Note of Linear Algebra

Yzy

2020 年 9 月 7 日

目录

1 Vector and Linear Combination

scalar multiplication(标量乘法): $2\boldsymbol{v}$ vector addition(矢量加法): $\boldsymbol{vw}$
xy plane(二维平面) xyz space(三维空间)

$$\text{列向量 } v = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} \text{转置 } \textit{(transpose)} \rightarrow \text{行向量 } v = \begin{bmatrix} 1 & 1 & -1 \end{bmatrix}$$

假设有三维向量 $\boldsymbol{u},\boldsymbol{v},\boldsymbol{w}$, 系数 c, d, e, 在不考虑它们是零向量的情况下:

1. $c\boldsymbol{u} + d\boldsymbol{v} + e\boldsymbol{w}$ 覆盖整个三维空间
2. $c\boldsymbol{u} + d\boldsymbol{v}$ 覆盖三维空间中的某一个平面
3. $c\boldsymbol{u}$ 覆盖三维空间中的某一条直线

故欲使 $c\boldsymbol{u} + d\boldsymbol{v} = \boldsymbol{w}$ 有解，$\boldsymbol{w}$ 必须在 $c\boldsymbol{u} + d\boldsymbol{v}$ 构成的平面上!
整数 whole numbers 等间距点 equally spaced points
几何 geometric lattice 点阵
点积 dot product(内积 inner product): $\boldsymbol{v} \cdot \boldsymbol{w}$ 即为对应元素相乘再求和!
长度 $\|\boldsymbol{v}\| = \sqrt{\boldsymbol{v} \cdot \boldsymbol{v}} = (v_1^2 + v_2^2 + \cdots + v_n^2)^{1/2}$
单位向量 (unit vector) $\boldsymbol{u} = \frac{\boldsymbol{v}}{\|\boldsymbol{v}\|}$ 且与 $\boldsymbol{v}$ 同方向!
方阵:

- 非奇异矩阵 (可逆矩阵) nonsingular(invertible) matrix
- 奇异矩阵 (非可逆矩阵) singular matrix

2 Solving Linear Equations

2.1 vectors and linear equations

$\boldsymbol{A}_{m\times n}\boldsymbol{x}_{n\times 1}=\boldsymbol{b}$ 分别从行和列的角度来看

- 行：从 m 个方程 n 个未知数得到的“平面”是处于 **n 维空间**
- 列：从 n 列且行数为 m 的列向量得到的线性组合是处于 **m 维空间**

矩阵乘法 (行图像形式)：

$$\begin{bmatrix}2 & 0\\0 & 5\end{bmatrix}\begin{bmatrix}row1\\row2\end{bmatrix}=\begin{bmatrix}2*row1+0*row2\\0*row1+5*row2\end{bmatrix}=\begin{bmatrix}2*row1\\5*row2\end{bmatrix}$$

$$AB=\begin{bmatrix}\boldsymbol{a_1}\\\boldsymbol{a_2}\end{bmatrix}B=\begin{bmatrix}\boldsymbol{a_1}\cdot B\\\boldsymbol{a_2}\cdot B\end{bmatrix}=\begin{bmatrix}a_{11}*\boldsymbol{b_1}+a_{12}*\boldsymbol{b_2}\\a_{21}*\boldsymbol{b_1}+a_{22}*\boldsymbol{b_2}\end{bmatrix}$$

这里 $\boldsymbol{a_i}$ 与 $\boldsymbol{b_i}$ 均为行向量。
用这种乘法可以快速理解左乘置换矩阵 P_{ij} 为什么能改变行的顺序

$$AB=A\begin{bmatrix}\boldsymbol{b_1} & \boldsymbol{b_2} & \boldsymbol{b_3}\end{bmatrix}=\begin{bmatrix}A\boldsymbol{b_1} & A\boldsymbol{b_2} & A\boldsymbol{b_3}\end{bmatrix}=\begin{bmatrix}\boldsymbol{a_1}\\\boldsymbol{a_2}\\\boldsymbol{a_3}\end{bmatrix}B=\begin{bmatrix}\boldsymbol{a_1}B\\\boldsymbol{a_2}B\\\boldsymbol{a_3}B\end{bmatrix}$$

这里 $\boldsymbol{b_i}$ 为列向量，$\boldsymbol{a_i}$ 为行向量。
列图像形式: 矩阵 A 乘矩阵 B 的**每列**，得到的结果 AB 的每一列都是 **A 的列的线性组合**。
行图像形式: 矩阵 A 的**每行**乘矩阵 B，则 AB 的每一行均为 **B 的行的线性组合** **列乘行形式**:

$$\begin{bmatrix}col1 & col2 & col3\\. & . & .\\. & . & .\end{bmatrix}\begin{bmatrix}row1\cdots\\row2\cdots\\row3\cdots\end{bmatrix}=(col1)(row1)+(col2)(row2)+(col3)(row3)$$

矩阵乘法在邻接矩阵中应用 2.4 worked example A
adjacency matrix

$$\boldsymbol{S}=\begin{bmatrix}0 & 1 & 1 & 0\\1 & 0 & 1 & 1\\1 & 1 & 0 & 1\\0 & 1 & 1 & 0\end{bmatrix}$$

其中，$(S^2)_{ij}$ 表示结点 i 与结点 j 之间，步长为 2 的路径数目。
$(S^2)_{ij}$ =(row i of S)·(column j of S)= $s_{i1}s_{1j}+s_{i2}s_{2j}+s_{i3}s_{3j}+s_{i4}s_{4j}$

2.2 矩阵的 LU 分解

$(E_{32}E_{31}E_{21})A=U\to A=(E_{21}^{-1}E_{31}^{-1}E_{32}^{-1})U\to A=LU$
$\boldsymbol{L},\boldsymbol{U}$ 分别是**下三角矩阵，上三角矩阵**。

从将 A 转换成行阶梯矩阵或行最简矩阵的目的以及过程中就可以发现，每次左乘的 E_ij 矩阵必须是**下三角矩阵**才能将矩阵 A 对角线左下的元素消为 0！所以矩阵 $\boldsymbol{L}$ 一定是**下三角矩阵**！
又因为

$$U=\begin{bmatrix} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_n \end{bmatrix}\begin{bmatrix} 1 & u_{12}/d_1 & u_{13}/d_1 & \cdot \\ & 1 & u_{23}/d_2 & \cdot \\ & & \ddots & \vdots \\ & & & 1 \end{bmatrix}$$

所以也可以写成 $A=LDU$,D 为主元对角矩阵，U 主对角线为 1。

2.3 转置 (transposes) 与置换 (permutations)

转置 transpose: $(A^T)_{ij}=A_{ij}$
若 LU 分解中的 A 是**对称的 (symmetric)**, 则记为 S, 且分解过程可写成 $A=LDU\to S=LDL^T$，其中 $U=L^T$
置换 permute 以及置换矩阵 permutation matrices P
P 即是对**单位矩阵 I** 进行**行变换或列变换**得到的矩阵。而不是只进行交换两行或两列的初等矩阵！
$P^{-1}=P^T$ 且 $PP^T=I$

提前做行交换再进行分解 PA=LU
若 A 是可逆的，且 A 在消元过程中需要做行交换，则可以在分解之前利用 P 提前做行变换 $\to PA$, 因此可以得到 $PA=LU$

3 向量空间 Space of Vectors

标准 n 维空间 R^n 可以理解为：$R \times R \times \cdots \times R$, 即 n 个 R 的笛卡尔积，即 n 维向量的所有可能情况！

向量空间必须满足的规则：该空间对**数乘，加法**封闭，即通过这两种操作所产生的向量依然在此空间。

注意：一个在三维空间中通过 (0,0,0) 的平面，是该三维空间的子空间，但不是 $\boldsymbol{R^2}$。因为按照定义,$\boldsymbol{R^2}$ 中的向量只能是 2 维的，但是该子空间里的向量是三维的，从定义上就不对。

3.1 子空间 subspaces

同样对**数乘，加法**封闭，且**必须包含零向量** (当 c=0 时，数乘 $c\boldsymbol{v} = \mathbf{0}$)

关于 $\boldsymbol{R^3}$ 的所有子空间：

- 所有过 (0,0,0) 的平面
- 所有过 (0,0,0) 的直线
- $\boldsymbol{Z}$ 原点，或零向量 (0,0,0)
- 整个空间 $\boldsymbol{R^3}$

3.2 列空间 column space

基于矩阵 A 的**列向量的所有线性组合**构成了列空间，这些组合是所有可能的 $A\boldsymbol{x}$, 它们填满了**列空间** $\boldsymbol{\mathcal{C}}(A)$。

$A_{m\times n}$ 有 n 个列向量，每个列向量有 m 维。故，这些列向量属于 $\boldsymbol{R^m}$ 空间。

A 的列空间是 $\boldsymbol{R^m}(not\boldsymbol{R^n})$ 的**一个子空间** (满足子空间要求)！

这一个概念至关重要,因为它<u>将 $A\boldsymbol{x} = \boldsymbol{b}$ 问题转化为 $\boldsymbol{b}$ 是否能由 A 的列向量线性表出的问题</u>！

方程组 $A\boldsymbol{x} = \boldsymbol{b}$ 有解当且仅当向量 $\boldsymbol{b}$ 处于 A 的列空间中！

$A, \quad [A \quad AB]$ **有相同的列空间**, 因为 AB **实质也是 A 的列向量的线性组合**！

3.3 子空间的交、并、和

子空间的并:$P \cup L$, 如平面 P 和一条直线 L 之间，它们之间的并集对线性运算 (加法) 不封闭. 所以子空间的并不是一个向量空间！

子空间的交:$P \cap L$, 平面 P 和直线 L 相交于原点，显然是 $\boldsymbol{R^3}$ 的子空间。
推广： 任意两个子空间的交集，仍是子空间！
子空间的和：P+L 各取一条向量，相加之后的向量处于平面和直线之间！脱离了并集构成的空间. 若两个子空间均为直线，则它们的和为一个过原点的平面。**子空间的加法不仅包含它们的并，也包含了它们的线性组合**。例如：并就是集合的并集，而和是在并的基础之上把所有线性运算的结果都包含进去。比如 x 轴并 y 轴就是那两个轴，而它们的和得把空的地方都补上，是整个平面。
$dim(S) + dim(U) = dim(S \cap U) + dim(S + U)$

3.4 零空间 Nullspaces

零空间 Nullspaces $\boldsymbol{N}(A)$, 行最简矩阵 (reduced row echelon form) $R = \boldsymbol{rref}(A)$,$\boldsymbol{N}(A) = \boldsymbol{N}(R)$
$A_{m\times n}\boldsymbol{x} = \boldsymbol{0}$ 的**所有解向量的线性组合**构成了零空间 $\boldsymbol{N}(A)$, 因为解向量 $\boldsymbol{x}$ 是 n 维的，所以该零空间是 n 维空间 $\boldsymbol{R}^n$ 的一个**子空间**。
$m > n$ 对方程组添加新的方程（行），会对解向量变得更加严格，原零空间不会变大！
$n > m$ 方程组至少有一个自由变量，即至少有一个非零解。
零空间是子空间，且它的**维度是自由变量的个数** (非主元 pivot 个数，即列数 n 减去主元个数)

3.5 秩 rank

rank of A : 主元 pivot 数目，记为 r。 每个自由列都是主列的线性组合！

- A 有 r 个线性无关列 (行)
- A 的列空间 (行空间) 的维度是 r
- n-r 是 $A_{m\times n}\boldsymbol{x} = \boldsymbol{0}$ 的解空间的维度

这里说的**维度** $\rightarrow$ 构成向量空间的**基的线性无关向量的个数**！
特殊的—秩为 1 的矩阵

秩为 1 的矩阵可以写成 (列向量乘行向量): $A = \boldsymbol{u}\boldsymbol{v}^T$. 　如:

$$\begin{bmatrix}1 & 3 & 10\\ 2 & 6 & 20\\ 3 & 9 & 30\end{bmatrix} = \begin{bmatrix}1\\ 2\\ 3\end{bmatrix}\begin{bmatrix}1 & 3 & 10\end{bmatrix}$$

