbf{b}\right]=n$，且此时有 $\mathbf{x}_1\neq\mathbf{x}_2$，且 $A\mathbf{x}_1=\mathbf{b},A\mathbf{x}_2=\mathbf{b}$

则，$A\left(\mathbf{x}_1-\mathbf{x}_2\right)=0$，则

$\left[\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,...,\boldsymbol{\alpha}_n\right]\left(\mathbf{x}_1-\mathbf{x}_2\right)=0$，且 $\mathbf{x}_1-\mathbf{x}_2\neq\mathbf{0}$

则，$\left[\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,...,\boldsymbol{\alpha}_n\right]$ 线性相关，

则 $r\left[\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,...,\boldsymbol{\alpha}_n\right]<n$（性质 1. 19）

则，$r\left(A\right)=r\left[\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,...,\boldsymbol{\alpha}_n\right]<n$，与已知条件矛盾，所以必有 $\mathbf{x}_1=\mathbf{x}_2$。

（2-1）有无穷多组解的充要条件是：

$$r(A) = r\left[A;\mathbf{b}\right] < n$$

推论 2.1：齐次线性方程组 $A\mathbf{x}=\mathbf{0}$，总有解。若 A 为 $m\times n$ 阶矩阵，则它有非零解的充要条件是：

$$r(A) < n$$

理解：显然 $\mathbf{x}=\mathbf{0}$ 一定是它的一个解。如果 $r(A)=n$，则根据线性方程组有唯一解的条件可知，此时 $A\mathbf{x}=\mathbf{0}$ 只能有唯一解，也就是 $\mathbf{x}=\mathbf{0}$。所以，$A\mathbf{x}=\mathbf{0}$ 要想有非零解，一定要求 $r(A)<n$。另一方面，如果 $r(A)<n$，则 $A\mathbf{x}=\mathbf{0}$ 有无穷多组解，当然肯定有非零解。

推论 2.2：若齐次线性方程组 $A\mathbf{x}=\mathbf{0}$ 中方程的个数小于未知量的个数，即 $m<n$，则它必有非零解。若 $m=n$，则它有非零解的充要条件是 $|A|=0$

理解：第一部分是显然的。对于第二个结论，根据推论 2.1，

$A\mathbf{x}=\mathbf{0}$ 有非零解 $\Leftrightarrow r(A)<n \Leftrightarrow$ 当 A 是方阵时，$|A|=0$（性质 1.15）

推论 2.3：齐次线性方程组 $A\mathbf{x}=\mathbf{0}$（A 为 n 阶方阵）的系数行列式 $|A|\neq 0$，则它只有零解。

证明：$\mathbf{x}=\mathbf{0}$ 一定是线性方程组 $A\mathbf{x}=\mathbf{0}$ 一个解。当 $|A|\neq 0$ 时，也就是 $r(A)=n$；由线性方程组解理论知道，这时 $A\mathbf{x}=\mathbf{0}$ 有唯一解，当然只能是 $\mathbf{x}=\mathbf{0}$。

线性方程组解的结构

[1] 若 γ_1,γ_2 是方程组 $A\mathbf{x}=\mathbf{b}$ 的任意两个解向量，则 $\gamma_1-\gamma_2$ 是对应的齐次线性方程组 $A\mathbf{x}=\mathbf{0}$ 的解向量

[2] 设 γ_0 是 $A\mathbf{x}=\mathbf{b}$ 的一个解向量，η 是对应的齐次线性方程组的任一解向量，则 $\gamma_0+\eta$ 仍是 $A\mathbf{x}=\mathbf{b}$ 的一个解向量

[3] $A\mathbf{x}=\mathbf{b}$ 的任一解向量 γ 都可以表示成 $\gamma=\gamma_0+\eta$，其中 γ_0 是 $A\mathbf{x}=\mathbf{b}$ 的某个解向量，η 是对应的齐次线性方程组的某个解向量。

证明：

证明[1]式

$A(\gamma_1-\gamma_2)=A\gamma_1-A\gamma_2=\mathbf{b}-\mathbf{b}=\mathbf{0}$，则 $\gamma_1-\gamma_2$ 是 $A\mathbf{x}=\mathbf{0}$ 的一个解向量。

证明[2]式

$A(\gamma_0+\eta)=A\gamma_0+A\eta=\mathbf{b}+0=\mathbf{b}$，则 $\gamma_0+\eta$ 是 $A\mathbf{x}=\mathbf{b}$ 的一个解向量。

证明[3]式

根据题设我们知道，γ 和 γ_0 都是 $A\mathbf{x}=\mathbf{b}$ 的解向量。因此，由[1]可知 $\gamma-\gamma_0$ 是对应的齐次线性方程组 $A\mathbf{x}=\mathbf{0}$ 的解向量。且，$\gamma=\gamma_0+(\gamma-\gamma_0)$，我们记 $\eta\equiv\gamma-\gamma_0$，则 $\gamma=\gamma_0+\eta$。

设有齐次线性方程组 $A\mathbf{x}=\mathbf{0}$，其中 A 是 $m\times n$ 矩阵，设 $\eta_1,\eta_2\ldots,\eta_s$ 是方程组的 s 个解向量，易知它们的线性组合 $\sum_{i=1}^{s}k_i\eta_i$ 仍是原方程组的解，其中 $k_1,k_2\ldots,k_s$ 为任意常数。所以，当齐次线性方程组有非零解时，它必有无穷多组解。然而，齐次线性方程组的解是 n 维向量，它们当中的线性无关的解向量的个数不可能多于 n。因此，必然存在一个与全体解向量等价的线性无关的向量组，使得任何一个解向量都能由它们线性表出。这样的一个线性无关的向量组称为 $A\mathbf{x}=\mathbf{0}$ 的一个基础解系。

定理 2.1：设 A 是 $m\times n$ 矩阵，若 $r(A)<n$，则齐次线性方程组 $A\mathbf{x}=\mathbf{0}$ 存在一个由 *n-r*（*r* 是系数矩阵 A 的秩）个线性无关的解向量 $\eta_1,\eta_2\ldots,\eta_{n-r}$ 构成的基础解系，它们的线性组合 $\tilde{\eta}=k_1\eta_1+k_2\eta_2+\ldots+k_{n-r}\eta_{n-r}$，其中 $k_1,k_2\ldots,k_{n-r}$ 为任意常数，给出了 $A\mathbf{x}=\mathbf{0}$ 的所有解。

注：基础解系是线性无关的，意味着里面不包含零向量。实际上，$A\mathbf{x}=\mathbf{0}$ 的所有解向量构成了一个线性空间，这个线性空间的维度为 *n-r*，且基础解系 $\eta_1,\eta_2\ldots,\eta_{n-r}$ 构成了这个线性空间的一组基。

定理 2.2：若非齐次线性方程组 $A\mathbf{x}=\mathbf{b}$ 满足 $r(A)=r(A\ \mathbf{b})=r$，$\eta_1=\eta_2=\ldots=\eta_{n-r}$ 是对应的齐次线性方程组 $A\mathbf{x}=\mathbf{0}$ 的基础解系，γ_0 是 $A\mathbf{x}=\mathbf{b}$ 的某个解，

则：

$$\mathbf{x}=\gamma_0+k_1\eta_1+k_2\eta_2+\ldots+k_{n-r}\eta_{n-r}$$

给出了 $A\mathbf{x}=\mathbf{b}$ 的所有解（通解），其中 $k_1,k_2,\ldots,k_{n-r}$ 是任意常数。称 γ_0 是 $A\mathbf{x}=\mathbf{b}$ 的一个特解。

克莱姆（Cramer）法则

方程的个数和未知量的个数均为 n，即方程的形式为：

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + ... + a_{12}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n = b_2 \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 + ... + a_{nn}x_n = b_n \end{cases}$$

其有唯一解的充要条件为：系数行列式 $D \neq 0$，且其解为：

$$x_1 = \frac{D_1}{D}, x_2 = \frac{D_2}{D}, ...x_n = \frac{D_n}{D}$$

其中，D 为方程组的系数行列式，D_j 是把 D 的第 j 列各元素依次换成方程组右端的常数项所得到的行列式。

证明：把这个方程组写成标准形式 $A\mathbf{x} = \mathbf{b}, A \in \mathbb{R}^{n\times n}, \mathbf{b} \in \mathbb{R}^{n\times 1}$。

$|A| \neq 0 \Leftrightarrow r(A) = n$（性质 1.15）；并且，由于 A 中的列向量都是 n 维向量，则 A 中的列向量构成了 n 维线性空间的一组基，当然 $\mathbf{b}$ 可以由其线性表出，则 $r(A\ \mathbf{b}) = r(A)$，则我们有 $r(A) = r(A\ \mathbf{b}) = n$，则由线性方程组解理论可知，该方程组有唯一解。

3. 特征值分解

定义 3.1 特征值与特征向量

如果存在数λ与非零向量$\boldsymbol{\alpha}$，使得

$A\boldsymbol{\alpha}=\lambda\boldsymbol{\alpha}$

则称λ为$A\in\mathbb{R}^{n\times n}$的特征值，$\boldsymbol{\alpha}$为对应于特征值$\lambda$的特征向量。

注意：特征值λ是可以为 0 的，但特征向量必须是非零向量。

由定义可知，如果λ_0是A的特征值，$\boldsymbol{\alpha}_0$是对应的特征向量，则$\boldsymbol{\alpha}_0$就是齐次线性方程组

$(\lambda_0 E-A)\mathbf{x}=0$

的非零解。

若能找到λ_0使得齐次线性方程组$(\lambda_0 E-A)\mathbf{x}=0$有非零解，则$\lambda_0$是$A$的一个特征值，这个方程组的任何一个非零解都是$\lambda_0$对应的特征向量。因而，由齐次线性方程组解的理论可知：

（1）λ_0是矩阵A的特征值的充要条件为$|\lambda_0 E-A|=0$

（2）若$\alpha_1,\alpha_2,\ldots\alpha_s$都是矩阵$A$对应于特征值$\lambda_0$的特征向量，$k_1,k_2,\ldots k_s$为数，且$k_1\alpha_1+k_2\alpha_2+\ldots k_s\alpha_s\neq 0$，则$\sum_{i=1}^{s}k_i\alpha_i$也是矩阵$A$对应于特征值$\lambda_0$的特征向量。因而$n$阶矩阵$A$对应于特征值$\lambda_0$的全体特征向量**再添上零向量**构成一个线性空间，它是齐次线性方程组$(\lambda_0 E-A)\mathbf{x}=0$的解空间，称之为矩阵$A$对应于特征值$\lambda_0$的**特征子空间**，记作$V_{\lambda_0}$。$V_{\lambda_0}$的维数当然就是与$\lambda_0$对应的线性无关的特征向量的个数。

注意：我们平时经常说，与这个特征值对应的特征向量是某某，感觉上好像是与特征值对应的特征向量只有一个似的。实际上，务必要清楚，与特征值λ_0对应的特征向量加上零向量就构成了一个子空间，子空间中任意向量的线性组合（倍加与数乘）的结果，只要不是零向量，便都是与λ_0对应的特征向量，所以**与某个特征值对应的特征向量根本谈不上唯一性**。

关于矩阵的特征值的个数问题

引理 3.1：（代数基本定理）Every polynomial of degree n has exactly n (real and) complex roots, counted with multiplicity. Eqivalently, if $f(x)=x^n+a_{n-1}x^{n-1}+a_{n-2}x^{n-2}+\ldots+a_1x+a_0$ is a polynomial of degree n then f factors as $f(x)=(x-\lambda_n)(x-\lambda_{n-1})\ldots(x-\lambda_1)$ for (not nessarily distinct) complex numbers $\lambda_1 \ldots, \lambda_n$。

n 阶方阵 A 的特征方程为$|\lambda E - A|=0$，是关于 λ 的 n 阶多项式。根据引理 3.1，在复数域中，它会有 n 个根（k 重根算作 k 个根）。因此，n 阶方阵 A 的特征值有 n 个。