$A\boldsymbol{x} = \mathbf{0}, \rightarrow \boldsymbol{u}(\boldsymbol{v}^T\boldsymbol{x}) = \mathbf{0} \rightarrow \boldsymbol{v}^T\boldsymbol{x} = \mathbf{0}$ 可以看出，行 (列) 空间是一条直线，且解空间中所有的向量都要垂直于向量 $\boldsymbol{v}^T$, 故解空间是一个垂直于向量的平面 $\boldsymbol{v}^T$。

A 的按秩分解规律

$A_{m\times n}, r(A) = r$，欲将 A 分解为 $(m\times r)(r\times n)$ 的矩阵:

$A = (A\textbf{的主元列})(R\ \textbf{前}r\textbf{行}) = (COL)(ROW)$, R 为 A 的行最简矩阵。把这个过程看成是 **A 的主元列的线性组合**，R 前 r 行包含了单位阵 I 和自由变量列。

3.6 线性相关性 Independence，基 Basis，维数 Dimension

一个空间的维度 dimension 是组成这个空间的基的线性无关的向量个数！$\mathbf{R}^n$ 的**维数是 n, 且向量的维数也是 n**

$\mathbf{C}(A)$—>r 维；$\mathbf{N}(A)$—>$n-r$ 维。

3.7 矩阵空间 Matrix Spaces 5th P182

一个向量空间 M 包含了所有 2*2 矩阵，它的维度是 4！　它的一个**基**为：$A_1, A_2, A_3, A_4 = \begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix}, \begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}, \begin{bmatrix}0 & 0\\ 1 & 0\end{bmatrix}, \begin{bmatrix}0 & 0\\ 0 & 1\end{bmatrix}$

3.8 函数空间 Function Spaces

二阶微分方程 $\frac{d^2y}{dx^2}+y=0$ 只考虑实数范围,有两个特解:$y = \sin x and y = \cos x$

它的通解即为这两个特解的线性组合：$\mathrm{y} = c_1\cos\mathrm{x} + c_2\sin\mathrm{x}$。

这类似于零空间，将解看成空间中的元素，全部解构成解空间，满足线性运算封闭条件。**维数为 2**

$y'' = 0$ 的解空间即类似零空间，为 $y = cx + d$。

而 $y'' = 2$ 不形成子空间，因为右边的 2 不是 0，无法形成零空间。它的全

部解为特解 $y(x)=x^2$ 加上零空间 $y=cx+d$ 即为 $y(x)=x^2+cx+d$。
空间 $\boldsymbol{Z}$ 只包含零向量，它的维度为 **0**。空集 empty set 是 $\boldsymbol{Z}$ 的一个基。若将一个零向量加入到一个基中，则会**破坏基中的线性无关性**。

3.9 四个基本子空间

- **关联矩阵** incidence matrix 可以用来表示有向图，有环则说明结点代表的向量是**线性相关的**, 有树则**线性无关**
- 列空间 $\mathcal{C}(A)$，行空间 $\mathcal{C}(A^T)$，零空间 $\mathcal{N}(A)$，左零空间 $\mathcal{C}(A^T)$

若求出了 A 的行阶梯或行最简矩阵 R，再求出所用的行变换矩阵 E，则可以通过 E 来求出左零空间的基，即 $A^Tx=0$ 的解。

$$\boldsymbol{EA}=\boldsymbol{R}=\begin{bmatrix}1 & 0 & 0\\ -1 & 1 & 0\\ -2 & -2 & 1\end{bmatrix}\begin{bmatrix}1 & 0 & 3\\ 1 & 1 & 7\\ 4 & 2 & 20\end{bmatrix}=\begin{bmatrix}1 & 0 & 3\\ 0 & 1 & 4\\ 0 & 0 & 0\end{bmatrix}$$

因为 A^T 的列是 A 的行，所以可以从上面直接看出 3-2=1，即解空间维数唯一。且 R 的第三行为零向量，由 E 的第三行乘上 A 的每行向量得到！
所以有 $A^T(row_3(E)^T)=\mathbf{0}$, **所以可以不用再求解一遍** $A^Tx=0$，**可以直接从 E 和 R 看出左零空间的基！**
故有：若 $\boldsymbol{EA}=\boldsymbol{R}$, 则 E 的**后** $\boldsymbol{m-r}$ **行**是 A 的左零空间的基！
注意: AB 的所有行都是 B 行的组合, 因此 AB 的行空间包含在（可能等于）B 的行空间中, 所以 rank（AB）<rank（B）。AB 的所有列是 A 的列的组合, 因此 AB 的列空间包含在（可能等于）A 的列空间中, 所以 rank（AB）<rank（A）。
若 A,B 的四个基本子空间都相同，则 $\boldsymbol{B=cA, c\neq 0}$

3.10 每个秩 r 的矩阵 =r 个秩 1 矩阵之和

对应上面矩阵 $\boldsymbol{EA}=\boldsymbol{R}$

$$A=\begin{bmatrix}u_1 & u_2 & u_3\end{bmatrix}\begin{bmatrix}v_1^{\mathrm{T}}\\ v_2^{\mathrm{T}}\\ \text{zero row}\end{bmatrix}=u_1v_1^{\mathrm{T}}+u_2v_2^{\mathrm{T}}=(\text{ rank } 1)+(\text{ rank } 1)$$

A 的主元列 $\boldsymbol{u_1}, \boldsymbol{u_2}$, R 的前 r 行 (主元行)$\boldsymbol{v_1^T}, \boldsymbol{v_2^T}$.

4 正交性 orthogonality

4.1 4 个子空间的正交性

当 b 在列空间之外时 (当我们想要求解 Ax = b 而不能这样做时) 那么 A^T 的这个零空间就会自成一体。它包含“最小二乘”解决方案中的误差 $e=b-Ax$。最小二乘法是本章线性代数的关键应用。

同一个向量空间的两个子空间 V,W 有 **V 与 W 是正交的**: $\boldsymbol{v}^{\mathrm{T}}\boldsymbol{w}=0$ for all $\boldsymbol{v}$ in $\boldsymbol{V}$ and all $\boldsymbol{w}$ in $\boldsymbol{W}$

注意：一定是子空间中的每个向量！比如三维空间中的两个垂直的平面，因为它们的交线不能垂直于自己，所以这两个平面不是正交的！

若有一个向量能同时处于两个正交子空间中，则这个向量必定是**零向量**, 零向量垂直于自己。

当 $\dim V+\dim W>\dim(\text{ whole space })$，$V$ 和 W **不可能正交**!

Every vector x in the nullspace is perpendicular to every row of A, because $Ax=0$.

The nullspace $N(A)$ and the row space $C\left(A^T\right)$ are orthogonal subspaces of $\mathbf{R}^n$.$C(A^T)\perp N\left(A\right)$

2 proof:

- $A\boldsymbol{x}=\begin{bmatrix}\text{row } 1\\ \vdots\\ \text{row } m\end{bmatrix}[\boldsymbol{x}]=\begin{bmatrix}0\\ \vdots\\ 0\end{bmatrix}$,Every row has a zero dot product with $\boldsymbol{x}$. Then $\boldsymbol{x}$ is also perpendicular to every combination of the rows.
- 行空间中的向量是行的线性组合 $A^T\boldsymbol{y}$, $\boldsymbol{x}$ 则是零空间的向量，则有 $\boldsymbol{x}^{\mathrm{T}}\left(A^{\mathrm{T}}\boldsymbol{y}\right)=(A\boldsymbol{x})^{\mathrm{T}}\boldsymbol{y}=\mathbf{0}^{\mathrm{T}}\boldsymbol{y}=0$.

Every vector y in the nullspace of A^{T} is perpendicular to every column of A

The left nullspace $N\left(A^{\mathrm{T}}\right)$ and the column space $C(A)$ are orthogonal in $\mathbf{R}^m$, $C(A)\perp N\left(A^{\mathrm{T}}\right)$.

4.2 正交补 orthogonal complements

定义：一个子空间 V 的**正交补包含了每一条垂直于 V 的向量**，这个正交空间记为：$V^{\perp}$.
例子：零空间是行空间的正交补，每一条垂直于行空间的向量都满足 $A\boldsymbol{x}=\mathbf{0}$，且均处于零空间中。它们满足 $r+(n-r)=n$。列空间和左零空间同理！

- 零空间 $\mathbf{N}(A)$ 是行空间 $\boldsymbol{C}\left(A^{\mathrm{T}}\right)$ (in $\mathbf{R}^{n}$) 的正交补—$\mathbf{N}(A)=(\boldsymbol{C}\left(A^{\mathrm{T}}\right))^{\perp}$
- 左零空间 $N\left(A^{\mathrm{T}}\right)$ 是列空间 $C(A)$ (in $\mathbf{R}^{m}$) 的正交补—$\mathbf{N}(A^{T})=(\boldsymbol{C}\,(A))^{\perp}$

正交补的两个子空间的**交集只能为 0**。
互补的含义
每个向量 $\boldsymbol{x}$ 都可以分成行空间分量 $\boldsymbol{x}_r$ 和零空间分量 $\boldsymbol{x}_n$.
$\boldsymbol{x}=\boldsymbol{x}_r+\boldsymbol{x}_n, \quad A\boldsymbol{x}=A\boldsymbol{x}_r+A\boldsymbol{x}_n.$

- 零空间分量变成零向量 $A\boldsymbol{x}_n=\mathbf{0}$
- 行空间分量变成了列空间向量 $A\boldsymbol{x}_r=A\boldsymbol{x}$

这说明了，一个向量左乘一个矩阵 A，会变成列空间中的一个向量！而且：列空间中的向量 $\boldsymbol{b}$ 来自有且仅有一个行空间中的向量 $\boldsymbol{x}_r$。

- 一个原空间 S 可以划分为行空间和零空间, 或者列空间和左零空间, 一**般地说可以划分为两个正交补子空间！**
- 正交补子空间空间的特点是**两个正交子空间维度之和为 dim(S)**

正交补子空间的基底构成了原空间的基底！由行空间和零空间的基底可得，$r+(n-r)=n$ 个向量，这 r 个向量是线性无关的，因此它们构成了 $\mathbf{R}^{n}$。如果所有 n 个向量的组合给出 $\boldsymbol{x}_r+\boldsymbol{x}_n=\mathbf{0}$, 则 $\boldsymbol{x}_r=-\boldsymbol{x}_n$ 在两个子空间中。所以 $\boldsymbol{x}_r=\boldsymbol{x}_n=\mathbf{0}$。行空间基底和零空间基底的所有系数必须为零。这证明了 n 个向量的线性无关性。
练习：若 $A^{\mathrm{T}}A\boldsymbol{x}=0$ then $A\boldsymbol{x}=\mathbf{0}$. 因为 $A\boldsymbol{x}$ 处于 A 的列空间中，同时也处于 A^{T} 的零空间中，即 A 的左零空间中，故 $A\boldsymbol{x}$ 只能是**零向量**。

4.3 投影 projections

投影矩阵 Projection matrix

$$P_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{将向量 } \boldsymbol{b} \text{ 投影到 z 轴}$$

$$\boldsymbol{p}_1 = P_1\boldsymbol{b} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ z \end{bmatrix}$$

$$P_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{将向量 } \boldsymbol{b} \text{ 投影到 } xy-plane$$

$$p_2 = P_2 b = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x \\ y \\ 0 \end{bmatrix}$$

投影后的 z 轴与 xy 平面是正交补,$dim(zaxis)+dim(xy-plane) = dim(\boldsymbol{R}^3)$.
所以此三维空间中的每个向量 $\boldsymbol{b}$ 都可以由这两个子空间的向量构成。
即 $\boldsymbol{p}_1 + \boldsymbol{p}_2 = \boldsymbol{b}$，投影矩阵也有 $P_1 + P_2 = I$。
利用 $A\boldsymbol{x} = \boldsymbol{b}$ 是将 x 投影至 A 的列空间这一概念, 可以解释上述投影原理：将 z 轴和 xy-plane 的基分别放至一个对应 $\boldsymbol{R}^m$ 空间的 $m \times m$ 矩阵 P_1, P_2 中，这两个矩阵分别乘上向量 $\boldsymbol{b}$ 即表示将向量 $\boldsymbol{b}$ 投影至它们的列空间，正好分别对应 z 轴和 xy-plane。