引理 3.2：如果 n 阶方阵 A 是个实矩阵，则其特征多项式 f 的系数都是实数。在这种情况下，它的复数特征值必然会以共轭复数的形式成对出现。

证明：如果 λ 是 A 的一个复数特征值，根据特征值定义，λ 是 f 的一个复数根，也就是说，

$$f(\lambda) = a_n\lambda^n + a_{n-1}\lambda^{n-1} + \ldots + a_1\lambda + a_0 = 0$$

则，

$$\begin{aligned}
0 &= \overline{f(\lambda)} \\
&= \overline{a_n\lambda^n + a_{n-1}\lambda^{n-1} + \ldots + a_1\lambda + a_0} \\
&= a_n\overline{\lambda}^n + a_{n-1}\overline{\lambda}^{n-1} + \ldots + a_1\overline{\lambda} + a_0 = f\left(\overline{\lambda}\right)
\end{aligned}$$

即，$\overline{\lambda}$ 也是 f 的一个复数根，则 $\overline{\lambda}$ 也是 A 的一个复数特征值。

我们也可以从另外一个角度来证明，

$A\mathbf{v} = \lambda\mathbf{v} \Rightarrow \overline{A\mathbf{v}} = \overline{\lambda\mathbf{v}} \Rightarrow \overline{A}\overline{\mathbf{v}} = \overline{\lambda}\overline{\mathbf{v}} \Rightarrow A\overline{\mathbf{v}} = \overline{\lambda}\overline{\mathbf{v}}$，则 $\overline{\lambda}$ 也是 A 的一个复数特征值。

定理 3.1：设 λ_0 为 n 阶矩阵 A 的 k 重特征值，则其特征子空间的维数 $\dim V_{\lambda_0} \le k$

定理 3.2：设 $\lambda_1, \lambda_2, \ldots, \lambda_s$ 是 n 阶矩阵 A 的 s 个不同的特征值，$\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_s$ 分别是它们对应的特征向量，则 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_s$ 线性无关。（证明见交大教材）

定义 3.2 矩阵的相似

设 A、B 为 n 阶矩阵，P 为 n 阶可逆矩阵，且 $P^{-1}AP = B$，则称矩阵 A 相似于矩阵 B，记为 A~B

相似关系具有：自反性、传递性、对称性

性质 3.1：若 A~B，则$|\lambda E - A| = |\lambda E - B|$，因而 A, B 有相同的特征值，相同的行列式和相同的迹

证明：

$$\begin{aligned}
&|\lambda E - B| = |\lambda P^{-1}EP - P^{-1}AP| = |P^{-1}(\lambda E)P - P^{-1}AP| \\
&= |P^{-1}||(\lambda E - A)||P| = |(\lambda E - A)|
\end{aligned}$$

则 A、B 有相同的特征值、行列式与迹。

性质 3.2： 若 $A\sim B$，且矩阵 A 可逆，则矩阵 B 也可逆，且 $A^{-1}\sim B^{-1}$

证明：由性质 3.1，当 $A\sim B$ 时，$|A|=|B|$，所以当 $|A|\neq 0$ 时必有 $|B|\neq 0$，即 A 可逆时，B 也可逆。设 P 为可逆矩阵，且 $B=P^{-1}AP$，则 $B^{-1}=\left(P^{-1}AP\right)^{-1}=P^{-1}A^{-1}P$，即 $A^{-1}\sim B^{-1}$

定理 3.3： n 阶矩阵 A 相似于对角矩阵的充要条件是 A 有 n 个线性无关的特征向量。（证明见交大教材）

推论 3.1： 若 n 阶矩阵 A 有 n 个不同的特征值，则 A 必能相似于对角矩阵

证明：设 A 的 n 个不同的特征值为 $\lambda_1,\lambda_2,\ldots,\lambda_n$，其对应的特征向量为 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\ldots,\boldsymbol{\alpha}_n$，根据定理 3.2 可知，$\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\ldots,\boldsymbol{\alpha}_n$ 线性无关，也就意味着 A 有 n 个线性无关的特征向量。根据定理 3.3 可知，A 必能相似于对角矩阵。

如果 n 阶矩阵 A 能够相似于对角矩阵，它的特征值分解为，

$P^{-1}AP=\Lambda$，其中，$\Lambda=\begin{bmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \vdots & \\ & & & \lambda_n\end{bmatrix}$ 为由 A 的特征值构成的对角矩阵，P 为与之对应的特征向量组成的矩阵。

容易验证，$P^{-1}AP=\Lambda$ 得出 $AP=P\Lambda$，即，

$A\left(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\ldots,\boldsymbol{\alpha}_n\right)=\left(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\ldots,\boldsymbol{\alpha}_n\right)\begin{bmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n\end{bmatrix}$，根据 $A\boldsymbol{\alpha}_i=\lambda_i\boldsymbol{\alpha}_i$，容易知道其成立。

注：由以上分析可知，如果 A 相似于对角矩阵 Λ，则 Λ 中的对角元素都必为 A 的特征值。

定理 3.4： 设 A 为 n 阶实对称矩阵，则存在正交矩阵 Q，使得 $Q^{-1}AQ=Q^TAQ=\Lambda$，

其中 $\Lambda=\begin{bmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n\end{bmatrix}$，$\lambda_1,\lambda_2,\ldots,\lambda_n$ 为 A 的特征值

这个定理说明 n 阶实对称矩阵一定有 n 个线性无关的特征向量。

推论 3.2： 实对称矩阵每个特征值的几何重数等于代数重数。

关于特征值分解的唯一性问题

如果 n 阶矩阵 A 能够相似于对角矩阵，则存在可逆矩阵 P 使得 $A = P\Lambda P^{-1}$，其中 Λ 为特征值对角矩阵。这样的分解不具有唯一性，也就是说满足这个条件的 P 是不唯一的。这是因为 P 的第 i 列 $\boldsymbol{\alpha}_i$ 是 A 的特征向量，则 $\boldsymbol{A\alpha}_i = \lambda_i \boldsymbol{\alpha}_i$，然而 $A(k\boldsymbol{\alpha}_i) = \lambda_i (k\boldsymbol{\alpha}_i)$ 一定成立，也即是说 P 的第 i 列可以替换为 $k\boldsymbol{\alpha}_i\,(k \neq 0)$。如果从特征值所对应的子空间的性质来说，$\lambda_i$ 所形成的特征子空间中的任意一个非零向量都是对应于 λ_i 的特征向量，因此与 λ_i 对应的特征向量不具有唯一性。

如果 A 为实对称矩阵，则 A 必能正交相似于由 A 的特征值所构成的对角矩阵，也就是 $A = Q\Lambda Q^T$，其中 Q 为正交矩阵。在这种情况下，即使 Λ 中的元素是按照大小顺序排布好的，A 的这种分解也不具有唯一性。比如，如果 A 的某个特征值 λ_i 是个 k 重特征值，这样与其对应的正交特征向量就需要有 k 个，而这 k 个特征向量在 Q 中的排列顺序组合是不唯一的。比如：

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

更加本质的理解：k 重特征值对应的特征子空间必须是 k 维的，而这个 k 维的线性空间的单位正交基是不唯一的。更强调一点，即使是一维线性空间，里面只有一个线性无关的特征向量，它所对应的单位向量也不是唯一的，$\dfrac{\boldsymbol{\alpha}}{\|\boldsymbol{\alpha}\|}$ 和 $-\dfrac{\boldsymbol{\alpha}}{\|\boldsymbol{\alpha}\|}$ 都是 $\boldsymbol{\alpha}$ 所张成的一维线性空间的单位向量。综上所述，矩阵的相似对角分解在任何情况下都不具有唯一性，**但化成的对角矩阵如果按照对角元大小排好序则是唯一的，这是因为这个对角矩阵上的对角元素必然是 A 的特征值，而 A 的特征值集合是由 A 唯一确定的，所以在对角元素排好序的情况下 A 所相似于的对角矩阵唯一。**

关于正交矩阵

定义 3.3 正交矩阵

若 A 为 n 阶实矩阵且满足 $AA^T = E$，则 A 被称为正交矩阵。

结论 3.1： 若 A 为正交矩阵，则 A 的行列式 $\det(A) = \pm 1$

证明：A 为正交矩阵，因此 $AA^T = E$，两边计算行列式，有，

$\det(E) = \det(AA^T) = \det(A)\det(A^T) = (\det(A))^2 = 1$，因此 $\det(A) = \pm 1$。

结论 3.2： 若 A 为正交矩阵，则 A 的特征值的模一定为 1

证明：设 λ 是 A 的特征值，$\mathbf{v}$ 是与之对应的特征向量，则 $A\mathbf{v} = \lambda\mathbf{v}$

则，

$$\|A\mathbf{v}\|^2=(\overline{A\mathbf{v}})\cdot(A\mathbf{v})=(\overline{A\mathbf{v}})^T(A\mathbf{v})=(\overline{A}\overline{\mathbf{v}})^T(A\mathbf{v})=(A\overline{\mathbf{v}})^T(A\mathbf{v})=\overline{\mathbf{v}}^T A^T A\mathbf{v}$$
$$=\overline{\mathbf{v}}^T\mathbf{v}=\|\mathbf{v}\|^2$$
另一方面，
$$\|\lambda\mathbf{v}\|^2=(\overline{\lambda\mathbf{v}})^T\lambda\mathbf{v}=(\overline{\lambda}\overline{\mathbf{v}})^T\lambda\mathbf{v}=\overline{\lambda}\overline{\mathbf{v}}^T\lambda\mathbf{v}=\overline{\lambda}\lambda\overline{\mathbf{v}}^T\mathbf{v}=\overline{\lambda}\lambda\|\mathbf{v}\|^2$$

由于$\|A\mathbf{v}\|^2=\|\lambda\mathbf{v}\|^2$，因此，$\overline{\lambda}\lambda\|\mathbf{v}\|^2=\|\mathbf{v}\|^2$，则$\overline{\lambda}\lambda=1$

结论 3.3： 若 A 为正交矩阵，则 A 的实特征值只可能是 1 或者-1。
证明：由结论 3.2 显然得证。

结论 3.4： 若 $A_{n\times n}$ 为正交矩阵且 n 为奇数，且 $\det(A)=1$，则 A 的特征值中必包含 1

证明：
因为正交阵特征值的模均为 1（结论 3.2），且复特征值按共轭关系成对出现（引理 3.2），所以若 1 不是 A 的特征值，那么 A 的特征值只有-1，以及成对出现的复特征值。注意到 A 是奇数阶的，所以除去成对出现的复特征值后必有奇数个特征值-1。这样，利用矩阵 A 的所有特征值之积就等于矩阵 A 的行列式 det$A$ 可知：这奇数个-1 与成对出现的复特征值之积应为 det$A$=1。但是，奇数个-1 的乘积为-1，成对出现的复特征值之积为 1，它们的乘积（行列式）应该是-1，与 det$A$=1 矛盾。因此假设不成立，即 1 必为 A 的一个特征值。.

定义 3.4 矩阵的平方根

如果 $PP=A$，则称 P 是 A 的平方根，记为 $P=A^{\frac{1}{2}}$
如果 A 是半正定矩阵，则这样的 P 比较容易构造。A 必定可以正交相似于对角矩阵（A 为半正定矩阵，则 A 首先必须是实对称矩阵，由定理 3.4 可知，A 必可正交相似于对角矩阵），

$A=U\begin{bmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n\end{bmatrix}U^T$，且 $\lambda_i\geq 0, i=1,\ldots,n$，$U$ 为正交矩阵，

则 P 可被构造为：$P=U\begin{bmatrix}\sqrt{\lambda_1} & & & \\ & \sqrt{\lambda_2} & & \\ & & \ddots & \\ & & & \sqrt{\lambda_n}\end{bmatrix}U^T$

容易验证这样的 P 满足我们的要求，

$$PP=U\begin{bmatrix}\sqrt{\lambda_1}&&&\\&\sqrt{\lambda_2}&&\\&&\ddots&\\&&&\sqrt{\lambda_n}\end{bmatrix}U^TU\begin{bmatrix}\sqrt{\lambda_1}&&&\\&\sqrt{\lambda_2}&&\\&&\ddots&\\&&&\sqrt{\lambda_n}\end{bmatrix}U^T=U\begin{bmatrix}\sqrt{\lambda_1}&&&\\&\sqrt{\lambda_2}&&\\&&\ddots&\\&&&\sqrt{\lambda_n}\end{bmatrix}\begin{bmatrix}\sqrt{\lambda_1}&&&\\&\sqrt{\lambda_2}&&\\&&\ddots&\\&&&\sqrt{\lambda_n}\end{bmatrix}U^T$$

$$=U\begin{bmatrix}\lambda_1&&&\\&\lambda_2&&\\&&\ddots&\\&&&\lambda_n\end{bmatrix}U^T=A$$

且可知，

$$P^{-1}=U\begin{bmatrix}\frac{1}{\sqrt{\lambda_1}}&&&\\&\frac{1}{\sqrt{\lambda_2}}&&\\&&\ddots&\\&&&\frac{1}{\sqrt{\lambda_n}}\end{bmatrix}U^T$$，且$\left(P^{-1}\right)^T=P^{-1}$

结论 3.5：如果 A 是半正定矩阵，则

$$tr\left(A^{\frac{1}{2}}\right)=tr\left(\begin{bmatrix}\sqrt{\lambda_1}&&&\\&\sqrt{\lambda_2}&&\\&&\ddots&\\&&&\sqrt{\lambda_n}\end{bmatrix}\right)=\sum_{i=1}^{n}\sqrt{\lambda_n}=\sum_{i=1}^{r}\sqrt{\lambda_i}$$，其中 $rank(A)=r$，λ_i 是 A 的特征值。

证明：

由于 A 是半正定矩阵，则当然 A 是实对称矩阵，且 $rank(A)=r$，由本章结论 4.3 可知，A 有且仅有 r 个非零特征值，又由于 A 的半正定性，这 r 个非零特征值都大于零。这样也就是说，A 有且仅有 r 个正特征值，$\lambda_i>0\left(i=1,\ldots,r\right)$，其余特征值均为零。

$$A^{\frac{1}{2}}=U\begin{bmatrix}\sqrt{\lambda_1}&&&\\&\sqrt{\lambda_2}&&\\&&\ddots&\\&&&\sqrt{\lambda_n}\end{bmatrix}U^T$$，也就意味着 $A^{\frac{1}{2}}$ 相似于矩阵 $\begin{bmatrix}\sqrt{\lambda_1}&&&\\&\sqrt{\lambda_2}&&\\&&\ddots&\\&&&\sqrt{\lambda_n}\end{bmatrix}$

因此，$tr\left(A^{\frac{1}{2}}\right)=\sum_{i=1}^{n}\sqrt{\lambda_n}=\sum_{i=1}^{r}\sqrt{\lambda_r}$ （性质 3.1）

结论 3.6： $A\in\mathbb{R}^{m\times n},B\in\mathbb{R}^{n\times m}$，则 $tr(AB)=tr(BA)$

结论 3.7 实对称矩阵的最大特征值的函数表示

设 A 为实对称矩阵，则 A 的最大特征值可表示为函数，

$$f(A)=\sup_{\mathbf{x}\neq\mathbf{0}}\frac{\mathbf{x}^TA\mathbf{x}}{\mathbf{x}^T\mathbf{x}}，或者 f(A)=\sup\left\{\mathbf{x}^TA\mathbf{x}\,|\,\|\mathbf{x}\|_2=1\right\} \tag{3-1}$$

证明：

不失一般性，我们假定 A 为 2 阶实对称矩阵。则 A 可以进行正交对角分解，即，$A\alpha_1=\lambda_1\alpha_1,A\alpha_2=\lambda_2\alpha_2$，且 $\|\alpha_1\|=\|\alpha_2\|=1,\alpha_1^T\alpha_2=0$，且 α_1,α_2 构成了二维向量空间的一组正交基。设 $\mathbf{x}$ 可以被 α_1,α_2 线性表出为 $\mathbf{x}=c_1\alpha_1+c_2\alpha_2$，则

$$\begin{aligned}\mathbf{x}^TA\mathbf{x}&=\left(c_1\alpha_1^T+c_2\alpha_2^T\right)A\left(c_1\alpha_1+c_2\alpha_2\right)=\left(c_1\alpha_1^TA+c_2\alpha_2^TA\right)\left(c_1\alpha_1+c_2\alpha_2\right)\\&=\left(c_1\lambda_1\alpha_1^T+c_2\lambda_2\alpha_2^T\right)\left(c_1\alpha_1+c_2\alpha_2\right)=c_1^2\lambda_1+c_2^2\lambda_2\end{aligned},$$

而，$\mathbf{x}^T\mathbf{x}=\left(c_1\alpha_1^T+c_2\alpha_2^T\right)\left(c_1\alpha_1+c_2\alpha_2\right)=c_1^2+c_2^2$，则，

$$\frac{\mathbf{x}^TA\mathbf{x}}{\mathbf{x}^T\mathbf{x}}=\frac{c_1^2\lambda_1+c_2^2\lambda_2}{c_1^2+c_2^2}\le\frac{\left(c_1^2+c_2^2\right)\max\left\{\lambda_1,\lambda_2\right\}}{c_1^2+c_2^2}=\max\left\{\lambda_1,\lambda_2\right\},$$

因此，$f(A)=\sup\limits_{\mathbf{x}\neq\mathbf{0}}\dfrac{\mathbf{x}^TA\mathbf{x}}{\mathbf{x}^T\mathbf{x}}$ 表示的是矩阵 A 的最大特征值。

显然，如果限定 $\|\mathbf{x}\|=1$ 时，$f(A)=\sup\left\{\mathbf{x}^TA\mathbf{x}\,|\,\|\mathbf{x}\|_2=1\right\}$ 。

与特征值相关的概念理解问题：

1) 特征值是不是只能对方阵计算？非方阵能否计算特征值？
2) 实矩阵的特征值、特征向量一定是实数吗？
3) 特征值能为零吗？特征向量能为零向量吗？
4) n 阶方阵一定有 n 个特征值吗？
5) 假设 n 阶矩阵 A 已经确定，它的特征值集合 $\left\{\lambda_i\right\}_{i=1}^n$ 是唯一确定的吗？
6) 如果矩阵 A 相似于对角矩阵 Σ，那么 Σ 中的对角元素一定是 A 的特征值吗？
7) n 阶矩阵 A 特征值对角分解，$A=P\Sigma P^{-1}$，在什么情况下 Σ 的形式是唯一的？在什么情况下 P 的形式是唯一的？
8) 实矩阵与实特征值对应的特征向量一定是实特征向量吗？

   不一定。特征向量是 $(A-\lambda E)\mathbf{x}=0$ 的非零解，当 A 和 λ 都为实数时，$\mathbf{x}$ 一定可以取到实向量；但这不是说 $\mathbf{x}$ 一定必须是实向量，显然如果实向量 $\mathbf{x}$ 是满足条件的向量，那么复向量 $i\mathbf{x}$ 也满足条件。

4. 奇异值分解

首先需要铺垫几个预备结论。

结论 4.1： 如果 $r\left(A_{m\times n}\right)=r$，则 $r\left(A^{T}A\right)=r\left(AA^{T}\right)=r$

证明：只需要证明 $A^{T}A\mathbf{x}=\mathbf{0}$ 与 $A\mathbf{x}=\mathbf{0}$ 同解即可。

如果 $\mathbf{x}$ 为 $A\mathbf{x}=\mathbf{0}$ 的解，即 $A\mathbf{x}=0$，显然必有 $A^{T}A\mathbf{x}=0$，也即 $\mathbf{x}$ 为 $A^{T}A\mathbf{x}=\mathbf{0}$ 的解。

如果 $\mathbf{x}$ 为 $A^{T}A\mathbf{x}=\mathbf{0}$ 的解，则有 $\mathbf{x}^{T}A^{T}A\mathbf{x}=0$，也即 $(A\mathbf{x})^{T}A\mathbf{x}=\mathbf{0}$，则向量 $A\mathbf{x}$ 的长度为 0，当且仅当向量为零向量，即 $A\mathbf{x}=\mathbf{0}$，则 $\mathbf{x}$ 为 $A\mathbf{x}=\mathbf{0}$ 的解。同理可以证明 $r\left(AA^{T}\right)=r(A)=r$。

结论 4.2： A 为 $m\times n$ 矩阵，则 $A^{T}A$ 与 AA^{T} 有相同的非零特征值

证明：设 λ 为 $A^{T}A$ 的特征值，则有 $A^{T}A\boldsymbol{\alpha}=\lambda\boldsymbol{\alpha}$，其中 $\boldsymbol{\alpha}$ 为对应的特征向量。左右同乘 A，$AA^{T}A\boldsymbol{\alpha}=\lambda A\boldsymbol{\alpha}$，即 $AA^{T}(A\boldsymbol{\alpha})=\lambda(A\boldsymbol{\alpha})$，则可知 λ 也为矩阵 AA^{T} 的特征值，对应的特征向量为 $A\boldsymbol{\alpha}$。

结论 4.3： 如果 A 为 n 阶实对称矩阵且 $r(A)=r$，则 A 必有且仅有 r 个不为零的特征值

证明：由于 A 为实对称矩阵，则它必可以（正交）相似于以它的特征值为对角元的对角矩阵，

$$A=P\begin{bmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n\end{bmatrix}P^{-1}$$

其中，P,P^{-1} 均为可逆矩阵（正交矩阵），因此，它们不会改变矩阵 $\begin{bmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n\end{bmatrix}$ 的秩（本章推论 1.5），也就是说，对角矩阵 $\begin{bmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n\end{bmatrix}$ 与 A 的秩相同（实际上，也可以从矩阵等价

的角度理解。A 等价于对角矩阵，则 A 与对角矩阵的秩相同）。因此，$\begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$ 秩为 r。

显然，由于 $\begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$ 为对角矩阵，若它的秩为 r，则它必有且仅有 r 个不为零的对角元，也就意味着 A 有且仅有 r 个不为零的特征值。

结论 4.4： 如果 A 为 $m \times n$ 矩阵，则 $A^T A$ 与 AA^T 均为半正定矩阵

证明：$\forall \mathbf{x} \in \mathbb{R}^{n\times 1} \neq \mathbf{0}$，$0 \leq (A\mathbf{x})^T (A\mathbf{x}) = \mathbf{x}^T A^T A\mathbf{x}$，则 $A^T A$ 为半正定矩阵。

$\forall \mathbf{x} \in \mathbb{R}^{m\times 1} \neq \mathbf{0}$，$0 \leq (A^T\mathbf{x})^T (A^T\mathbf{x}) = \mathbf{x}^T AA^T\mathbf{x}$，则 AA^T 为半正定矩阵。

结论 4.5： 有 $m \times n$ 矩阵 A，且 $rank(A) = r$，则 $(A^T A)_{n\times n}$ 有且仅有 r 个大于零的特征值，且其他特征值均为零。