4.3.1 b 投影到直线 a 上

$\boldsymbol{b}$ 投影到 $\boldsymbol{a}$ 上的投影向量为 $\boldsymbol{p}$, 必有 $\boldsymbol{p} = \widehat{x}\boldsymbol{a}$, 其中 $\hat{x}$ 是倍数。
误差向量 $\boldsymbol{e} = \boldsymbol{b} - \boldsymbol{p} = \boldsymbol{b} - \widehat{x}\boldsymbol{a}$，且 $\boldsymbol{e} \perp \boldsymbol{a}$。 利用它们点乘为 0 求出倍数：
$a \cdot (b - \widehat{x}a) = 0 —> a \cdot b - \widehat{x}a \cdot a = 0 —> \widehat{\boldsymbol{x}} = \frac{\boldsymbol{a} \cdot \boldsymbol{b}}{\boldsymbol{a} \cdot \boldsymbol{a}} = \frac{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}}{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}}$
所以，向量 $\boldsymbol{b}$ 在 $\boldsymbol{a}$ 上的投影为 $\boldsymbol{p} = \widehat{\boldsymbol{x}}\boldsymbol{a} = \frac{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}}{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}}\boldsymbol{a}$。
再求所用的投影矩阵 P
$\boldsymbol{p} = \boldsymbol{a}\widehat{\boldsymbol{x}} = \boldsymbol{a}\frac{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}}{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}} = P\boldsymbol{b} \quad \therefore P = \frac{aa^{\mathrm{T}}}{a^{\mathrm{T}}a}, rank(P) = 1$
$P^2 = P$，投影两次到 P 的列空间中还是一样的。
矩阵 $I - P$ 同样对向量 $\boldsymbol{b}$ 作投影，$(I - P)\boldsymbol{b} = \boldsymbol{b} - \boldsymbol{p} = \boldsymbol{e}$, 这说明该矩阵将 b 投影到 b 的垂直部分 e。

当矩阵 P 将 b 投影至一个子空间 V 中，I-P 会将 b 投影到一个垂直于 V 的子空间 W 中，上述的 I-P 将 b 投影至垂直于向量 $\boldsymbol{a}$ 的平面中!

4.3.2 b 投影到子空间上

假设有 n 个向量 $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ in $\mathbf{R}^m$,且它们线性无关,$A_{m\times n} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n]$。找到最接近 $\boldsymbol{b}$ 的线性组合 $\boldsymbol{p} = \widehat{x}_1\boldsymbol{a}_1 + \cdots + \widehat{x}_n\boldsymbol{a}_n = A\widehat{\boldsymbol{x}}$ 即为将 $\boldsymbol{b}$ 投影至 $\boldsymbol{a's}$ 扩展的子空间。
若 n=1, 则说明这是投影至一条直线上，$\widehat{x} = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}/\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}$
若 n>1,则说明这是投影至一个子空间上,则最佳的组合系数是 $\widehat{\boldsymbol{x}} = (\widehat{x}_1, \ldots, \widehat{x}_n)$
同样有误差向量 $\boldsymbol{e} = \boldsymbol{b} - \boldsymbol{p} = \boldsymbol{b} - A\widehat{x}$ 垂直于 $\boldsymbol{a's}$ 构成的子空间，即 $\perp \boldsymbol{a}_1, \ldots, \perp \boldsymbol{a}_n$，利用点积为 0 构造等式：

$$\begin{matrix} \boldsymbol{a}_1^{\mathrm{T}}(\boldsymbol{b} - A\widehat{\boldsymbol{x}}) = 0 \\ \vdots \\ \boldsymbol{a}_n^{\mathrm{T}}(\boldsymbol{b} - A\widehat{\boldsymbol{x}}) = 0 \end{matrix} \rightarrow \begin{bmatrix} -\boldsymbol{a}_1^{\mathrm{T}}- \\ \vdots \\ -\boldsymbol{a}_n^{\mathrm{T}}- \end{bmatrix} \begin{bmatrix} \boldsymbol{b} - A\widehat{\boldsymbol{x}} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \end{bmatrix} \rightarrow A^{\mathrm{T}}(\boldsymbol{b} - A\widehat{\boldsymbol{x}}) = \mathbf{0}$$

将该等式展开，得到一个重要的式子：$A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$——**正规方程组**
又 $\because A_{m\times n}$ 的列向量线性无关，所以 A^TA 是可逆的 (证明见下面)。由此可以求得:
系数向量 $\widehat{\boldsymbol{x}}_{n\times 1} = \left(A^{\mathrm{T}}A\right)^{-1}A^{\mathrm{T}}\boldsymbol{b}$
$\therefore$ 投影向量 $\boldsymbol{p}_{m\times 1} = A\widehat{\boldsymbol{x}} = A\left(A^{\mathrm{T}}A\right)^{-1}A^{\mathrm{T}}\boldsymbol{b}$
$\therefore$ 投影矩阵 $P_{m\times m} = A\left(A^{\mathrm{T}}A\right)^{-1}A^{\mathrm{T}}$
这与 n=1 即投影到一条直线上时对应的公式非常类似!

同样，$P^2 = P$，做两次投影不改变第一次的投影!

注意，公式中的 $P = A\left(A^{\mathrm{T}}A\right)^{-1}A^{\mathrm{T}}$ 不能拆成 $P = AA^{-1}\left(A^{\mathrm{T}}\right)^{-1}A^{\mathrm{T}}$，因为 A 有可能不是方阵，则不存在 A 的逆!

P222 证明了当 A 不是方阵时，$rank(A^TA) = rank(A)$。(同解方程组，即相同的零空间！)

If $A^{\mathrm{T}}Ax = 0$ then $Ax = 0$ 的向量空间解法：如果 $A^{\mathrm{T}}A\boldsymbol{x} = \mathbf{0}$，则 $A\boldsymbol{x}$ 位于 A 的左零空间 $\boldsymbol{N}(A^T)$ 中，但是 $A\boldsymbol{x}$ 总是在 A 的列空间 $\boldsymbol{C}(A)$ 中。要同时在这两个正交互补空间中，$A\boldsymbol{x}$ 必须为零。所以 A 和 A^TA 具有相同的零空间。
投影矩阵性质：$P = A\left(A^{\mathrm{T}}A\right)^{-1}A^{\mathrm{T}}$

- $P^{\mathrm{T}} = P$

- $P^2 = P$
- $P\boldsymbol{b} = \boldsymbol{p}$
- P 将 b 投影到 A 的列空间 $\boldsymbol{C}(A)$，$I - P$ 将 b 投影到 A 的左零空间 $\boldsymbol{N}(A^T)$

更一般地，在 $\boldsymbol{R}^m$ 中，若 $\boldsymbol{b}$ 投影到其中 n 个正交补子空间中 V_i(n)00009m)，且 $\sum_{i=1}^{n} dim(V_i) = m$，则有 $\sum_{i=1}^{n} \boldsymbol{p}_i = \boldsymbol{b}, \sum_{i=1}^{n} P_i = I$。
例如，三维空间中将 b 分别投影到一组基底，则投影向量便是这组基底构成 b 的线性组合，和为 b；投影矩阵之和便是单位矩阵。

4.4 最小二乘逼近 Least Squares Approximations

4.2 练习倒数第二题 Kalman filter 卡尔曼滤波器
假设 $A_{m\times n}$ 列向量线性无关，从 Ax 是将 x 投影到 A 的列空间的角度来看 (也即**投影角度**)：

- 当 $A\boldsymbol{x} = \boldsymbol{b}$ 有解时，向量 $\boldsymbol{b}$ 处于 A 的列空间中，误差 $\boldsymbol{e} = \boldsymbol{b} - A\boldsymbol{x} = zero$，所以 $\boldsymbol{x}$ 就是要求的解；
- 当 $A\boldsymbol{x} = \boldsymbol{b}$ 无解时，向量 $\boldsymbol{b}$ 处于 A 的列空间外，误差 $\boldsymbol{e} = \boldsymbol{b} - A\boldsymbol{x}$ 是必然存在的，但是如果能让误差 $\boldsymbol{e}$ 尽可能地小，那便称 $\hat{\boldsymbol{x}}$(投影系数) 是最小二乘解。

当求解 $A\boldsymbol{x} = \boldsymbol{b}$ 这种无解问题的最小二乘解时，这里直接给出其中一种**基于投影**的做法：

可以利用上一小节的投影法中误差向量 $\boldsymbol{e}$ 垂直于投影空间 (列空间) 这个条件求得最小二乘解。也即**等式两边左乘** A^T，然后再解 $A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$。

至于为什么该条件能够使其误差向量最小？详见下面一小节。

4.4.1 最小化误差 minimizing the Error

最小二乘法的矩阵以及代数直接推导见 (`https://www.cnblogs.com/pinard/p/5976811.html`)

三种方式给出最佳 $\boldsymbol{x}$ 即投影系数 $\hat{\boldsymbol{x}}$。

- 几何：通过误差向量 $\boldsymbol{e}$ 与所有列向量 (线性无关) 垂直这个条件，得出最佳误差 $\boldsymbol{e}(= \boldsymbol{b} - \boldsymbol{p})$，所以最佳系数为 $\boldsymbol{p} = A\widehat{\boldsymbol{x}}$ 中的 $\hat{\boldsymbol{x}}$。

- 代数：(**正规方程组法 normal eqution**) $\boldsymbol{b}=\boldsymbol{p}+\boldsymbol{e}$ 将向量 $\boldsymbol{b}$ 分成两部分，其中 $\boldsymbol{p}$ 在列空间中，$\boldsymbol{e}$ 与列空间垂直并处于左零空间。所以 $A\boldsymbol{x}=\boldsymbol{b}=\boldsymbol{p}+\boldsymbol{e}$ 是不可解的；$A\widehat{\boldsymbol{x}}=\boldsymbol{p}$ 是有解的 (通过去掉误差向量 $\boldsymbol{e}$ 并求解 $A^{\mathrm{T}}A\widehat{\boldsymbol{x}}=A^{\mathrm{T}}\boldsymbol{b}$ 得到 $\widehat{\boldsymbol{x}}=\left(A^{\mathrm{T}}A\right)^{-1}A^{\mathrm{T}}\boldsymbol{b}$)。

  将 Ax 视为 b 尝试投影到 A 列空间的投影向量 $\boldsymbol{p}$，则有误差向量 (非最小)$\boldsymbol{e}=A\boldsymbol{x}-\boldsymbol{b}$，若要使误差最小，即使误差长度 $E=\|Ax-b\|^2$ 最小，则取 $x=\widehat{\boldsymbol{x}}$ 满足要求。

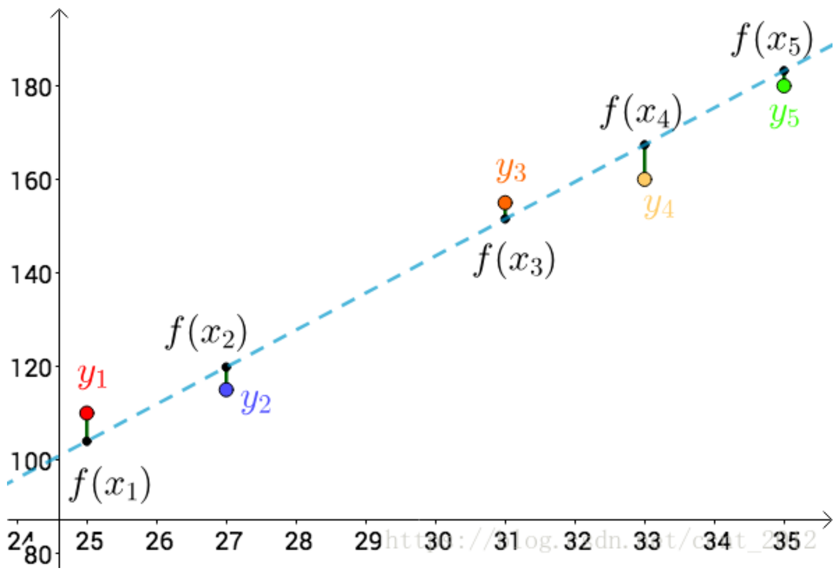
- 微积分：通过令误差函数 $E=\|Ax-b\|^2$(写成函数形式) 对各个未知数 (即 $\widehat{\boldsymbol{x}}$ 的各个分量) 的一阶偏导为零 (式子要除以 2)，来求出矩阵 A^TA。