证明：由于 $r(A) = r$，则根据结论 4.1 可知，$r(A^T A) = r$。更进一步，$A^T A$ 为秩为 r 的实对称矩阵，由结论 4.3 可知，$A^T A$ 必有且仅有 r 个不为零的特征值（其余特征值均为零）。由结论 4.4 可知，$A^T A$ 是个半正定矩阵，因此它的特征值全部为非负数。这样，由于已经知道了 $A^T A$ 有且仅有 r 个不为零的特征值，则这 r 个不为零的特征值必然都大于零。证毕。

奇异值分解定理

任何一个矩阵 $A_{m\times n}$ 都可以分解为形式，

$$A_{m\times n} = U_{m\times m} \Sigma_{m\times n} V_{n\times n}^T$$

其中，$rank(A) = r$，U, V 为正交矩阵，$\Sigma_{m\times n} = \begin{bmatrix} \Sigma_r & \boldsymbol{O}_{r\times(n-r)} \\ \boldsymbol{O}_{(m-r)\times r} & \boldsymbol{O}_{(m-r)\times(n-r)} \end{bmatrix}_{m\times n}$，

$$\Sigma_r = diag(\sigma_1, \sigma_2, ..., \sigma_r)$$

$\sigma_1, \sigma_2, ..., \sigma_r > 0$ 称为矩阵 A 的奇异值。

注意：矩阵 A 的奇异值必为正数。

一般我们将$\sigma_1,\sigma_2,...,\sigma_r$按照从大到小的顺序排列，如此，$\sum_{m\times n}$便是被$A$唯一确定的。

下面我们来理解一下A的奇异值分解与特征值分解之间的关联。

我们现在考虑$\left(A^T A\right)_{n\times n}$的正交对角分解。由于它是实对称矩阵，其必然可以进行正交对角分解，$\left(A^T A\right)_{n\times n} = V'\Lambda V'^T$。如果我们再附加条件，$\Lambda$中的对角元（均为$A^T A$的特征值）按照从大到小的顺序依次排列，则必有，

$$\Lambda_{n\times n} = \begin{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_r \end{bmatrix} & \boldsymbol{O}_{r\times(n-r)} \\ \boldsymbol{O}_{(n-r)\times r} & \boldsymbol{O}_{(n-r)\times(n-r)} \end{bmatrix} \tag{4-1}$$

其中，$\lambda_1 \ge \lambda_2 \ge ... \ge \lambda_r > 0$，且这样的$\Lambda$是唯一确定的，但正交矩阵$V'$不是唯一的。

我们再从奇异值分解的结果看看$A^T A$是什么样子的。

如果$A = U\sum V^T$，则

$$\left(A^T A\right)_{n\times n} = \left(U\sum V^T\right)^T U\sum V^T = V\sum^T U^T U\sum V^T = V\left(\sum^T\sum\right)_{n\times n} V^T$$

$$= V\begin{bmatrix} \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_r^2 \end{bmatrix} & \boldsymbol{O}_{r\times(n-r)} \\ \boldsymbol{O}_{(n-r)\times r} & \boldsymbol{O}_{(n-r)\times(n-r)} \end{bmatrix} V^T \tag{4-2}$$

显然，式（4-2）也是矩阵$\left(A^T A\right)_{n\times n}$的某个具体正交对角分解形式。且当对角元按序排列时，矩阵相似对角化之后的对角矩阵具有唯一性，因此必有，

$$\begin{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_r \end{bmatrix} & \boldsymbol{O}_{r\times(n-r)} \\ \boldsymbol{O}_{(n-r)\times r} & \boldsymbol{O}_{(n-r)\times(n-r)} \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_r^2 \end{bmatrix} & \boldsymbol{O}_{r\times(n-r)} \\ \boldsymbol{O}_{(n-r)\times r} & \boldsymbol{O}_{(n-r)\times(n-r)} \end{bmatrix}$$

则有矩阵A的奇异值$\sigma_i = \sqrt{\lambda_i}, 1 \le i \le r$，$A^T A$的正交相似对角分解产生的矩阵$V'$也就是矩阵$A$的右奇异矩阵$V$（虽然不具有唯一性）。

从以上推导过程中容易知道，$\sum_{m\times n}=\begin{bmatrix}\sum_r & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{O}\end{bmatrix}$实际上是被 A 唯一确定的，但右奇异矩阵 V 不是唯一的。

从上述讨论过程中，我们得出如下结论：

结论 4.6：有矩阵 $A_{m\times n}$，$rank\left(A_{m\times n}\right)=r$，则它的 r 个奇异值$\left\{\sigma_i\right\}_{i=1}^{r}$是$\left(A^TA\right)_{n\times n}$的对应的特征值$\left\{\lambda_i\right\}_{i=1}^{r}$的平方根，即$\sigma_i=\sqrt{\lambda_i}\left(i=1,\dots,r\right)$；$A$ 的奇异值分解的右奇异矩阵 V 就是$\left(A^TA\right)_{n\times n}$进行正交对角分解时产生的正交矩阵。

基于与上面完全类似的推导过程，我们也可以得出矩阵 $A_{m\times n}$ 的奇异值分解与$\left(AA^T\right)_{m\times m}$的特征值分解的关系。

结论 4.7：$\left(AA^T\right)_{m\times m}$ 实际上与$\left(A^TA\right)_{n\times n}$具有相同的非零特征值（由结论 4.2 得到），$\lambda_1\ge\lambda_2\ge\dots\ge\lambda_r>0$，则 A 的 r 个奇异值$\left\{\sigma_i\right\}_{i=1}^{r}$也是$\left(AA^T\right)_{m\times m}$的对应的特征值$\left\{\lambda_i\right\}_{i=1}^{r}$的平方根，即$\sigma_i=\sqrt{\lambda_i}$。另外，$\left(AA^T\right)_{m\times m}=(U\Sigma V)(U\Sigma V)^T=U\Sigma VV^T\Sigma^TU^T=U\Sigma\Sigma^TU^T$。因此，$A$ 的左奇异矩阵 U 也就是$\left(AA^T\right)_{m\times m}$进行正交相似对角分解产生的正交矩阵（虽然不具有唯一性）。

结论 4.8：如果矩阵 $A_{n\times n}$ 是半正定矩阵（当然也是实对称），则它的正交对角特征值分解与它的奇异值分解是相同的。也就是说，如果某种分解形式是 A 的一个正交对角特征值分解，则也是 A 的一个奇异值分解。

证明：

我们先证明这种情况下，A 的奇异值就是 A 的特征值。

假设$rank\left(A\right)=r$，由于 A 是半正定矩阵，由结论 4.5 可知，A 有且仅有 r 个正特征值$\lambda_1\ge\lambda_2\ge\dots\ge\lambda_r>0$，其它特征值均为零。因此，$A$ 可正交相似对角分解为

$$A=U\begin{bmatrix}\begin{bmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_r\end{bmatrix} & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{O}\end{bmatrix}U^T，\text{则 } AA^T=U\begin{bmatrix}\begin{bmatrix}\lambda_1^2 & & & \\ & \lambda_2^2 & & \\ & & \ddots & \\ & & & \lambda_r^2\end{bmatrix} & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{O}\end{bmatrix}U^T$$

则由结论 4.6 可知，A 的奇异值就是$\sqrt{\lambda_1^2},\dots,\sqrt{\lambda_r^2}$，即为$\lambda_1,\dots\lambda_r$

我们再来证明 A 的左奇异矩阵就是 U。这可以由结论 4.7 直接得到。
我们再来证明 A 的右奇异矩阵也是 U。由于 A 是实对称矩阵，

因此，$A^T A = AA^T = U \begin{bmatrix} \begin{bmatrix} \lambda_1^2 & & & \\ & \lambda_2^2 & & \\ & & \ddots & \\ & & & \lambda_r^2 \end{bmatrix} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{bmatrix} U^T$

由结论 4.6 可知，A 的右奇异矩阵就是 $A^T A$ 进行正交对角分解所产生的正交矩阵，即为 U。

奇异值分解的简洁形式

$$A_{m\times n} = U_{m\times m} \Sigma_{m\times n} V_{n\times n}^T = [\mathbf{u}_1,\dots,\mathbf{u}_r \mid \mathbf{u}_{r+1},\dots,\mathbf{u}_m] \begin{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \sigma_2 & \\ & & \ddots\sigma_r \end{bmatrix} & \mathbf{O}_{r\times(n-r)} \\ \mathbf{O}_{(m-r)\times r} & \mathbf{O}_{(m-r)\times(n-r)} \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_r^T \\ \hline \mathbf{v}_{r+1}^T \\ \vdots \\ \mathbf{v}_n^T \end{bmatrix} \quad (4\text{-}3)$$

$$= [\mathbf{u}_1,\dots,\mathbf{u}_r] \begin{bmatrix} \sigma_1 & & \\ & \sigma_2 & \\ & & \ddots\sigma_r \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_r^T \end{bmatrix} = U_{m\times r} \Sigma_{r\times r} V_{n\times r}^T$$

这种形式的 SVD 称为 SVD 的简洁（economy-sized）模式。

矩阵乘法的外积（outer product）表达

如果 X 是 $m\times k$ 矩阵，x_i 是它的列；Y 是 $k\times n$ 矩阵，它的行为 y_i^T，则

$$XY = \sum_{i=1}^{k} x_i y_i^T \quad (4\text{-}4)$$

容易证明，每一个子矩阵 $x_i y_i^T$ 都是一个秩为 1 的矩阵。

现在，通过矩阵的外积分解，我们可以得到奇异值分解的矩阵和形式。

令 $X = [\mathbf{u}_1,\dots,\mathbf{u}_r] \begin{bmatrix} \sigma_1 & & \\ & \sigma_2 & \\ & & \ddots\sigma_r \end{bmatrix} = \left[\sigma_1\mathbf{u}_1, \sigma_2\mathbf{u}_2, \dots, \sigma_r\mathbf{u}_r\right]$

令 $Y=\begin{bmatrix}\mathbf{v}_1^T\\ \mathbf{v}_2^T\\ \vdots\\ \mathbf{v}_r^T\end{bmatrix}$，进一步由 4-3 可知，

$$A=XY=\left[\sigma_1\mathbf{u}_1,\sigma_2\mathbf{u}_2,\ldots,\sigma_r\mathbf{u}_r\right]\begin{bmatrix}\mathbf{v}_1^T\\ \mathbf{v}_2^T\\ \vdots\\ \mathbf{v}_r^T\end{bmatrix}=\sum_{i=1}^{r}\sigma_i\mathbf{u}_i\mathbf{v}_i^T \tag{4-5}$$

下面说明奇异值分解在最小二乘法中的应用。
对于最小二乘问题，

$$\mathbf{x}^*=\arg\min_{\mathbf{x}}\left\|A_{m\times n}\mathbf{x}_{n\times 1}-\mathbf{b}_{m\times 1}\right\|_2^2$$

解 $\mathbf{x}$ 需要满足 $A^TA\mathbf{x}=A^T\mathbf{b}$，当 $rank\left(A\right)=n$ 时，可推出 $rank\left(A^TA\right)=n$（结论 4.1），则 A^TA 才可逆，此时 $\mathbf{x}=\left(A^TA\right)^{-1}A^T\mathbf{b}$（此时，对于方程组 $A^TA\mathbf{x}=A^T\mathbf{b}$ 来说，$r\left(A^TA\right)=r\left(A^TA;A^T\mathbf{b}\right)=n$，因此此方程组有唯一解）；如果 $rank\left(A\right)\neq n$，就不能得出结论了。此时有两种可能性：1）$r\left(A^TA\right)\neq r\left(A^TA;A^T\mathbf{b}\right)$，这时 $A^TA\mathbf{x}=A^T\mathbf{b}$ 没有解，又形成了一个新的最小二乘问题！；2）$r\left(A^TA\right)=r\left(A^TA;A^T\mathbf{b}\right)<n$，这时 $A^TA\mathbf{x}=A^T\mathbf{b}$ 有无穷多组解，接下去我们还需要想办法找到它的一个特解。