$A\boldsymbol{x}=\boldsymbol{b}$(A 的列向量线性无关)：

- **有解：** 将 $\boldsymbol{x}$ 分解成 $\boldsymbol{x}_r+\boldsymbol{x}_n$ 来求解！
- **无解：** 将 $\boldsymbol{b}$ 分解成 $\boldsymbol{b}=\boldsymbol{p}+\boldsymbol{e}$ 来求解，即求解 $A\widehat{\boldsymbol{x}}=\boldsymbol{p}$，而误差 $\boldsymbol{e}=\boldsymbol{b}-\boldsymbol{p}$ 是不可避免的！

拟合直线： 有多个原数据点 (t,b)，$\boldsymbol{p}$ 是用来代替原数据高向量 $\boldsymbol{b}$ 的高度向量，$\boldsymbol{e}$ 是拟合的高度与原数据点高度之差的向量！

$$A\boldsymbol{x}=\boldsymbol{b}\rightarrow\begin{array}{l}C+Dt_1=b_1\\C+Dt_2=b_2\\\vdots\\C+Dt_m=b_m\end{array}\quad with\quad A=\begin{bmatrix}1&t_1\\1&t_2\\\vdots&\vdots\\1&t_m\end{bmatrix}\quad\widehat{\boldsymbol{x}}=(\boldsymbol{C},\boldsymbol{D})\quad e_i=b_i-C-Dt_i$$



square.png square.bb

Gram-Schmidt 施密特正交化（4.3 最后一题）：先将 A 的列正交化（即将 $x_i := x_i - 1/n\sum_{i=1}^{n} x_i$），然后便可以得到对角矩阵 $A^T A$，这样求 $\widehat{\boldsymbol{x}}$ 就很方便！

4.4.2 若 A 的列向量线性相关

因为 A 的列向量线性相关，若把 b 投影到 A 的列空间中，即求解 $A\widehat{\boldsymbol{x}} = \boldsymbol{p}$, 则此时会产生多个解，即有多条直线可以选，但是不知道哪条是最佳的！

涉及到奇异值分解 SVD(7.4)，7.4 中会利用 A 的伪逆”pseudoinverse” 来选择一个最短解 $\widehat{\boldsymbol{x}}$。

这里直接给出结果：若 A 的**列线性无关**，它的伪逆是 $L = \left(A^{\mathrm{T}}A\right)^{-1}A^{\mathrm{T}}$

拟合抛物线：

同样有 m 个点，$(t_1, b_1), \ldots, (t_m, b_m)$，利用 $C + Dt + Et^2$ 来拟合这些点！

$$\begin{aligned} C + Dt_1 + Et_1^2 &= b_1 \\ &\vdots \\ C + Dt_m + Et_m^2 &= b_m \end{aligned} \qquad \begin{array}{l} \text{is } A\boldsymbol{x} = \boldsymbol{b} \text{ with} \\ \text{the } m \text{ by } 3 \text{ matrix} \end{array} \qquad A = \begin{bmatrix} 1 & t_1 & t_1^2 \\ \vdots & \vdots & \vdots \\ 1 & t_m & t_m^2 \end{bmatrix}$$

最接近的抛物线利用 $\widehat{\boldsymbol{x}} = (C, D, E)$ 来满足正规方程组 $A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$。即给出了最小二乘解（具有最小均方误差 MSE）。

解耦 uncouple: 指在线性方程组中只通过一个方程就可以解出一个变量的值。
耦合 couple: 指单个方程里的多个变量只能通过整个方程组来解出。
递归最小二乘 recursive least squares: 对旧均值增加一个新数据后，不用重新计算所有值之和再除以总数，$\{b_1, \ldots, b_n\} + b_{n+1}$，$\widehat{x}_{\text{new}} = \widehat{x}_{\text{old}} + \frac{1}{n+1}\left(b_{n+1} - \widehat{x}_{\text{old}}\right)$

4.5 规范（标准）正交基和施密特正交化 Orthonormal Bases and Gram-Schmidt

当 $\boldsymbol{q}_i^{\mathrm{T}}\boldsymbol{q}_j = \begin{cases} 0 & \text{when } i \neq j \quad \text{orthogonal vectors} \\ 1 & \text{when } i = j \quad (\text{unit vectors: } \|\boldsymbol{q}_i\| = 1) \end{cases}$ 向量 $q_1, \ldots, q_n$ 是规范正交的。

若一个矩阵的列均为规范正交向量，则记为 $\boldsymbol{Q}$，有 $Q^{\mathrm{T}}Q = I$。且若 Q 是**方阵**，则称其为**正交矩阵 orthogonal matrix**，有 $Q^{\mathrm{T}} = Q^{-1}$：**转置 = 逆**!
正交矩阵的作用:

- **旋转 Rotation** （二维下）$Q = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ 将向量旋转 θ 度；
  $Q^{\mathrm{T}} = Q^{-1} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$ 将向量旋转 $-\theta$ 度。
- **置换 Permutation** 每个置换矩阵都是正交矩阵，它的逆 = 转置！
- **反射 Reflection** 设 $\boldsymbol{u}$ 是任意一个单位列向量，且 $Q = I - 2uu^{\mathrm{T}}$，则有 $Q^{\mathrm{T}} = I - 2uu^{\mathrm{T}} = Q$，$Q^{\mathrm{T}}Q = I - 4\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}} + 4\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}} = I$，反射矩阵可以将向量 u 转为 $-u$。

一个向量乘上正交矩阵, 它的**长度，角度都不改变**：$\|Q\boldsymbol{x}\|^2 = (Q\boldsymbol{x})^{\mathrm{T}}(Q\boldsymbol{x}) = \boldsymbol{x}^{\mathrm{T}}Q^{\mathrm{T}}Q\boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}}I\boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} = \|\boldsymbol{x}\|^2$，即有 $\|Q\boldsymbol{x}\| = \|\boldsymbol{x}\|$；也能使点积结果不变：$(Q\boldsymbol{x})^{\mathrm{T}}(Q\boldsymbol{y}) = \boldsymbol{x}^{\mathrm{T}}Q^{\mathrm{T}}Q\boldsymbol{y} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}$

4.5.1 用规范正交基做投影，Q 代替 A

对于投影到子空间，通常利用 A^TA，A^TA 中的条目为 $\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{a}_j$，即 A 的基的点积。
现假设 A 的基是规范正交的，则 A^TA 简化为 $Q^TQ = I$(解耦了)，且:

- $Q^TQ\widehat{\boldsymbol{x}}=Q^{\mathrm{T}}\boldsymbol{b}\rightarrow\widehat{\boldsymbol{x}}=Q^{\mathrm{T}}\boldsymbol{b}$
- $\boldsymbol{p}=Q\widehat{\boldsymbol{x}}=QQ^{\mathrm{T}}\boldsymbol{b}$
- $P=Q(Q^TQ)^{-1}Q^{\mathrm{T}}\rightarrow P=QQ^T$

$$\boldsymbol{p}=QQ^{\mathrm{T}}\boldsymbol{b}=\left[\begin{array}{ccc}| & & | \\ \boldsymbol{q}_1 & \cdots & \boldsymbol{q}_n \\ | & & |\end{array}\right]\left[\begin{array}{c}\boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{b} \\ \vdots \\ \boldsymbol{q}_n^{\mathrm{T}}\boldsymbol{b}\end{array}\right]=\boldsymbol{q}_1\left(\boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{b}\right)+\cdots+\boldsymbol{q}_n\left(\boldsymbol{q}_n^{\mathrm{T}}\boldsymbol{b}\right)$$

当 Q 为方阵时，列空间便为整个空间，$Q^{\mathrm{T}}=Q^{-1}$，$\boldsymbol{x}=Q^{-1}\boldsymbol{b}$，最小二乘解是精确的，$\boldsymbol{b}$ 在整个空间上的投影是它自己，$P=QQ^{\mathrm{T}}=I$。
$\boldsymbol{p}=\boldsymbol{b}=QQ^{\mathrm{T}}\boldsymbol{b}=\boldsymbol{q}_1\left(\boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{b}\right)+\boldsymbol{q}_2\left(\boldsymbol{q}_2^{\mathrm{T}}\boldsymbol{b}\right)+\cdots+\boldsymbol{q}_n\left(\boldsymbol{q}_n^{\mathrm{T}}\boldsymbol{b}\right)$，这说明了每个 $\boldsymbol{b}$ 是沿着 Q 的基方向的每个 $\boldsymbol{b}$ 的分量之和！$\widehat{\boldsymbol{x}}_i=\boldsymbol{q}_i^T\boldsymbol{b}$, $\boldsymbol{q}_i\widehat{\boldsymbol{x}}_i$ **是向量 b 在第 i 个基上的投影！故每个基上的投影加起来便是原向量 b！（用于 QR 分解的理解！）**
$QQ^{\mathrm{T}}=I$ 是傅里叶变换基础，这些变换将向量或者函数分解成垂直片断，再通过上式利用逆变换将这些片段放到一起组成向量或者函数 f。

4.5.2 施密特正交化

思想：新向量减去它在已有的正交向量上的投影向量，即求误差 e，这就是新的正交向量！最后对每个正交向量除以自己的长度，进行归一化！

4.5.3 $A_{m\times n}=Q_{m\times m}R_{m\times n}$ 分解

具体表述以及推导过程参考`https://blog.csdn.net/u010945683/article/details/45972819`(**QR 分解解析**)，里面将从施密特正交化求正交基的方法中将转换前和转换后的向量用线性组合表示出来，然后再写成矩阵乘法形式，清晰易懂。

将上述的 A 与 Q(规范正交矩阵) 用 R(上三角矩阵) 联系起来，**A—) 00009 施密特正交化—>Q**, 每一步 $\boldsymbol{a}_1,\ldots,\boldsymbol{a}_k$ 都是 $q_1,\ldots,q_k$ 的线性组合，后面的 $\boldsymbol{q}_{k+1}\ldots$ 均不涉及。正交化过程中用到的转换矩阵便是 $T,R=T^{-1}$。

$$A=QR=\left[\begin{array}{ccc}\boldsymbol{a} & \boldsymbol{b} & \boldsymbol{c}\end{array}\right]=\left[\begin{array}{ccc}\boldsymbol{q}_1 & \boldsymbol{q}_2 & \boldsymbol{q}_3\end{array}\right]\left[\begin{array}{ccc}\boldsymbol{q}_1^T\boldsymbol{a} & \boldsymbol{q}_1^T\boldsymbol{b} & \boldsymbol{q}_1^T\boldsymbol{c} \\ & \boldsymbol{q}_2^T\boldsymbol{b} & \boldsymbol{q}_2^T\boldsymbol{c} \\ & & \boldsymbol{q}_3^T\boldsymbol{c}\end{array}\right]$$

简而言之，$A=QR$ 就是 Gram-Schmidt 正交化的产物, 左乘一个 Q^T 得：$\boldsymbol{R}=\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{A}$，其中 R 为上三角矩阵！(因为 $Q^TQ=I$)

4.5.4 Problem set

4. (a) Any $Q_{m\times n}$ with $n<m$ has $QQ^{\mathrm{T}}\neq I$, 如 $Q=\begin{bmatrix}1&0\\0&1\\0&0\end{bmatrix}$, $QQ^{\mathrm{T}}=$

$$\begin{bmatrix}1&0&0\\0&1&0\\0&0&0\end{bmatrix}\neq I$$

若 A 能 QR 分解，则 $A^{\mathrm{T}}A=R^{\mathrm{T}}R$ 下三角乘上三角。

5 行列式 (方阵) Determinants

5.1 性质 properties

- A 可逆，则 $\boldsymbol{det}(A) \neq 0$
- $det(A^{-1}) = 1/(\det A)$

反对称矩阵 skew-symmetric matrix $A^{\mathrm{T}} = -A$

5.2 排列和代数余子式 Permutations and Cofactors

求行列式的三种方法

- 消元法 Pivot formula 将矩阵化为三角矩阵，即进行 LU 分解。
- 逆序数法 Big formula 每个 $n \times n$ 矩阵的行列式都有 $n!$ 个项，每个项里的元素均为不同行不同列的元素，元素的行列不能有重合!
- 代数余子式法 Cofactors $\det A = a_{11}(a_{22}a_{33} - a_{23}a_{32}) + a_{12}(a_{23}a_{31} - a_{21}a_{33}) + a_{13}(a_{21}a_{32} - a_{22}a_{31})$，括号里的便是代数余子式。$\det A = a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in}$，$C_{ij} = (-1)^{i+j}\det M_{ij}$。