而利用奇异值分解，我们可以放松对 A 的秩的要求，使得在更加宽泛的条件下，得到方程系统的最小二乘解（当然，不能保证解的唯一性，唯一性是要靠线性方程组的解理论来保证的）。
设 A 的奇异值分解为 $A=U_{m\times m}\Sigma_{m\times n}V_{n\times n}^T$，则

$$A\mathbf{x}-\mathbf{b}=U\Sigma V^T\mathbf{x}-\mathbf{b}=U\left(\Sigma V^T\mathbf{x}\right)-U\left(U^T\mathbf{b}\right)=U\left(\Sigma V^T\mathbf{x}-U^T\mathbf{b}\right)\triangleq U\left(\Sigma\mathbf{y}_{n\times 1}-\mathbf{c}_{m\times 1}\right)$$

其中，$\mathbf{y}_{n\times 1}=V^T\mathbf{x},\mathbf{c}_{m\times 1}=U^T\mathbf{b}$，而由于 U 是正交矩阵，可以保持向量长度，因此，

$\left\|A\mathbf{x}-\mathbf{b}\right\|=\left\|U\left(\Sigma\mathbf{y}-\mathbf{c}\right)\right\|=\left\|\Sigma\mathbf{y}_{n\times 1}-\mathbf{c}_{m\times 1}\right\|$，因此我们只需要找到 $\mathbf{y}$ 使得 $\left\|\Sigma\mathbf{y}_{n\times 1}-\mathbf{c}_{m\times 1}\right\|$ 最小即可，然后 $\mathbf{x}=V\mathbf{y}$。

假设$\Sigma=\begin{bmatrix}\begin{bmatrix}\sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r\end{bmatrix} & \boldsymbol{O}_{r\times(n-r)} \\ \boldsymbol{O}_{(m-r)\times r} & \boldsymbol{O}_{(m-r)\times(n-r)}\end{bmatrix}_{m\times n}$，则，

$$\Sigma\mathbf{y}_{n\times1}=\begin{bmatrix}\begin{bmatrix}\sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r\end{bmatrix} & \boldsymbol{O}_{r\times(n-r)} \\ \boldsymbol{O}_{(m-r)\times r} & \boldsymbol{O}_{(m-r)\times(n-r)}\end{bmatrix}_{m\times n}\begin{bmatrix}y_1\\y_2\\\vdots\\y_n\end{bmatrix}=\begin{bmatrix}\sigma_1y_1\\\sigma_2y_2\\\vdots\\\sigma_ry_r\\0\\\vdots\\0\end{bmatrix}_{m\times1}$$，所以

$$\Sigma\mathbf{y}_{n\times1}-\mathbf{c}_{m\times1}=\begin{bmatrix}\sigma_1y_1-c_1\\\sigma_2y_2-c_2\\\vdots\\\sigma_ry_r-c_r\\-c_{r+1}\\\vdots\\-c_m\end{bmatrix}_{m\times1}$$，

因此，只要让$y_i=\dfrac{c_i}{\sigma_i},1\le i\le r$（此时$y_{r+1},...,y_n$可以是任意值），则$\left\|\Sigma\mathbf{y}_{n\times1}-\mathbf{c}_{m\times1}\right\|$便可取得最小长度$\left(\sum_{i=r+1}^{m}c_i^2\right)^{1/2}$。

也就是，

$$\mathbf{y}=\begin{bmatrix}\begin{bmatrix}\frac{1}{\sigma_1} & & & \\ & \frac{1}{\sigma_2} & & \\ & & \ddots & \\ & & & \frac{1}{\sigma_r}\end{bmatrix} & \boldsymbol{O}_{r\times(m-r)} \\ \boldsymbol{O}_{(n-r)\times r} & \boldsymbol{O}_{(n-r)\times(m-r)}\end{bmatrix}_{n\times m}\begin{bmatrix}c_1\\c_2\\\vdots\\c_m\end{bmatrix}_{m\times1}=\begin{bmatrix}c_1/\sigma_1\\c_2/\sigma_2\\\vdots\\c_r/\sigma_r\\0\\\vdots\\0\end{bmatrix}_{n\times1}\triangleq\Sigma^+\mathbf{c}_{m\times1}$$

（这只是满足条件的$\mathbf{y}$当中的一个，因为$y_{r+1},...,y_n$可以是任意值。）

其中，Σ^{+}代表了对矩阵Σ做转置并把非零对角元取倒数的操作。因此，最终的结果为，

$\mathbf{x}=V\mathbf{y}=V\Sigma^{+}U^{T}\mathbf{b}$，其中$A^{+}\triangleq V\Sigma^{+}U^{T}$称为$A$的 Moore-Penrose 广义逆。

注意：在以上的推导过程中，我们并没有对A的秩做出限制，最终得到的$\mathbf{x}$是一个具有更加普适范围的最小二乘解。

下面举几个例子，看看特征值分解或奇异值分解的用途。
在 PCA 中，我们其实也用到了特征值分解、或者奇异值分解。

$\mathbf{X}=\left\{\mathbf{x}_{1},\mathbf{x}_{2},\ldots,\mathbf{x}_{n}\right\}\in\mathbb{R}^{d\times n}$是一个数据集，要找它的主成分，我们需要先计算其协方差矩阵，

我们假设 $\mathbf{X}$ 为去均值化之后的数据，则其协方差矩阵为$\mathbf{C}=\dfrac{1}{n-1}\mathbf{X}\mathbf{X}^{T}$。我们在阅读文献时，有些文献说要对$\mathbf{C}$进行特征值分解，有些文献说要对$\mathbf{C}$进行奇异值分解。显然，通过本节的学习我们已经知道：$\mathbf{C}$是实对称矩阵，且更进一步$\mathbf{C}$是半正定矩阵，所以$\mathbf{C}$的特征值分解与它的奇异值分解实际上是一回事。

我们的 homography 估计问题，最终形式化成了$A\mathbf{x}=\mathbf{0}$的问题，变形为

$\mathbf{x}^{*}=\arg\min_{\mathbf{x}}\|A\mathbf{x}\|_{2}^{2},st.,\|\mathbf{x}\|_{2}^{2}=1$，

拉格朗日乘子形式为，$\mathbf{x}^{*}=\arg\min_{\mathbf{x}}\left[\|A\mathbf{x}\|_{2}^{2}+\lambda\left(1-\|\mathbf{x}\|_{2}^{2}\right)\right]$

求驻点，可得$A^{T}A\mathbf{x}=\lambda\mathbf{x}$，则$E\left(\mathbf{x}\right)=\|A\mathbf{x}\|_{2}^{2}=\mathbf{x}^{T}A^{T}A\mathbf{x}=\mathbf{x}^{T}\lambda\mathbf{x}=\lambda$

则$\mathbf{x}$必为与$A^{T}A$的最小的特征值所对应的特征向量。且显然，$A^{T}A$是半正定矩阵，同样的，对它进行特征值分解与进行奇异值分解是相同的。

5. 范数

定理 5.1 范数函数的性质

设Ω是一个线性空间。一个函数$f(\mathbf{x}):\Omega\to[0,+\infty)$为范数，则它要满足

1）正定性，$f(\mathbf{x})\ge 0$，$\forall\mathbf{x}\in\Omega$；且$f(\mathbf{x})=0\Leftrightarrow\mathbf{x}=\mathbf{0}$

2）正齐次性，$\forall\mathbf{x}\in\Omega,\forall t\in\mathbb{R},\ f(t\mathbf{x})=|t|f(\mathbf{x})$，

3）三角不等式，$\forall\mathbf{x},\mathbf{y}\in\Omega,\ f(\mathbf{x}+\mathbf{y})\le f(\mathbf{x})+f(\mathbf{y})$

称二元体$(\Omega,f(.))$为赋范线性空间。

5.1 向量的范数

在本节以下的讨论中，$\mathbf{x}=(x_1,x_2,...,x_n)^T\in\mathbb{R}^{n\times1}$。

定义 5.1 l_p范数

向量的 l_p范数，$p\ge1$，$\|\mathbf{x}\|_p=\left(\sum_{i=1}^n|x_i|^p\right)^{1/p}$

当p取一些具体的值时，我们可以得到一些常见的向量范数形式。

定义 5.2 l_1-范数

p=1, l_1-norm of the vector **x**: $\|\mathbf{x}\|_1=\sum_{i=1}^n|x_i|$

定义 5.3 l_2-范数

p=2, l_2-norm of the vector **x**: $\|\mathbf{x}\|_2=\left(\sum_{i=1}^n|x_i|^2\right)^{1/2}=\left(\mathbf{x}^T\mathbf{x}\right)^{1/2}$

定义 5.4 l_∞-范数

$p=\infty$，l_∞-norm of the vector **x**: $\|\mathbf{x}\|_\infty=\max(|x_1|,|x_2|,...,|x_n|)$

我们可以证明，$\lim\limits_{p\to\infty}\left(\sum_{i=1}^n|\mathbf{x}_i|^p\right)^{1/p}=\max\{|x_1|,|x_2|,...,|x_n|\}$

证明：设$m=\max\{|x_1|,|x_2|,...,|x_n|\}$

一方面，$\left(\sum_{i=1}^{n}\left|\mathbf{x}_i\right|^p\right)^{1/p} \le \left(\sum_{i=1}^{n} m^p\right)^{1/p} = \left(nm^p\right)^{1/p} = n^{1/p}m$，且

$\lim_{p\to\infty}\left(n^{1/p}m\right)=m$

另一方面，$\left(\sum_{i=1}^{n}\left|\mathbf{x}_i\right|^p\right)^{1/p} \ge \left(m^p\right)^{1/p} = m$，

则根据函数极限的夹逼准则，$\lim_{p\to\infty}\left(\sum_{i=1}^{n}\left|\mathbf{x}_i\right|^p\right)^{1/p} = m$

定理 5.2 Holder 不等式

$\left|\mathbf{x}^T\mathbf{y}\right| \le \left\|\mathbf{x}\right\|_p \left\|\mathbf{y}\right\|_q$, where $\frac{1}{p}+\frac{1}{q}=1$

Holder 不等式的一个推论就是 Cauchy-Schwarz 不等式$\left|\mathbf{x}^T\mathbf{y}\right| \le \left\|\mathbf{x}\right\|_2 \left\|\mathbf{y}\right\|_2$

5.2 矩阵的范数

定义 5.5 矩阵的内积

同阶矩阵才能定义内积，定义为对应位置元素相乘再求和，即

$X,Y\in\mathbb{R}^{m\times n}$，它们的内积为，

$$\left\langle X,Y\right\rangle = \sum_{i=1}^{m}\sum_{j=1}^{n}X_{ij}Y_{ij} = tr\left(X^TY\right) = tr\left(XY^T\right)$$

显然，X,Y 的内积也可以看作是把 X 和 Y 按照相同方式展开成列向量，再计算向量的内积。

定义 5.6 矩阵的 Frobenius 范数

矩阵 A 的 Frobenius 范数定义为矩阵所有元素平方和的平方根，即

$$\left\|A\right\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}a_{ij}^2}$$

显然，我们有一些推论成立，

$$\left\|A\right\|_F = \left(tr\left(A^TA\right)\right)^{\frac{1}{2}} = \left(tr\left(AA^T\right)\right)^{\frac{1}{2}} = \sqrt{\sum_{i=1}^{r}\lambda_i} = \sqrt{\sum_{i=1}^{r}\sigma_i^2}$$

其中，$rank(A)=r$，$\left\{\lambda_i\right\}_{i=1}^{r}$ 是矩阵 A^TA 的 r 个正特征值（其余特征值均为零），$\left\{\sigma_i\right\}_{i=1}^{r}$ 为 A 的 r 个奇异值。

证明：

第一、二个等号。

$$\|A\|_F = \left(\langle A, A\rangle\right)^{\frac{1}{2}} = \left(tr\left(A^T A\right)\right)^{\frac{1}{2}} = \left(tr\left(AA^T\right)\right)^{\frac{1}{2}}$$

由于 $A^T A$（AA^T）为半正定矩阵，且 $rank\left(A^T A\right) = r$（$rank\left(AA^T\right) = r$），则由本章结论 4.5 可知，$A^T A$（$AA^T$）有 r 个正特征值 $\left\{\lambda_i\right\}_{i=1}^r$ 且其余特征值均为零。则有，

$$tr\left(A^T A\right) = tr\left(A^T A\right) = \sum_{i=1}^{r} \lambda_i 。$$

更进一步，由本章结论 4.6 可知，$\lambda_i = \sigma_i^2$，则 $\sum_{i=1}^{r} \lambda_i = \sum_{i=1}^{r} \sigma_i^2$，证毕。

定义 5.7 矩阵的 l_1 范数

矩阵每一列元素的绝对值求和，然后取最大值，即 $\|A\|_1 = \max\limits_{1\le j\le n}\left(\sum_{i=1}^{m}\left|a_{ij}\right|\right)$

定义 5.8 矩阵的无穷范数

矩阵每一行元素的绝对值求和，然后取最大值，即 $\|A\|_\infty = \max\limits_{1\le i\le m}\left(\sum_{j=1}^{n}\left|a_{ij}\right|\right)$

定义 5.9 矩阵的谱范数，也称为矩阵的 l_2 范数

矩阵的谱范数（spectral norm）为矩阵的最大奇异值

$$\|A\|_2 = \sigma_{\max}(A) = \sqrt{\lambda_{\max}(A^T A)} = \sqrt{\lambda_{\max}(AA^T)}$$

矩阵的谱范数有如下的函数表示法。

矩阵 $X \in \mathbb{R}^{m\times n}$ 的最大的奇异值称为矩阵的谱范数，它可以表示为函数，

$$f(X) = \sup\left\{\mathbf{u}^T X\mathbf{v} \mid \|\mathbf{u}\|_2 = 1, \|\mathbf{v}\|_2 = 1\right\},$$

证明：

$$f(X) = \sup\left\{\mathbf{u}^T X\mathbf{v} \mid \|\mathbf{u}\|_2 = 1, \|\mathbf{v}\|_2 = 1\right\}$$

$$= \left[\sup\left\{\left(\mathbf{u}^T X\mathbf{v}\right)^T\left(\mathbf{u}^T X\mathbf{v}\right) \mid \|\mathbf{u}\|_2 = 1, \|\mathbf{v}\|_2 = 1\right\}\right]^{\frac{1}{2}}$$

$$= \left[\sup\left\{\mathbf{v}^T X^T \mathbf{u}\mathbf{u}^T X\mathbf{v} \mid \|\mathbf{u}\|_2 = 1, \|\mathbf{v}\|_2 = 1\right\}\right]^{\frac{1}{2}} \qquad (1)$$

$$= \left[\sup\left\{\lambda_{\max}\left\{X^T \mathbf{u}\mathbf{u}^T X\right\} \mid \|\mathbf{u}\|_2 = 1\right\}\right]^{\frac{1}{2}} \qquad (2)$$

$$= \left[\sup\left\{\lambda_{\max}\left\{\mathbf{u}^T XX^T \mathbf{u}\right\} \mid \|\mathbf{u}\|_2 = 1\right\}\right]^{\frac{1}{2}} \qquad (3)$$

$$= \left[\sup\left\{\mathbf{u}^T XX^T \mathbf{u} \mid \|\mathbf{u}\|_2 = 1\right\}\right]^{\frac{1}{2}} \qquad (4)$$

$$= \sqrt{\lambda_{\max}\left(XX^T\right)} = \sigma_{\max}(X) \qquad (5)$$

（1）到（2），当 $\mathbf{u}$ 是固定的时候，变化 $\mathbf{v}$ 对 $\mathbf{v}^T X^T \mathbf{u}\mathbf{u}^T X\mathbf{v}$ 求最大，根据结论 3.7 可知，得到的是矩阵 $X^T\mathbf{u}\mathbf{u}^T X$ 最大的特征值；（2）到（3），记 $A \triangleq X^T\mathbf{u}$，则 $AA^T = X^T\mathbf{u}\mathbf{u}^T X$；根据结论 4.2 可知，$AA^T$ 与 $A^T A$ 具有相同非零特征值，因此 $\lambda_{\max}\left(X^T\mathbf{u}\mathbf{u}^T X\right)=\lambda_{\max}\left(AA^T\right)=\lambda_{\max}\left(A^T A\right)=\lambda_{\max}\left(\mathbf{u}^T XX^T\mathbf{u}\right)$。（3）到（4），由于 $\mathbf{u}^T XX^T\mathbf{u}$ 实际上是个标量，因此 $\lambda_{\max}\left(\mathbf{u}^T XX^T\mathbf{u}\right)=\mathbf{u}^T XX^T\mathbf{u}$。（4）到（5），再一次运用结论 3.7。

定义 5.10 矩阵的核（nuclear）范数

矩阵的核范数定义为矩阵所有奇异值之和，即 $\|A\|_* = \sum_{i=1}^{r}\sigma_i = tr\left(\left(A^T A\right)^{\frac{1}{2}}\right) = tr\left(\left(AA^T\right)^{\frac{1}{2}}\right)$，

其中 $rank(A)=r$，$\left\{\sigma_i\right\}_{i=1}^{r}$ 为 A 的奇异值。

证明：

第一个等号是核范数的定义。

再来理解最后一个等号。我们已经知道 $A^T A$ 与 AA^T 具有相同的非零特征值，且由于它们都是半正定矩阵，因此他们的非零特征值都是正的，$\left\{\lambda_i\right\}_{i=1}^{r}>0$

由本章结论 3.5 可知，

$tr\left(\left(A^T A\right)^{\frac{1}{2}}\right)=\sum_{i=1}^{r}\sqrt{\lambda_i}$，$tr\left(\left(AA^T\right)^{\frac{1}{2}}\right)=\sum_{i=1}^{r}\sqrt{\lambda_i}$，则有 $tr\left(\left(A^T A\right)^{\frac{1}{2}}\right)=tr\left(\left(AA^T\right)^{\frac{1}{2}}\right)$

而由本章结论 4.6 可知，$\sum_{i=1}^{r}\sqrt{\lambda_i}=\sum_{i=1}^{r}\sigma_i$

6. 线性空间

设 P 是包含 0 和 1 的数集，若 P 对加，减，乘和除（0 不作除数）运算是封闭的，则称 P 是一个数域。

定义 6.1：线性空间

设 V 是非空集合，P 是数域。在 V 的元素间定义了一种运算，称为加法：$\forall \alpha, \beta \in V, \exists$唯一的 $\delta \in V$ 与它们对应，称为 α 与 β 的和，记作 $\delta = \alpha + \beta$；加法满足

（1）$\alpha + \beta = \beta + \alpha$

（2）$(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$

（3）V 中有一个零元素 0，$\forall \alpha \in V, \alpha + 0 = \alpha$

（4）$\forall \alpha \in V, \exists \beta \in V$，称为 α 的负元素，记为-α，使得 $\alpha + (-\alpha) = 0$

另外，在 P 和 V 之间还定义了一种运算称为数乘，$\forall \alpha \in V, k \in P$ 有唯一的元素 $\eta \in V$ 与之对应，称为 k 与 α 的数量乘积，记为 $\eta = k\alpha$；数乘满足

（5）$1\alpha = \alpha$

（6）$k(l\alpha) = l(k\alpha) = (kl)\alpha$

加法和数乘还满足

（7）$(k + l)\alpha = k\alpha + l\alpha$

（8）$k(\alpha + \beta) = k\alpha + k\beta$

定义了这两种运算的集合 V 称为数域 P 上的线性空间。

定义 6.2 线性空间中向量的线性相关

设 V 是数域 P 上的线性空间，$\alpha_1, \alpha_2, \ldots, \alpha_m$ 是 V 中向量，若存在 P 中不全为 0 的数 $k_1, k_2, \ldots, k_m$ 使得 $k_1\alpha_1 + k_2\alpha_2 + \ldots + k_m\alpha_m = 0$，则称 $\alpha_1, \alpha_2, \ldots, \alpha_m$ 线性相关，否则称它们线性无关。

注意：上述定义中提到的 V 中的“向量”，是一个推广之后的广泛的概念（比如它的具体形式可以是矩阵，可以是函数等等），并不一定是我们平时所说的向量，这里 V 中的向量只是在表达 V 中的元素。

维数、基与坐标

设 V 是数域 P 上的线性空间，$\alpha_1, \alpha_2, \ldots, \alpha_n$ 是 V 中 n 个线性无关的向量，若 V 中任一向量 α 均可以由 $\alpha_1, \alpha_2, \ldots, \alpha_n$ 线性表示，则称线性空间 V 是 n 维线性空间，$\dim V = n$，$\alpha_1, \alpha_2, \ldots, \alpha_n$ 称为 V 的一组基。

显然，线性空间作为一个向量集合，基是它的极大无关组，维数就是这个极大无关组的秩。

7. 向量与矩阵的求导

7.1. 向量和矩阵函数对标量变量的导数

定义 7.1：

n 维向量函数 $\mathbf{f}(t)=\left(f_1(t),f_2(t),...,f_n(t)\right)^T$，对标量自变量 t 的导数定义为：

$$\frac{d\mathbf{f}(t)}{dt}=\left[\frac{df_1(t)}{dt},\frac{df_2(t)}{dt},...,\frac{df_n(t)}{dt}\right]^T$$

定义 7.2：$n\times m$ 维矩阵函数

$$F(t)=\begin{bmatrix} f_{11}(t) \ \ f_{12}(t)\dots f_{1m}(t) \\ f_{21}(t) \ \ f_{22}(t)\dots f_{2m}(t) \\ \vdots \\ f_{n1}(t) \ \ f_{n2}(t)\dots f_{nm}(t) \end{bmatrix}=[f_{ij}(t)]_{n\times m}$$，它对标量自变量 t 的导数定义为：

$$\frac{df}{dt}=\begin{bmatrix} \frac{df_{11}(t)}{dt}\,\frac{df_{12}(t)}{dt},...,\frac{df_{1m}(t)}{dt} \\ \frac{df_{21}(t)}{dt}\,\frac{df_{22}(t)}{dt},...,\frac{df_{2m}(t)}{dt} \\ \vdots \\ \frac{df_{n1}(t)}{dt}\,\frac{df_{n2}(t)}{dt},...,\frac{df_{nm}(t)}{dt} \end{bmatrix}=\left[\frac{df_{ij}(t)}{dt}\right]_{n\times m}$$