5.3 Cramer's Rule, Inverses, and Volumes

Cramer's Rule(可逆方阵): $\begin{bmatrix} & A & \\ & & \end{bmatrix} \begin{bmatrix} x_1 & 0 & 0 \\ x_2 & 1 & 0 \\ x_3 & 0 & 1 \end{bmatrix} = \begin{bmatrix} b_1 & a_{12} & a_{13} \\ b_2 & a_{22} & a_{23} \\ b_3 & a_{32} & a_{33} \end{bmatrix} =$
B_1

$(\det A)(x_1) = \det B_1 \quad$ or $\quad x_1 = \frac{\det B_1}{\det A}$，$x_2, x_3$ 同理

故若 $det(A) \neq 0, A\boldsymbol{x} = \boldsymbol{b}$ 则有：$x_1 = \frac{\det B_1}{\det A} \quad x_2 = \frac{\det B_2}{\det A} \quad \ldots \quad x_n = \frac{\det B_n}{\det A}$

A^{-1} 包含 A 的代数余子式，同样利用 cramer 法则解决 $AA^{-1} = I$:

过程：$AA^{-1} = A\begin{bmatrix} \boldsymbol{col1} & \boldsymbol{col2} & \boldsymbol{col3} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$，将其中每一个等式单独拿出来利用 Cramer 法则进行计算, 例如 $A\boldsymbol{x} = (1, 0, 0)$

$$\begin{vmatrix} 1 & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & a_{32} & a_{33} \end{vmatrix} \quad \begin{vmatrix} a_{11} & 1 & a_{13} \\ a_{21} & 0 & a_{23} \\ a_{31} & 0 & a_{33} \end{vmatrix} \quad \begin{vmatrix} a_{11} & a_{12} & 1 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & 0 \end{vmatrix}$$

所以每个 $|B_j|$ 都是 A 的代数余子式！故有：$(A^{-1})_{ij} = \frac{C_{ji}}{\det A}$ and $A^{-1} = \frac{C^{\mathrm{T}}}{\det A}$

二阶行列式是平行四边形的面积，是三角形面积的一半！

通过面积 $s = \boldsymbol{ab}\sin\theta$ 证明。

矩阵的行列式其实是该矩阵变换下的图形面积或体积的伸缩因子。
(Jacobian matrix 雅可比矩阵)

叉乘 Cross Product

叉乘只用于三维空间中，结果为向量：$\boldsymbol{u} = (u_1, u_2, u_3)$ $\boldsymbol{v} = (v_1, v_2, v_3)$

$$\boldsymbol{u}\times\boldsymbol{v} = \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix} = (u_2v_3 - u_3v_2)\,\boldsymbol{i} + (u_3v_1 - u_1v_3)\,\boldsymbol{j} + (u_1v_2 - u_2v_1)\,\boldsymbol{k}$$

向量 $\boldsymbol{u}\times\boldsymbol{v}$ 垂直于 $\boldsymbol{u}$ 和 $\boldsymbol{v}$。此外，$|\boldsymbol{u}\times\boldsymbol{v}| = \|\boldsymbol{u}\|\|\boldsymbol{v}\||\sin\theta|$ 在数值上等于由向量 $\boldsymbol{u}$ 和向量 $\boldsymbol{v}$ 构成的平行四边形的面积。

三重积（混合积）triple product

混合积是一个标量 scalar，且它是这三个向量构成的体积！

$$(\boldsymbol{u}\times\boldsymbol{v})\cdot\boldsymbol{w} = \begin{vmatrix} w_1 & w_2 & w_3 \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix} = \begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix}$$

当 $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ 均位于同一平面时，混合积为零。

6 特征值和特征向量 Eigenvalues and Eigenvectors

6.1 特征值 Eigenvalues

$A\boldsymbol{x} = \lambda\boldsymbol{x}$, 其中 $\boldsymbol{x}$ 是特征向量，λ 是特征值（可以为 0，说明特征向量处于零空间中）!
特征向量乘上 A 不改变方向，其他向量会改变方向。但是**其他向量可以用特征向量的线性组合来表示!**
这里引入**马尔科夫矩阵 Markov matrix**: 它的每一列相加为 1，最大特征值为 1，对应特征向量处于一种稳定状态（当 A 乘上 A^k 时,A 的所有列都会达到的一个稳定状态）
$A^{99}\begin{bmatrix} .8 \\ .2 \end{bmatrix}$ is really $\boldsymbol{x}_1 + (.2)\left(\frac{1}{2}\right)^{99}\boldsymbol{x}_2 = \begin{bmatrix} .6 \\ .4 \end{bmatrix} + \begin{bmatrix} \text{very} \\ \text{small} \\ \text{vector} \end{bmatrix}$（这里将 A 的其中一列乘上 A^{99}）

6.1.1 特征值方程

$A\boldsymbol{x} = \lambda\boldsymbol{x} \to (A - \lambda I)\boldsymbol{x} = \mathbf{0}$，也就是说特征向量 $\boldsymbol{x}$ 构成了零空间，说明方程组必有非零解，即行列式为 0。
λ 为特征值当且仅当 $(A - \lambda I)$ 是奇异的（不可逆），即 $\det(A - \lambda I) = 0$。
特征多项式 (characteristic polynomial) $\det(A - \lambda I)$ 仅包含 λ，当矩阵 A 是 $n \times n$ 时，有 n 个特征值（可能重复）。
计算过程省略。。。

6.1.2 行列式和迹 determinant and trace

- $det(A) = \lambda_1\lambda_2\ldots\lambda_n$
- $trace(A) = \lambda_1 + \lambda_2 + \cdots + \lambda_n = a_{11} + a_{22} + \cdots + a_{nn}$
- 三角阵对角线上的元素即为它的特征值!

上述前两条证明需要用到推广韦达定理,行列式展开式,具体证明见 chrome 收藏:mit 线性代数

6.1.3 AB 和 A+B 的特征值

错误证明：$AB\boldsymbol{x} = A\beta\boldsymbol{x} = \beta A\boldsymbol{x} = \beta\lambda\boldsymbol{x}$，这是**因为特征向量 $\boldsymbol{x}$ 不一定同时是 A 和 B 的特征向量!**

同理，A+B 的特征值也不一定是 $\lambda+\beta$。上述情况只有在 $\boldsymbol{x}$ 同时为 A，B 的特征向量时才成立!

当且仅当 $AB = BA$ 时，A,B 同时共享相同的 n 个特征向量!

对称矩阵的特征值为实数，反对称矩阵的特征值为虚数!

6.2 对角化, 相似

6.2.1 对角化 Diagonalization

若矩阵 A 有 n 个线性无关的特征向量，将它们放入特征向量矩阵 X 中，则有 $X^{-1}AX = \Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$，大写 lambda Λ 的对角线是特征值!

解释 $AX = X\Lambda$：$AX = A\begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} = \begin{bmatrix} \lambda_1 x_1 & \cdots & \lambda_n x_n \end{bmatrix} = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} = X\Lambda$

故 $AX = X\Lambda \rightarrow X^{-1}AX = \Lambda \quad \text{or} \quad A = X\Lambda X^{-1}$。$A$ 和 Λ 的特征值相同，特征向量不同!

$A^k = (X\Lambda X^{-1})(X\Lambda X^{-1})\dots(X\Lambda X^{-1}) = X\Lambda^k X^{-1}$

只有当 **A 有 n 个线性无关的特征向量时，才能对角化**。P306(316) 利用构造线性组合证明了不同特征值对应的特征向量线性无关!

6.2.2 相似矩阵：相同特征值

假设 Λ 不变，改变特征矩阵 X，这样会得到所有能相似对角化成 Λ。

相似定义：存在可逆矩阵 B，若 $A = BCB^{-1}$，则称 A 与 C 相似（C 不一定是对角阵）。

事实上，若 A，C **相似**，则它们一定**有相同的特征值**!

Proof: 假设 $C\boldsymbol{x} = \lambda\boldsymbol{x}$，则对于 $A = BCB^{-1}$ 可以找到特征向量 $B\boldsymbol{x}$ 使得:

$(BCB^{-1})(B\boldsymbol{x}) = BC\boldsymbol{x} = B\lambda\boldsymbol{x} = \lambda(B\boldsymbol{x})$，即 $A(B\boldsymbol{x}) = \lambda(B\boldsymbol{x})$。

6.2.3 若尔当型标准型 Jordan Form

Jordan Block: $J_i = \begin{pmatrix} \lambda_i & 1 & & & & \\ & \lambda_i & 1 & & & \\ & & \lambda_i & \ddots & & \\ & & & \ddots & 1 & \\ & & & & \lambda_i & 1 \\ & & & & & \lambda_i \end{pmatrix}_{r_i \times r_i}$，对角线上是重复的特征值，每个特征值上面有 1，其余元素都是 0.

对于任一 n 阶复数矩阵 A，都与一个若尔当型矩阵 J 相似，其中 $J = \begin{bmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_{r_i} \end{bmatrix}$，且 $1 + 2 + \cdots + r_i = n$，每一个若尔当块对应一个特征向量，所以 A 的若尔当块数 = 特征向量数。(若有 n 个特征向量，n 个分块，即 $r_i = n$，则说明矩阵 A 可以对角化即 $J = \Lambda$)

在不计 Jordan 块次序的前提下，A 的 Jordan 标准型矩阵是唯一确定的。

Jordan Form 矩阵是指最接近对角阵但是又无法对角化的矩阵! 比如 $\begin{bmatrix} 4 & 1 \\ 0 & 4 \end{bmatrix}$ 它不可能对角化 (因为 $\Lambda = 4I$ 利用谱定理得到的总是 $4I$)。

同样有两个不能对角化的矩阵，它们有相同特征值 0，相同数量的特征向量:

$\left[\begin{array}{ccc|c} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ \hline 0 & 0 & 0 & 0 \end{array}\right]$ 和 $\left[\begin{array}{cc|cc} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$，**尽管它们特征值相同，但是由于它们的 Jordan block 分块不同，所以不相似!**

6.2.4 Fibonacci Numbers

斐波那契数列 $0, 1, 1, 2, 3, 5, 8, 13, \ldots$ 有递推式 $F_{k+2} = F_{k+1} + F_k$，可以利用矩阵方程 $\mathbf{u}_{k+1} = A\boldsymbol{u}_k$ 来快速计算。

设 $\boldsymbol{u}_k = \begin{bmatrix} F_{k+1} \\ F_k \end{bmatrix}$ 递推式 $\begin{matrix} F_{k+2} = F_{k+1} + F_k \\ F_{k+1} = F_{k+1} \end{matrix}$ 可以表示为 $\boldsymbol{u}_{k+1} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \boldsymbol{u}_k$

所以每求一步新的斐波那契数就乘上一个矩阵 $A=\begin{bmatrix}1&1\\1&0\end{bmatrix}$，**差分方程** $\boldsymbol{u}_{100}=A^{100}\boldsymbol{u}_0$ 即 $u_0=\begin{bmatrix}1\\0\end{bmatrix}$, $u_1=\begin{bmatrix}1\\1\end{bmatrix}$, $u_2=\begin{bmatrix}2\\1\end{bmatrix}$, $u_3=\begin{bmatrix}3\\2\end{bmatrix}$, $\ldots$, $u_{100}=\begin{bmatrix}F_{101}\\F_{100}\end{bmatrix}$

为了简化矩阵乘法，使其能快速求解，**利用 A 的特征值：**

$A-\lambda I=\begin{bmatrix}1-\lambda&1\\1&-\lambda\end{bmatrix}\rightarrow det(A-\lambda I)=\lambda^2-\lambda-1=0\rightarrow$

$Eigenvalues:\lambda_1=\frac{1+\sqrt{5}}{2}\approx 1.618\quad \lambda_2=\frac{1-\sqrt{5}}{2}\approx -0.618$

$Eigenvectors:\boldsymbol{x}_1=(\lambda_1,1)\quad \boldsymbol{x}_2=(\lambda_2,1)$ 然后将 $\boldsymbol{u}_0$ **表示为两个特征向量的线性组合**：$\boldsymbol{u}_0=\frac{\boldsymbol{x}_1-\boldsymbol{x}_2}{\lambda_1-\lambda_2}\rightarrow\begin{bmatrix}1\\0\end{bmatrix}=\frac{1}{\lambda_1-\lambda_2}\left(\begin{bmatrix}\lambda_1\\1\end{bmatrix}-\begin{bmatrix}\lambda_2\\1\end{bmatrix}\right)$

再计算 $\boldsymbol{u}_{100}=A^{100}\boldsymbol{u}_0$，即将 $\boldsymbol{u}_0$ 乘上 A^{100}，就有：$\boldsymbol{u}_{100}=\frac{(\lambda_1)^{100}\boldsymbol{x}_1-(\lambda_2)^{100}\boldsymbol{x}_2}{\lambda_1-\lambda_2}$。

而且 $\lambda_2^{100}\approx 0$，所以第 100 个数为 $\frac{\lambda_1^{100}-\lambda_2^{100}}{\lambda_1-\lambda_2}$ 近似于 $\frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^{100}$。

每个 F_k 都是整数，且比例 F_101/F_100 是非常接近 $\frac{1+\sqrt{5}}{2}$ 的。

6.2.5 矩阵的幂

矩阵的对角化分解非常适合用来计算矩阵的幂，例如上一小节的差分方程:$\boldsymbol{u}_{100}=A^{100}\boldsymbol{u}_0$ 中，$A^k\boldsymbol{u}_0=(X\Lambda X^{-1})\cdots(X\Lambda X^{-1})\boldsymbol{u}_0=(X\Lambda^kX^{-1})\boldsymbol{u}_0$。

下面分步解释特征值在其中是如何起作用的 (详细说明 P310(320))：

1. 写成线性组合用特征向量表出 $\boldsymbol{u}_0=c_1\boldsymbol{x}_1+\cdots+c_n\boldsymbol{x}_n,\quad \boldsymbol{u}_0=X\boldsymbol{c}_{n\times 1}\quad \boldsymbol{c}_{n\times 1}=X^{-1}\boldsymbol{u}_0$

2. 每个特征向量 $\boldsymbol{x}_i$ 乘上 $(\lambda_i)^k$，即有 $\Lambda^kX^{-1}\boldsymbol{u}_0$

3. 在乘上 X 得到 $c_i(\lambda_i)^k\boldsymbol{x}_i$，全部加起来便是 $X\Lambda^kX^{-1}\boldsymbol{u}_0$

所以最终解的形式：$\boldsymbol{u}_k=c_1(\lambda_1)^k\boldsymbol{x}_1+\cdots+c_n(\lambda_n)^k\boldsymbol{x}_n=\sum c_i(\lambda_i)^k\boldsymbol{x}_i$。

6.2.6 不可对角化的矩阵 Nondiagonalizable Matrices

也就是有重复的特征值 λ，但是对应 λ 的**特征向量个数 GM 不等于特征值重复的次数 AM!**

欧拉公式 Euler's formula: $e^{i\theta}=cos\theta+isin\theta$

- A,B 均为方阵时，AB,BA 有相同特征值值
- 当 $A_{m\times n}, B_{n\times m}$ 时，AB，BA 有相同的非零特征值

证明见百度 (利用特征值公式)~

6.3 微分方程组 Systems of Differential Equations

A 的相似对角化不仅适用于计算 A^k，也适用于计算微分方程 $d\boldsymbol{u}/dt = A\boldsymbol{u}$。此章节主要是：**将线性常系数微分方程转换为线性代数！**
普通微分方程 $\frac{du}{dt} = u \to u(t) = \boldsymbol{Ce^t}$， $\frac{du}{dt} = \lambda u \to u(t) = \boldsymbol{Ce^{\lambda t}}$，当 time t=0 时，有 $u(0) = c$，说明了解的系数 $\boldsymbol{C}$。
但以上只是从一个方程出发，**若有多个微分方程联立共同确定几个具有同一自变量的函数，这构成了微分方程组（耦合），它的解是关于共同自变量的函数**。**未知数 (函数) 可以写成向量**，从给定的初始条件 $\boldsymbol{u(0)}$ 开始。
例如：$\begin{cases} \frac{du_1}{dt} = -u_1 + 2u_2 \\ \frac{du_2}{dt} = u_1 - 2u_2 \end{cases}$ 写成向量形式：$\frac{d}{dt}\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} -1 & 2 \\ 1 & -2 \end{bmatrix}\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \to \frac{d\boldsymbol{u}}{dt} = A\boldsymbol{u}$

即 $\frac{d\boldsymbol{u}}{dt} = \boldsymbol{u}$ 从向量 $\boldsymbol{u}(0) = \begin{bmatrix} u_1(0) \\ \cdots \\ u_n(0) \end{bmatrix}$ 开始！利用 $A\boldsymbol{x} = \lambda\boldsymbol{x}$ 找到纯指数解 $e^{\lambda t}\boldsymbol{x}$。

6.3.1 du/dt=Au 的解

$d\boldsymbol{u}/dt = A\boldsymbol{u}$ 的解纯指数解为 $\boldsymbol{u}(t) = e^{\lambda t}\boldsymbol{x}$，其中 λ 是 A 的特征值，$\boldsymbol{x}$ 是 A 的特征向量。$\frac{d\boldsymbol{u}}{dt} = \lambda e^{\lambda t}\boldsymbol{x}$ 与 $A\boldsymbol{u} = Ae^{\lambda t}\boldsymbol{x}$ 一致。
求解过程：

1. 将初始状态 $\boldsymbol{u}(0)$ 表示为 A 特征向量的线性组合 $c_1\boldsymbol{x}_1 + \cdots + c_n\boldsymbol{x}_n$(需要 A 能相似对角化，即有 n 个线性无关的特征向量)
2. 对每个特征向量 $\boldsymbol{x}_i$ 乘上对应的 $e^{\lambda_i t}$
3. 所以微分方程组 $\frac{d\boldsymbol{u}}{dt} = A\boldsymbol{u}$ 的通解为：$\boldsymbol{u}(t) = c_1e^{\lambda_1 t}\boldsymbol{x}_1 + \cdots + c_ne^{\lambda_n t}\boldsymbol{x}_n$

若 A 有重复特征值且 GM！=AM，即特征向量个数不等于重复数，则通解中会包含 $te^{\lambda t}\boldsymbol{x}$。

6.3.2 二阶微分方程的解

高数中：$m\frac{d^2y}{dt^2}+b\frac{dy}{dt}+ky=0 \quad \text{becomes} \quad (m\lambda^2+b\lambda+k)\,e^{\lambda t}=0$，且利用 $y=e^{\lambda t}$ 来解微分方程，即 $y_1=e^{\lambda_1 t}$ and $y_2=e^{\lambda_2 t}$。($\lambda_1\lambda_2$ 是特征方程的两个根！)

线代中：利用特征值和特征矩阵来解方程，先对该二阶微分方程 $m\frac{d^2y}{dt^2}+b\frac{dy}{dt}+ky=0$ 构造方程组, 设 $m=1$(不设也没关系),$\boldsymbol{u}=(y,y')$ **降阶处理**：

1. $\begin{cases} dy/dt=y' \\ dy'/dt=-ky-by' \end{cases} \quad \rightarrow \frac{d}{dt}\begin{bmatrix} y \\ y' \end{bmatrix}=\begin{bmatrix} \mathbf{0} & \mathbf{1} \\ -\boldsymbol{k} & -\boldsymbol{b} \end{bmatrix}\begin{bmatrix} y \\ y' \end{bmatrix}=A\boldsymbol{u}$,
   第一个式子是不必要的（为了构建矩阵），第二个式子把 y'' 连接到 y' 和 y，所以宏观看来这是 $\boldsymbol{u}'$ 转为 $\boldsymbol{u}$。

2. $A-\lambda I=\begin{bmatrix} -\lambda & 1 \\ -k & -b-\lambda \end{bmatrix}$ 有行列式 $\lambda^2+b\lambda+k=0$，这与高数中的一致！求出特征值 λ_1,λ_2 特征向量 $\boldsymbol{x}_1=\begin{bmatrix} 1 \\ \lambda_1 \end{bmatrix} \quad \boldsymbol{x}_2=\begin{bmatrix} 1 \\ \lambda_2 \end{bmatrix}$

3. 得到方程组通解：$u(t)=c_1e^{\lambda_1 t}\begin{bmatrix} 1 \\ \lambda_1 \end{bmatrix}+c_2e^{\lambda_2 t}\begin{bmatrix} 1 \\ \lambda_2 \end{bmatrix}$

4. 方程组通解第一个元素：$y=c_1e^{\lambda_1 t}+c_2e^{\lambda_2 t}$ 就是该二阶微分方程的通解。

5. 第二个元素便是 dy/dt。

求解 3 阶…n 阶微分方程同理, 仿照上述方法构造矩阵即可。

6.3.3 2×2 矩阵的稳定性

微分方程 $d\boldsymbol{u}/dt=A\boldsymbol{u}$ 的通解 $\boldsymbol{u}(t)$，当 $t\rightarrow\infty$ 时，通解是趋于 0? 通解 $\boldsymbol{u}(t)$ 是从 $e^{\lambda t}\boldsymbol{x}$ 得来的，所以：

- 若特征值 λ 是实数，则 $\lambda<0$ 才有 $e^{\lambda t}\rightarrow 0$
- 若特征值 λ 是复数，则 $\lambda=r+is$ 的实部 $r<0$ 才有 $e^{\lambda t}\rightarrow 0$。
  注意：$e^{\lambda t}$ 分解成 $e^{rt}e^{ist}$，因为 $e^{ist}=\cos st+i\sin st$(欧拉公式) 有 $|e^{ist}|^2=\cos^2 st+\sin^2 st=1$，所以 e^{ist} 固定为 1，不影响解的值!

对于 2×2 矩阵来说，A 所有特征值实部均为负数方程的解才会趋于稳定，可以用矩阵的迹 trace(A)<0 和行列式 det(A)>0 来判断。

6.3.4 解耦与矩阵指数 uncouple and exponential of a matrix

一般来说，微分方程组中各个同自变量的函数（即未知数）是耦合的，利用 A 的特征值和特征向量进行**对角化**就可以进行**解耦**。
首先定义**矩阵指数** e^{At}, 把指数展开成幂级数的方式，**当指数是矩阵时也一样**！(特征向量矩阵 X，且 $A=X\Lambda X^{-1}$):

1. 幂级数公式：$e^{\mathrm{x}}=1+\mathrm{x}+\frac{x^2}{2}+\frac{x^3}{6}+\cdots+\frac{x^n}{\mathrm{n}!}\cdots$ （收敛域为全体实数)
2. 扩展到矩阵计算中，I 代替 1，矩阵代替 x：
$e^{At}=I+At+\frac{(At)^2}{2}+\frac{(At)^3}{6}+\ldots\cdots\cdots+\frac{(At)^n}{n!}+\cdots\cdots$
3. 对角化 A：$e^{At}=I+X\Lambda X^{-1}t+\frac{1}{2}\left(X\Lambda X^{-1}t\right)\left(X\Lambda X^{-1}t\right)+\cdots$
$=X\left[I+\Lambda t+\frac{1}{2}(\Lambda t)^2+\cdots\right]X^{-1}$
4. 对比 e 的幂级数展开式可知：$e^{At}=Xe^{\Lambda t}X^{-1}$

矩阵分析中有证明矩阵的泰勒展开式与函数的泰勒展开式的形式一样！

$$\Lambda=\begin{bmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_n\end{bmatrix}\quad e^{\Lambda t}=\begin{bmatrix}e^{\lambda_1 t} & & \\ & \ddots & \\ & & e^{\lambda_n t}\end{bmatrix}$$