7.2. 标量函数对矩阵变量的导数

定义 7.3：

一个函数 $f:\mathbb{R}^{m\times n}\to\mathbb{R}$ 把一个 $m\times n$ 的矩阵映射成一个实数。我们定义函数 f 对矩阵变量 A 的导数为：

$$\frac{df(A)}{dA}=\begin{bmatrix} \frac{\partial f}{\partial A_{11}}\,\frac{\partial f}{\partial A_{12}}\cdots\frac{\partial f}{\partial A_{1n}} \\ \vdots \\ \frac{\partial f}{\partial A_{m1}}\,\frac{\partial f}{\partial A_{m2}}\cdots\frac{\partial f}{\partial A_{mn}} \end{bmatrix}$$

例子：

假设 $A=\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$，函数 $f(A)=\frac{3}{2}A_{11}+5A_{12}^2+A_{21}A_{22}$，则

$$\frac{df(A)}{dA}=\begin{bmatrix}\frac{3}{2} & 10A_{12}\\ A_{22} & A_{21}\end{bmatrix}$$

7.3. 标量函数对向量变量的导数

定义 7.4：

设标量函数 $f(\mathbf{x}),\mathbf{x}=\left(x_1,x_2,...,x_n\right)^T$ 是以列向量 $\mathbf{x}$ 为自变量的标量函数。则标量函数 $f(\mathbf{x})$ 对列向量自变量 $\mathbf{x}$ 的导数是列向量函数：

$$\frac{df}{d\mathbf{x}}=\left[\frac{\partial f}{\partial x_1},\frac{\partial f}{\partial x_2},...,\frac{\partial f}{\partial x_n}\right]^T$$

实际上 $f(\mathbf{x})$ 就是一般的多元函数，上面的定义就是多元函数的梯度。所以：标量函数对于列向量的导数是一个列向量。

同样，也可以定义标量函数对行向量的导数，

$$\frac{df}{d\mathbf{x}^T}=\left[\frac{\partial f}{\partial x_1},\frac{\partial f}{\partial x_2},...,\frac{\partial f}{\partial x_n}\right]$$

7.4. 向量函数对向量变量的导数

$\mathbf{x}$ 和 $\mathbf{y}$ 分别是 n 维和 m 维的列向量，其中 y_i 是以 $\mathbf{x}$ 为变量的 n 元函数。

由于 $\mathbf{y}$ 是列向量，我们可以直观的定义 $\mathbf{y}$ 对于行向量 $\mathbf{x}^T$ 的导数：

$$\frac{d\mathbf{y}}{d\mathbf{x}^T}=\begin{bmatrix}\frac{\partial y_1(\mathbf{x})}{\partial x_1},\frac{\partial y_1(\mathbf{x})}{\partial x_2},...,\frac{\partial y_1(\mathbf{x})}{\partial x_n}\\ \frac{\partial y_2(\mathbf{x})}{\partial x_1},\frac{\partial y_2(\mathbf{x})}{\partial x_2},...,\frac{\partial y_2(\mathbf{x})}{\partial x_n}\\ \vdots \\ \frac{\partial y_m(\mathbf{x})}{\partial x_1},\frac{\partial y_m(\mathbf{x})}{\partial x_2},...,\frac{\partial y_m(\mathbf{x})}{\partial x_n}\end{bmatrix}_{m\times n}$$

同样，如果 $\mathbf{y}^T$ 被定义为行向量，我们可以定义 $\mathbf{y}^T$ 对于列向量 $\mathbf{x}$ 的导数：

$$\frac{d\mathbf{y}^T}{d\mathbf{x}}=\begin{bmatrix}\frac{\partial y_1(\mathbf{x})}{\partial x_1},\frac{\partial y_2(\mathbf{x})}{\partial x_1},\ldots,\frac{\partial y_m(\mathbf{x})}{\partial x_1}\\ \frac{\partial y_1(\mathbf{x})}{\partial x_2},\frac{\partial y_2(\mathbf{x})}{\partial x_2},\ldots,\frac{\partial y_m(\mathbf{x})}{\partial x_2}\\ \vdots\\ \frac{\partial y_1(\mathbf{x})}{\partial x_n},\frac{\partial y_2(\mathbf{x})}{\partial x_n},\ldots,\frac{\partial y_m(\mathbf{x})}{\partial x_n}\end{bmatrix}_{n\times m}$$

例题：

$$\mathbf{y}=\begin{bmatrix}y_1(\mathbf{x})\\ y_2(\mathbf{x})\end{bmatrix},\mathbf{x}=\begin{bmatrix}x_1\\ x_2\\ x_3\end{bmatrix},y_1(\mathbf{x})=x_1^2-x_2,y_2(\mathbf{x})=x_3^2+3x_2$$

则行向量 $\mathbf{y}^T$ 对列向量 $\mathbf{x}$ 的导数矩阵为：

$$\frac{d\mathbf{y}^T}{d\mathbf{x}}=\begin{bmatrix}\frac{\partial y_1(\mathbf{x})}{\partial x_1} & \frac{\partial y_2(\mathbf{x})}{\partial x_1}\\ \frac{\partial y_1(\mathbf{x})}{\partial x_2} & \frac{\partial y_2(\mathbf{x})}{\partial x_2}\\ \frac{\partial y_1(\mathbf{x})}{\partial x_3} & \frac{\partial y_2(\mathbf{x})}{\partial x_3}\end{bmatrix}=\begin{bmatrix}2x_1 & 0\\ -1 & 3\\ 0 & 2x_3\end{bmatrix}$$

7.5. 常用结论

[1] $\frac{d\mathbf{y}^T(\mathbf{x})}{d\mathbf{x}}=\left(\frac{d\mathbf{y}(\mathbf{x})}{d\mathbf{x}^T}\right)^T$，其中 $\mathbf{x}$ 是 n 维列向量，$\mathbf{y}$ 是 m 维由函数组成的列向量。

这个等式可以直接由向量函数对向量变量的导数的定义得到。

[2] $\mathbf{x}$ 和 $\mathbf{a}$ 都为 n 维列向量，$\mathbf{x}$ 为自变量，$\mathbf{a}$ 为常量，则 $\frac{d\mathbf{a}^T\mathbf{x}}{d\mathbf{x}}=\frac{d\mathbf{x}^T\mathbf{a}}{d\mathbf{x}}=\mathbf{a}$

证明：$f(\mathbf{x})=\mathbf{a}^T\mathbf{x}$ 为一个标量函数。

设 $\mathbf{a}=(a_1,a_2,\ldots,a_n)^T$，则 $f(\mathbf{x})=a_1x_1+a_2x_2+\ldots+a_nx_n$，则，

$$\frac{d\mathbf{a}^T\mathbf{x}}{d\mathbf{x}}=\left(\frac{\partial(\mathbf{a}^T\mathbf{x})}{\partial x_1},\frac{\partial(\mathbf{a}^T\mathbf{x})}{\partial x_2},\ldots,\frac{\partial(\mathbf{a}^T\mathbf{x})}{\partial x_n}\right)^T=(a_1,a_2,\ldots,a_n)^T=\mathbf{a}$$

[3] A 为 $m\times n$ 矩阵，$\mathbf{x}$ 为 n 维列向量，$\frac{dA\mathbf{x}}{d\mathbf{x}^T}=A$

证明：

$$\mathbf{y}(\mathbf{x})=A\mathbf{x}=\begin{bmatrix}a_{11}a_{12}\cdots a_{1n}\\a_{21}a_{22}\cdots a_{2n}\\\vdots\\a_{m1}a_{m2}\cdots a_{mn}\end{bmatrix}\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}=\begin{bmatrix}a_{11}x_1+a_{12}x_2+\cdots+a_{1n}x_n\\a_{21}x_1+a_{22}x_2+\cdots+a_{2n}x_n\\\vdots\\a_{m1}x_1+a_{m2}x_2+\cdots+a_{mn}x_n\end{bmatrix}\triangleq\begin{bmatrix}y_1(\mathbf{x})\\y_2(\mathbf{x})\\\vdots\\y_m(\mathbf{x})\end{bmatrix}$$

$\mathbf{y}$ 是 m 维列向量，现在要求它对 n 维行向量 $\mathbf{x}^T$ 的导数。

$$\frac{d\mathbf{y}(\mathbf{x})}{d\mathbf{x}^T}=\begin{bmatrix}\frac{\partial y_1(\mathbf{x})}{\partial x_1}\frac{\partial y_1(\mathbf{x})}{\partial x_2}\cdots\frac{\partial y_1(\mathbf{x})}{\partial x_n}\\\frac{\partial y_2(\mathbf{x})}{\partial x_1}\frac{\partial y_2(\mathbf{x})}{\partial x_2}\cdots\frac{\partial y_2(\mathbf{x})}{\partial x_n}\\\vdots\\\frac{\partial y_m(\mathbf{x})}{\partial x_1}\frac{\partial y_m(\mathbf{x})}{\partial x_2}\cdots\frac{\partial y_m(\mathbf{x})}{\partial x_n}\end{bmatrix}=\begin{bmatrix}a_{11}a_{12}\cdots a_{1n}\\a_{21}a_{22}\cdots a_{2n}\\\vdots\\a_{m1}a_{m2}\cdots a_{mn}\end{bmatrix}=A$$

[4] A 为 $m*n$ 矩阵，$\mathbf{x}$ 为 n 维列向量，$\frac{d\left(\mathbf{x}^TA^T\right)}{d\mathbf{x}}=A^T$

证明：$\frac{d\left(\mathbf{x}^TA^T\right)}{d\mathbf{x}}=\frac{d(A\mathbf{x})^T}{d\mathbf{x}}$，根据结论[1]，$\frac{d(A\mathbf{x})^T}{d\mathbf{x}}=\left(\frac{dA\mathbf{x}}{d\mathbf{x}^T}\right)^T=A^T$

[5] A 为 $n\times n$ 矩阵，$\mathbf{x}$ 为 n 维列向量，$\frac{d\left(\mathbf{x}^TA\mathbf{x}\right)}{d\mathbf{x}}=\left(A+A^T\right)\mathbf{x}$

证明：可以知道 $\mathbf{x}^TA\mathbf{x}$ 是关于列向量 $\mathbf{x}$ 的标量函数。

不失一般性，我们假定 A 是 3 阶方阵，$\mathbf{x}$ 是 3 维列向量。

$$\mathbf{x}^TA\mathbf{x}=\left(x_1,x_2,x_3\right)\begin{bmatrix}a_{11}&a_{12}&a_{13}\\a_{21}&a_{22}&a_{23}\\a_{31}&a_{32}&a_{33}\end{bmatrix}\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}$$

$$=x_1^2a_{11}+x_1x_2a_{21}+x_1x_3a_{31}+x_1x_2a_{12}+x_2^2a_{22}+x_2x_3a_{32}+x_1x_3a_{13}+x_2x_3a_{23}+x_3^2a_{33}$$

则：

$$\frac{d\mathbf{x}^TA\mathbf{x}}{d\mathbf{x}}=\begin{bmatrix}2a_{11}x_1+\left(a_{12}+a_{21}\right)x_2+\left(a_{13}+a_{31}\right)x_3\\\left(a_{21}+a_{12}\right)x_1+2a_{22}x_2+\left(a_{23}+a_{32}\right)x_3\\\left(a_{31}+a_{13}\right)x_1+\left(a_{32}+a_{23}\right)x_2+2a_{33}x_3\end{bmatrix}$$

$$=\begin{bmatrix}2a_{11}&\left(a_{12}+a_{21}\right)&\left(a_{13}+a_{31}\right)\\\left(a_{21}+a_{12}\right)&2a_{22}&\left(a_{23}+a_{32}\right)\\\left(a_{31}+a_{13}\right)&\left(a_{32}+a_{23}\right)&2a_{33}\end{bmatrix}\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}=\left(\begin{bmatrix}a_{11}&a_{12}&a_{13}\\a_{21}&a_{22}&a_{23}\\a_{31}&a_{32}&a_{33}\end{bmatrix}+\begin{bmatrix}a_{11}&a_{21}&a_{31}\\a_{12}&a_{22}&a_{32}\\a_{13}&a_{23}&a_{33}\end{bmatrix}\right)\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}$$

$$=\left(A+A^T\right)\mathbf{x}$$

[6] $\dfrac{d\mathbf{a}^T\mathbf{X}\mathbf{b}}{d\mathbf{X}}=\mathbf{a}\mathbf{b}^T$，其中 $\mathbf{X}$ 是含有未知数的 $m\times n$ 矩阵。$\mathbf{a}$、$\mathbf{b}$ 分别为 m 和 n 维列向量。