解耦: 在之前的 $\frac{d\boldsymbol{u}}{dt}=A\boldsymbol{u}$ 中，有 u_1,u_2 耦合,

1. 设 $\boldsymbol{u}=X\boldsymbol{v}$, $\frac{d\boldsymbol{u}}{dt}=X\frac{d\boldsymbol{v}}{dt}=A\boldsymbol{u}=AX\boldsymbol{v}$
2. 提取出 $X\frac{d\boldsymbol{v}}{dt}=AX\boldsymbol{v}$，左右两边同时左乘上 X^{-1} 得到：
$\frac{d\boldsymbol{v}}{dt}=X^{-1}AX\boldsymbol{v}=\Lambda v$
3. 由此得到关于 $\boldsymbol{v}$ 的对角化方程组，不耦合—$\frac{\mathrm{d}v_1}{\mathrm{dt}}=\lambda_1 v_1,\frac{\mathrm{d}v_2}{\mathrm{dt}}=\lambda_2 v_2$
4. 最终有：$\boldsymbol{v}(\mathrm{t})=e^{\Lambda t}\boldsymbol{v}(0)$ 且 $\boldsymbol{u}(t)=e^{At}\boldsymbol{u}(0)=\mathrm{X}\mathrm{e}^{\Lambda t}X^{-1}\boldsymbol{u}(0)$

$$e^{At}\boldsymbol{u}(0)=Xe^{\Lambda t}X^{-1}\boldsymbol{u}(0)=\begin{bmatrix}\boldsymbol{x}_1 & \cdots & \boldsymbol{x}_n\end{bmatrix}\begin{bmatrix}e^{\lambda_1 t} & & \\ & \ddots & \\ & & e^{\lambda_n t}\end{bmatrix}\begin{bmatrix}c_1\\ \vdots\\ c_n\end{bmatrix}$$

$\boldsymbol{u}(0)=c_1\boldsymbol{x}_1+\cdots+c_n\boldsymbol{x}_n=X\boldsymbol{c}$,　　故它与之前提到的解微分方程组三步解

法得到的**通解** $\boldsymbol{u}(t)=\boldsymbol{c}_1 e^{\boldsymbol{\lambda}_1 t}\boldsymbol{x}_1+\cdots+\boldsymbol{c}_n e^{\boldsymbol{\lambda}_n t}\boldsymbol{x}_n$ **是一样的**!
P327(337) example4 解释了**为什么在解一个二阶微分方程式会出现:** e^t **和** te^t,其核心在于 A 有重复特征值且特征向量数目不够,则 $e^{At}=e^{It}e^{(A-I)t}=e^t[I+(A-I)t](e^{(A-I)t}$ 的幂级数展开式是收敛的,因为 $(A-I)^2$ 是零矩阵)
$\begin{bmatrix} y \\ y' \end{bmatrix}=e^t\left[I+\begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix}t\right]\begin{bmatrix} y(0) \\ y'(0) \end{bmatrix}$, $y(t)=e^t y(0)-te^t y(0)+te^t y'(0)$
Problem Set 26 证明了 e^{At} **为什么是可逆的!(所有特征值** $e^{\lambda t}>0$**)**

6.4 对称阵 symmetric matrices

key fact:

- 一个实对称阵只有实特征值
- 特征向量可以选择变成规范正交的 (orthonormal)(通过单位化), X 的每个列向量都是规范正交的话,则 X 变成了正交矩阵 (orthogonal)Q, 有 $Q^{-1}=Q^T$
- 实对称阵均可对角化

谱定理 Spectral Theorem: 每个对称阵都有分解:$S=Q\Lambda Q^{-1}=Q\Lambda Q^{\mathrm{T}}$
with $\quad Q^{-1}=Q^{\mathrm{T}}$, Λ 是实特征值,Q 是正交矩阵。
还证明了为什么实对称阵的特征值也是实数?为什么实对称阵的特征向量正交?
$S=Q\Lambda Q^{\mathrm{T}}=\begin{bmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 \end{bmatrix}\begin{bmatrix} \lambda_1 & \\ & \lambda_2 \end{bmatrix}\begin{bmatrix} \boldsymbol{q}_1^{\mathrm{T}} \\ \boldsymbol{q}_2^{\mathrm{T}} \end{bmatrix}$
$S=Q\Lambda Q^{\mathrm{T}}=\lambda_1 q_1 q_1^{\mathrm{T}}+\cdots+\lambda_n q_n q_n^{\mathrm{T}}$, 利用特征向量的规范正交性有:$S\boldsymbol{q}_i=\left(\lambda_1\boldsymbol{q}_1\boldsymbol{q}_1^{\mathrm{T}}+\cdots+\lambda_n\boldsymbol{q}_n\boldsymbol{q}_n^{\mathrm{T}}\right)\boldsymbol{q}_i=\lambda_i\boldsymbol{q}_i$

6.4.1 实对称阵的复特征值

对于任意实对称阵:$Sx=\lambda x\rightarrow S\overline{x}=\overline{\lambda}\overline{x}$,若是实对称阵则前两式相同,若是非对称阵则可有复特征值、复特征向量则前两式不同。
对于实数阵,复数的特征值和特征向量都是共轭对,$\lambda=a+ib\quad\overline{\lambda}=a-ib$,
If $A\boldsymbol{x}=\lambda\boldsymbol{x}\quad$ then $\quad A\overline{\boldsymbol{x}}=\overline{\lambda}\overline{\boldsymbol{x}}$

6.4.2 特征值 vs 主元 (行阶梯)

对于**对称阵**来说，**正特征值的数目 = 正主元数目** (也有特征值乘积 =det(S)= 主元乘积)

特殊情况：对称阵 $\boldsymbol{S}$ 的所有特征值 $\lambda_i > 0$ 当且仅当它的所有主元 pivots 是正的 (大于零)。该特殊情况在 6.5 正定矩阵中很重要！

举例证明:$S = \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix}$, 进行 LU 分解 $\rightarrow S = LU = LDL^{\mathrm{T}} = \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & \\ & -8 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix}$

S 的特征值为 4，-2。当 $L \rightarrow I$ 时，$S \rightarrow D$ 即 $D = IDI^T$ 且特征值为 1 和-8，这便是 S 的主元 pivot。

因为 S 在往 D 转变的过程中，特征值没有越过 0，所以特征值的符号一直没有改变！故**特征值的符号与主元符号相同！**

6.4.3 所有对称阵均可对角化

主要说明的是当出现重复特征值时，有没有对应数量的线性无关的特征向量来保证可对角化性？对称阵可以！

利用**舒尔定理 (Schur's Theorem)**：任意方阵 $A = QTQ^{-1}$ 且 T 是上三角型，Q 是正交矩阵，$\overline{Q}^{\mathrm{T}} = Q^{-1}$，所以 $T = Q^{\mathrm{T}}SQ$，当 $S^{\mathrm{T}} = S$ 时，上三角型 T 是对称的，故 T 一定是对角阵 Λ。

舒尔定理书里未证明！

惯性定理 Law of Inertia: S 与 $A^{\mathrm{T}}SA$ 一致 (A 为方阵且可逆)，且它们对应正，负，零特征值数量相同！

6.5 正定矩阵 Positive Definite Matrices

正定矩阵: 对称阵，且所有特征值 $\lambda > 0$。

正定，半定，负定

正定矩阵的判别**两种**方法：

- 矩阵左上行列式从 $1 \times 1 \rightarrow n \times n$ 均大于 0
- 矩阵的所有主元 (行阶梯) 均大于 0

6.5.1 基于能量的定义 energy-based definition

$S\boldsymbol{x}=\lambda\boldsymbol{x}, \lambda>0$, $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}=\|\boldsymbol{x}\|^2$, 所以 $\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x}=\lambda\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}$ **对任何非零向量 x 都是大于 0!** 而数 $x^{\mathrm{T}}Sx$ 在系统中称为能量, 如此便定义了正定矩阵!

定义: 对任意非零向量 $\boldsymbol{x}\neq 0$ 有 $\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x}>0$, 则 S 是正定的!

若 S，T 是正定的，则 S+T 也是正定的。

命题: $S=A^TA$ 是方阵且对称 (A 任意形状)，若 A 列向量线性无关, 则 S 是正定的!

证明: $\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x}=\boldsymbol{x}^{\mathrm{T}}A^{\mathrm{T}}A\boldsymbol{x}=(A\boldsymbol{x})^{\mathrm{T}}(A\boldsymbol{x})=\|A\boldsymbol{x}\|^2$。A 的列线性无关保证了当 $\boldsymbol{x}\neq 0$ 时 $A\boldsymbol{x}$ 不为 0。

Cholesky factorization(LU 三角分解在实对称阵条件下的变形): 实对称阵 $S=LU=LDL^T=L\sqrt{D}\sqrt{D}L^T=L\sqrt{D}(L\sqrt{D})^T=A^TA$, 且 A 的列向量线性无关！A 为上三角矩阵!

6.5.2 正半定矩阵 Positive Semidefinite Matrices

正半定矩阵的所有 $\lambda\geq 0$, 最小的特征值为 0，且对任意非零向量 $\boldsymbol{x}$ 有 $\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x}\geq 0$。该矩阵进行 A^TA 分解后，A 的列向量总是线性相关!

6.5.3 椭圆 The Ellipse $ax^2+2bxy+cy^2=1$

这里设 S 是 2×2 的对称阵，且正定
$\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x}=\begin{bmatrix} x & y \end{bmatrix}\begin{bmatrix} a & b \\ b & c \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}=ax^2+2bxy+cy^2=1$ 是一个斜椭圆。若要使其与坐标轴 X, Y 对齐, 则应利用主轴定理 (principal axis theorem)$S=Q\Lambda Q^{\mathrm{T}}$ 来生成对齐的椭圆!
$\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x}=\boldsymbol{x}^{\mathrm{T}}Q\Lambda Q^{\mathrm{T}}\boldsymbol{x}=(Q^{\mathrm{T}}\boldsymbol{x})^T\Lambda Q^{\mathrm{T}}\boldsymbol{x}=\boldsymbol{X}^{\mathrm{T}}\Lambda\boldsymbol{X}=1$
原 S 中倾斜椭圆的轴沿着 S 的特征向量，变换后椭圆的轴沿着 Λ 的特征向量！长短轴的端点由特征值给出。
椭圆 $\begin{bmatrix} x & y \end{bmatrix}Q\Lambda Q^{\mathrm{T}}\begin{bmatrix} x \\ y \end{bmatrix}=\begin{bmatrix} X & Y \end{bmatrix}\Lambda\begin{bmatrix} X \\ Y \end{bmatrix}=\lambda_1X^2+\lambda_2Y^2=1$
若 S 有特征值为负，则 $\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x}=1$ 变成一个双曲线。
若 S 是 $n\times n$ 的，则 $\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x}=1$ 是 $\mathbf{R}^n$ 中的 “椭球体”!

6.5.4 具体应用：求二元函数的无条件极值

高数中可知无条件极值有判别法：$AC-B^2$, A,C,B 分别对应 F 对 x,F 对 y，F 对 xy(yx) 的二阶导。

转为矩阵 $S=\begin{bmatrix} \partial^2F/\partial x^2 & \partial^2F/\partial x\partial y \\ \partial^2F/\partial y\partial x & \partial^2F/\partial y^2 \end{bmatrix}$

则上述判别公式可以转述为：**若 S 是正定的且 $\partial \boldsymbol{F}/\partial \boldsymbol{x}=\partial \boldsymbol{F}/\partial \boldsymbol{y}=\mathbf{0}$**，则 $F(x,y)$ 有最小值。

若是三元函数的极小值：先有 $\frac{\partial F}{\partial x}=\frac{\partial F}{\partial y}=\frac{\partial F}{\partial z}=0$ 成立

再有 $S=\begin{bmatrix} F_{xx} & F_{xy} & F_{xz} \\ F_{yx} & F_{yy} & F_{yz} \\ F_{zx} & F_{zy} & F_{zz} \end{bmatrix}$ 是正定的!

6.5.5 problem set

命题：当 S,T 均为正定矩阵，ST 未必是对称的，但是 ST 的特征值依然是正的!