证明：不失一般性，设 $\mathbf{a}=\begin{pmatrix}a_1\\a_2\end{pmatrix},\mathbf{b}=\begin{pmatrix}b_1\\b_2\\b_3\end{pmatrix},\mathbf{X}=\begin{bmatrix}x_{11}&x_{12}&x_{13}\\x_{21}&x_{22}&x_{23}\end{bmatrix}$，则：

$$f(\mathbf{X})=\begin{pmatrix}a_1&a_2\end{pmatrix}\begin{bmatrix}x_{11}&x_{12}&x_{13}\\x_{21}&x_{22}&x_{23}\end{bmatrix}\begin{pmatrix}b_1\\b_2\\b_3\end{pmatrix}=a_1b_1x_{11}+a_2b_1x_{21}+a_1b_2x_{12}+a_2b_2x_{22}+a_1b_3x_{13}+a_2b_3x_{23}$$

则：

$$f(\mathbf{X})=\begin{pmatrix}a_1&a_2\end{pmatrix}\begin{bmatrix}x_{11}&x_{12}&x_{13}\\x_{21}&x_{22}&x_{23}\end{bmatrix}\begin{pmatrix}b_1\\b_2\\b_3\end{pmatrix}=a_1b_1x_{11}+a_2b_1x_{21}+a_1b_2x_{12}+a_2b_2x_{22}+a_1b_3x_{13}+a_2b_3x_{23}$$

$$\frac{df(\mathbf{X})}{d\mathbf{X}}=\begin{bmatrix}\dfrac{\partial f}{\partial x_{11}}&\dfrac{\partial f}{\partial x_{12}}&\dfrac{\partial f}{\partial x_{13}}\\\dfrac{\partial f}{\partial x_{21}}&\dfrac{\partial f}{\partial x_{22}}&\dfrac{\partial f}{\partial x_{23}}\end{bmatrix}=\begin{bmatrix}a_1b_1&a_1b_2&a_1b_3\\a_2b_1&a_2b_2&a_2b_3\end{bmatrix}=\begin{pmatrix}a_1\\a_2\end{pmatrix}\begin{pmatrix}b_1&b_2&b_3\end{pmatrix}=\mathbf{a}\mathbf{b}^T$$

[7] $\dfrac{d\mathbf{a}^T\mathbf{X}^T\mathbf{b}}{d\mathbf{X}}=\mathbf{b}\mathbf{a}^T$，其中 $\mathbf{X}$ 是含有未知数的 $n\times m$ 矩阵。$\mathbf{a}$、$\mathbf{b}$ 分别为 m 和 n 维列向量。

证明：

$$\frac{d\mathbf{a}^T\mathbf{X}^T\mathbf{b}}{d\mathbf{X}}=\frac{d\left(\mathbf{b}^T\mathbf{X}\mathbf{a}\right)^T}{d\mathbf{X}}=\frac{d\mathbf{b}^T\mathbf{X}\mathbf{a}}{d\mathbf{X}}=\mathbf{b}\mathbf{a}^T$$

[8] $\dfrac{d\mathbf{x}^T\mathbf{x}}{d\mathbf{x}}=2\mathbf{x}$，其中 $\mathbf{x}$ 是 $n\times1$ 的列向量

证明：

$$f(\mathbf{x})=\mathbf{x}^T\mathbf{x}=x_1^2+x_2^2+\ldots+x_n^2,\quad \frac{df}{d\mathbf{x}}=\begin{bmatrix}\dfrac{\partial f}{\partial x_1}\\\dfrac{\partial f}{\partial x_2}\\\vdots\\\dfrac{\partial f}{\partial x_n}\end{bmatrix}=\begin{bmatrix}2x_1\\2x_2\\\vdots\\2x_n\end{bmatrix}=2\mathbf{x}$$

[9] $\dfrac{d\left(trAB\right)}{dA}=B^T$，其中 A 为 $m\times n$ 矩阵自变量，B 为 $n\times m$ 矩阵

证明：

$$AB=\begin{bmatrix} A_{11}\ A_{12}\dots A_{1n} \\ A_{21}\ A_{22}\dots A_{2n} \\ \vdots \\ A_{m1}\ A_{m2}\dots A_{mn} \end{bmatrix}\begin{bmatrix} B_{11}\ B_{12}\dots B_{1m} \\ B_{21}\ B_{22}\dots B_{2m} \\ \vdots \\ B_{n1}\ B_{n2}\dots B_{nm} \end{bmatrix}$$

则 $trAB=\sum_{j=1}^{n}A_{1j}B_{j1}+\sum_{j=1}^{n}A_{2j}B_{j2}+\ldots+\sum_{j=1}^{n}A_{mj}B_{jm}=\sum_{i=1}^{m}\sum_{j=1}^{n}A_{ij}B_{ji}$

则 $\frac{d\left(trAB\right)}{dA}=\begin{bmatrix} B_{11}\ B_{21},\ldots,B_{n1} \\ \vdots \\ B_{1n}\ B_{2n},\ldots,B_{nm} \end{bmatrix}=B^T$

[10] $\frac{df\left(A\right)}{dA^T}=\left(\frac{df\left(A\right)}{dA}\right)^T$

这个结论根据标量函数对矩阵变量的导数的定义很容易验证。

[11] $\frac{d\left|A\right|}{dA}=\left|A\right|\left(A^{-1}\right)^T$

证明：由伴随矩阵的理论，我们知道 $A^*=\left|A\right|A^{-1}$

又由 $\left|A\right|=\sum_j A_{ij}A_{ji}^*$，由于 A_{ji}^* 不依赖于 A_{ij}，则

$\frac{d\left|A\right|}{dA_{ij}}=A_{ji}^*$，这样就有 $\frac{d\left|A\right|}{dA}=\left(A^*\right)^T=\left|A\right|\left(A^{-1}\right)^T$

例 1：线性最小二乘法。$\mathbf{x}^*=\arg\min_{\mathbf{x}}\left\|A\mathbf{x}-\mathbf{b}\right\|_2^2, A\in\mathbb{R}^{m\times n},\mathbf{x}\in\mathbb{R}^{n\times 1},\mathbf{b}\in\mathbb{R}^{m\times 1}$，且

$rank\left(A\right)=n$

我们要计算 $\left\|A\mathbf{x}-\mathbf{b}\right\|_2^2$ 的驻点，即梯度为零的点。

方法 1：

$$\begin{aligned}\left\|A\mathbf{x}-\mathbf{b}\right\|_2^2&=\left(A\mathbf{x}-\mathbf{b}\right)^T\left(A\mathbf{x}-\mathbf{b}\right)\\ &=\left(\mathbf{x}^TA^T-\mathbf{b}^T\right)\left(A\mathbf{x}-\mathbf{b}\right)=\mathbf{x}^TA^TA\mathbf{x}-\mathbf{x}^TA^T\mathbf{b}-\mathbf{b}^TA\mathbf{x}+\mathbf{b}^T\mathbf{b}\\ &=\mathbf{x}^TA^TA\mathbf{x}-2\mathbf{x}^TA^T\mathbf{b}+\mathbf{b}^T\mathbf{b}\end{aligned}$$

则

$\frac{d\left\|A\mathbf{x}-\mathbf{b}\right\|_2^2}{d\mathbf{x}}=\frac{d\mathbf{x}^TA^TA\mathbf{x}}{d\mathbf{x}}-2\frac{d\mathbf{x}^TA^T\mathbf{b}}{d\mathbf{x}}=\left(A^TA+A^TA\right)\mathbf{x}-2A^T\mathbf{b}=\mathbf{0}$，则

$2A^TA\mathbf{x}=2A^T\mathbf{b}$，即 $\mathbf{x}=\left(A^TA\right)^{-1}A^T\mathbf{b}$

方法 2：链式求导

令 $\mathbf{e}(\mathbf{x}) = A\mathbf{x} - \mathbf{b}$， $E(\mathbf{x}) = \mathbf{e}^2(\mathbf{x})$，则

$$\frac{dE}{d\mathbf{x}^T} = \frac{dE}{d\mathbf{e}^T} \cdot \frac{d\mathbf{e}}{d\mathbf{x}^T} = \frac{d\mathbf{e}^T\mathbf{e}}{d\mathbf{e}^T} \cdot \frac{d(A\mathbf{x} - \mathbf{b})}{d\mathbf{x}^T} = 2\mathbf{e}^T \cdot A = 2(A\mathbf{x} - \mathbf{b})^T A$$ ，则

$$\frac{dE}{d\mathbf{x}} = \left(\frac{dE}{d\mathbf{x}^T}\right)^T = \left(2(A\mathbf{x} - \mathbf{b})^T A\right)^T = 2A^T(A\mathbf{x} - \mathbf{b})$$ ，则如果令 $\frac{dE}{d\mathbf{x}} = \mathbf{0}$，则有

$2A^T(A\mathbf{x} - \mathbf{b}) = \mathbf{0}$，即 $\mathbf{x} = \left(A^T A\right)^{-1} A^T\mathbf{b}$

例 2：岭回归（ridge regression）。

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} \|A\mathbf{x} - \mathbf{b}\|_2^2 + \lambda\|\mathbf{x}\|_2^2, A \in \mathbb{R}^{m\times n}, \mathbf{x} \in \mathbb{R}^{n\times 1}, \mathbf{b} \in \mathbb{R}^{m\times 1}, \lambda > 0$$

$$E(\mathbf{x}) = \|A\mathbf{x} - \mathbf{b}\|_2^2 + \lambda\|\mathbf{x}\|_2^2 = (A\mathbf{x} - \mathbf{b})^T(A\mathbf{x} - \mathbf{b}) + \lambda\mathbf{x}^T\mathbf{x}$$
$$= \left(\mathbf{x}^T A^T - \mathbf{b}^T\right)(A\mathbf{x} - \mathbf{b}) + \lambda\mathbf{x}^T\mathbf{x} = \mathbf{x}^T A^T A\mathbf{x} - \mathbf{x}^T A^T\mathbf{b} - \mathbf{b}^T A\mathbf{x} + \mathbf{b}^T\mathbf{b} + \lambda\mathbf{x}^T\mathbf{x}$$

则，

$$\frac{dE}{d\mathbf{x}} = \frac{d\mathbf{x}^T A^T A\mathbf{x}}{d\mathbf{x}} - 2\frac{d\mathbf{x}^T A^T\mathbf{b}}{d\mathbf{x}} + \lambda\frac{d\mathbf{x}^T\mathbf{x}}{d\mathbf{x}} = \left(A^T A + A^T A\right)\mathbf{x} - 2A^T\mathbf{b} + 2\lambda\mathbf{x} = \mathbf{0}$$ ，则

$$2A^T A\mathbf{x} + 2\lambda E\mathbf{x} = 2A^T\mathbf{b} \to \left(A^T A + \lambda E\right)\mathbf{x} = A^T\mathbf{b} \to \mathbf{x} = \left(A^T A + \lambda E\right)^{-1} A^T\mathbf{b}$$

（注意： $A^T A + \lambda E$ 必为正定矩阵）