证明：

广义特征值问题：$K\boldsymbol{x}=\lambda M\boldsymbol{x}$ 有 $ST=M^{-1}K$，则 $ST\boldsymbol{x}=\lambda\boldsymbol{x}\rightarrow(T\boldsymbol{x})^{\mathrm{T}}ST\boldsymbol{x}=(T\boldsymbol{x})^{\mathrm{T}}\lambda x$. Then $\lambda=\boldsymbol{x}^{\mathrm{T}}T^{\mathrm{T}}ST\boldsymbol{x}/\boldsymbol{x}^{\mathrm{T}}T\boldsymbol{x}>0$。(利用了 S，T 的正定性)

6.6 更正：傅里叶级数 fourier series 与函数空间的内积—在 10.5 中

假设一个 n 维空间的一组完整基是 n 个标准正交向量 $\boldsymbol{q}_1,\boldsymbol{q}_2,\ldots,\boldsymbol{q}_n$，那么在此空间中任意向量 $\boldsymbol{v}$ 可以由这组基线性表示：$\boldsymbol{v}=\boldsymbol{x}_1\boldsymbol{q}_1+\boldsymbol{x}_2\boldsymbol{q}_2+\ldots+\boldsymbol{x}_n\boldsymbol{q}_n$。

矩阵化：$Q\boldsymbol{x}=\boldsymbol{v}\rightarrow\boldsymbol{x}=Q^{-1}\boldsymbol{v}$，因为 $Q^T=Q^{-1}$，所以 $\boldsymbol{x}=Q^{-1}\boldsymbol{v}=Q^T\boldsymbol{v}$，所以向量 $\boldsymbol{v}$ 的各个分量为 $x_i=\boldsymbol{q}_i^T\boldsymbol{v}$(行乘列)。

函数正交

对比向量内积的形式，因为函数在其定义域是连续的，对应两个连续函数而言，**函数内积**就是在其定义域 [a,b] 上对 $f(x)\cdot g(x)$ 的积分：$\int_a^b f(x)\cdot g(x)dx$ 记作函数的内积！若函数内积为零则称这两个函数在定义域上**正交**!

fourier 级数:$f(x)=a_0+a_1\cos x+b_1\sin x+a_2\cos 2x+b_2\sin 2x+\ldots$, 作用在

函数空间上，$f(x)$ 相当于上述的向量 $\boldsymbol{v}$，再有**标准正交基** $1, \cos x, \sin x, \cos 2x, \sin 2x \ldots$
代替 $\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_n$
所以系数 $a_0, a_1, b_1 \ldots\ldots$ 对应向量 $\boldsymbol{v}$ 中的分量 $x_1, x_2, \ldots !$
求解系数 a1：仿照求解向量分量，将向量与对应的正交基做内积！这里将
$f(x)$ 与 $\cos x$ 做内积：$\int_0^{2\pi} f(x) \cos x dx = \int_0^{2\pi} (a_0 + a_1 \cos x + b_1 \sin x + a_2 \cos 2x + b_2 \sin 2x) \cos x dx$，
因为 $\int_0^{2\pi} a_1 \cos x^2 dx = a_1 \pi$，所以 $a_1 = \frac{\int_0^{2\pi} f(x) \cos x dx}{\pi}$。

7 Singular Values Decomposition

提前介绍后面要用到的性质。

[矩阵 $\boldsymbol{A^TA}$ 和 $\boldsymbol{AA^T}$(任意维度) 的性质及证明]`https://blog.csdn.net/Europe233/article/details/86720864`

性质：$\boldsymbol{A^TA}$ 和 $\boldsymbol{AA^T}$ 具有相同的非零特征值

证明 1:`https://blog.csdn.net/Europe233/article/details/86727078`

证明 2：

1. 假设 $\boldsymbol{x} \neq \boldsymbol{0}$，$\boldsymbol{A}^T\boldsymbol{Ax} = \lambda\boldsymbol{x}$，等式左右两边同时左乘一个 $\boldsymbol{A}$ 得 $\rightarrow (\boldsymbol{AA}^T)\boldsymbol{Ax} = \lambda\boldsymbol{Ax}$；

2. 同理，假设 $\boldsymbol{x} \neq \boldsymbol{0}$，$\boldsymbol{AA}^T\boldsymbol{x} = \lambda\boldsymbol{x}$，左右两边同时左乘一个 $\boldsymbol{A^T}$，得 $\rightarrow (\boldsymbol{A}^T\boldsymbol{A})\boldsymbol{A}^T\boldsymbol{x} = \lambda\boldsymbol{A^Tx}$。

3. 经检验，维度是匹配的。这说明在 $\lambda \neq 0$ 时上述式子成立，也即 $\boldsymbol{A^TA}$ 和 $\boldsymbol{AA^T}$ 具有相同的非零特征值。

7.1 SVD 的目的

从空间变换的角度出发

- 对称阵的谱分解 (特征值分解)：通过矩阵 $\boldsymbol{A}$ 将一组基映射到同一组基。

- SVD 的目的是：通过任意矩阵 $\boldsymbol{A}_{m\times n}$ 将一组单位正交基映射到另一组单位正交基。

SVD 的两种表示 (下面的向量和矩阵都没有加粗，自行辨识)：

$$Av_i = \sigma_i u_i$$

$$\begin{array}{c}(m \text{ by } n)(n \text{ by } r)\\ AV_r = U_r\Sigma_r \\ (m \text{ by } r)(r \text{ by } r)\end{array} \quad A\begin{bmatrix} v_1 & \cdots & v_r \end{bmatrix} = \begin{bmatrix} u_1 & \cdots & u_r \end{bmatrix}\begin{bmatrix} \sigma_1 & & & \\ & \cdot & & \\ & & \cdot & \\ & & & \sigma_r \end{bmatrix}$$

书上 7.2 的直接给出向量 u, v 是从四个基本子空间的正交单位基而来。这里给出个人理解：

- 直接从目标的表达式 $\boldsymbol{AV} = \boldsymbol{U\Sigma}$(此处 $\boldsymbol{V}, \boldsymbol{U}, \boldsymbol{\Sigma}$ 为方阵) 看出，每个向量 $\boldsymbol{u}_i$ 是矩阵 $\boldsymbol{A}$ 的列向量的线性组合，所以 $\boldsymbol{u}_i$ 处于 $\boldsymbol{A}$ 的**列空间**中!
- 倘若按照 SVD 的目的，从一组单位正交基映射到另一组单位正交基，那么矩阵 $\boldsymbol{U}, \boldsymbol{V}$ 便都是正交矩阵，所以对目标表达式左乘一个 U^T，右乘一个 V^T 然后再转置得到：$\boldsymbol{A}^T\boldsymbol{U} = \boldsymbol{V\Sigma}$。由此便可以按列向量展开，然后看出每个 $\boldsymbol{v}_i$ 向量都是矩阵 A 的行向量的线性组合，即 $\boldsymbol{v}_i$ 处于 $\boldsymbol{A}$ 的**行空间**中。
- 所以书上一开始就直接去 $\boldsymbol{A}$ 的列空间和行空间里面去找规范正交基
- 后面分别找零空间和左零空间的规范正交基是因为要找满足分别与先前两组规范正交基正交的另外两组基，并且恰好分别能将其扩展为方阵!

https://www.zhihu.com/question/22237507、https://blog.csdn.net/zhongkejingwang/article/details/43053513，的推导过程：
假设找到了一组正交的基 $\{v_1, v_2, \ldots, v_n\}$，通过矩阵 A 将这组正交基映射为 $\{Av_1, Av_2, \ldots, Av_n\}$，令映射后的向量两两正交即：$(\mathrm{A}v_i)^T Av_j = v_i^T A^T \mathrm{A}v_j = 0$。根据假设，存在 $v_i^T v_j = 0$，由此看出，如果**选用矩阵** A^TA **这个实对称阵的特征向量作为** v_i **的话，就满足要求**，也即 $v_i^T A^T \mathrm{A}v_j = v_i^T \lambda_j v_j = \lambda_j v_i^T v_j = \lambda_j v_i, v_j = 0$。所以这里便找到了另一组对应的正交基 $\{Av_1, Av_2, \ldots, Av_n\}$!

然后将映射后的正交基单位化：

- 因为 $(Av_i)^T Av_i = v_i^T (A^TA) v_i = v_i^T \lambda_i v_i = \lambda_i (v_i$ 是正交单位向量)
- 所以有 $||Av_i||^2 = \lambda_i \geq 0$
- 取单位向量：$u_i = \dfrac{Av_i}{||Av_i||} = \dfrac{1}{\sqrt{\lambda_i}} Av_i$

由此可得，$Av_i = \sigma_i u_i, \sigma_i($ 奇异值 $) = \sqrt{\lambda_i}, 0 \leq i \leq \mathrm{k}, \mathrm{k} = \mathrm{Rank(A)}$

注意：对 $u_i = \dfrac{Av_i}{||Av_i||} = \dfrac{1}{\sqrt{\lambda_i}} Av_i$ 左右两边同时左乘上 AA^T 得

- $AA^T u_i = \dfrac{1}{\sqrt{\lambda_i}} AA^T Av_i$
- 即 $AA^T u_i = \dfrac{1}{\sqrt{\lambda_i}} A(A^TA) v_i = \dfrac{1}{\sqrt{\lambda_i}} A\lambda_i v_i$
- 即 $AA^T u_i = \lambda_i \dfrac{1}{\sqrt{\lambda_i}} Av_i$

- 即 $AA^Tu_i = \lambda_i u_i$(根据假设定义 $Av_i = \sqrt{\lambda_i} u_i$)

所以，向量 u_i 是矩阵 AA^T 的特征向量，且特征值相同（对应最开始 AA^T, A^TA 的性质）

将 SVD 用向量写出来：$A = U\Sigma V^{\mathrm{T}} = u_1\sigma_1 v_1^{\mathrm{T}} + \cdots + u_r\sigma_r v_r^{\mathrm{T}}$，这便是 $\boldsymbol{A}$ 的满秩分解，将 $\boldsymbol{A}$ 分解为只有 $rank(\boldsymbol{A})$ 个矩阵之和！当然 $\boldsymbol{U}, \boldsymbol{V}$ 扩展成方阵也是，新增进去的向量不起作用，以及 $\boldsymbol{\Sigma}$ 多出来的部分里面值为 0。

7.2 瑞利熵 (Rayleigh quotient)

实数上的定义：

$$r(\boldsymbol{x}) = \frac{\boldsymbol{x}^{\mathrm{T}} S \boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}} \boldsymbol{x}}$$

可以参考`https://www.cnblogs.com/xingshansi/p/6702188.html?utm_source=itdadao&utm_medium=referral`A-普通瑞利熵部分，得出对该式求得最大值时的条件就是 $Sx = r(x)x$，也就是说当 $r(\boldsymbol{x})$ 为 S 的特征值，$\boldsymbol{x}$ 为 S 的特征向量时满足要求！与书上的表述一致。

7.3 PCA by SVD

PCA 详细推导见`https://www.cnblogs.com/pinard/p/6239403.html`。本小节主要是介绍 SVD 在 PCA 中的应用。

样本中心化：给定一个数据矩阵 $\boldsymbol{A_0}, size = (m \times n); n \to samples, m \to variables$。对 $\boldsymbol{A}_0$ 的每一行减去该行的均值，便得到矩阵 $\boldsymbol{A}$。可以求得，现在矩阵 $\boldsymbol{A}$ 的每行的均值为 0。然后根据**样本协方差矩阵**定义 (sample covariance matrix, 也叫自协方差矩阵) 的定义，得到 $\boldsymbol{S} = \dfrac{\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}}{n-1}$。以上定义以及推导参考`https://blog.csdn.net/Happy_code666/article/details/101716869`。

根据方差以及协方差的定义，可知对于已经样本中心化的矩阵 $\boldsymbol{A}_{2\times 2}$ 有：

- $\left(AA^{\mathrm{T}}\right)_{11}$ and $\left(AA^{\mathrm{T}}\right)_{22}$ 表示该两个样本的方差 s_1^2, s_2^2
- $\left(AA^{\mathrm{T}}\right)_{12}$ 表示该两个样本间的协方差 s_{12}(等同于 s_{21})

Example1 表明通过对矩阵 A 利用其样本协方差矩阵 S 做 SVD 分解，可以找到 A 的每列样本在相应的空间中是如何分布的。

- 样本矩阵 A_0 的总方差 T 等于 $tr(S)$，同时也等于 S 的特征值之和。Total variance $T = \sigma_1^2 + \cdots + \sigma_m^2 = s_1^2 + \cdots + s_m^2 =$ trace (diagonal sum)

- S 的第一个特征向量 $\boldsymbol{u}_1$ 指明了数据分布的主要方向，该方向占了总方差的 σ_1^2/T 部分。
- 当其他部分很小时，对于数据在对应空间的分布就不起明显的作用了。因为对 A 做 SVD 分解得到的起主要作用的奇异向量有 R 个，所以主要有 R 个方向。

注意：在 Chpater4 中提到最小二乘法是利用**竖直距离** (vertical offsets) 来拟合直线，而 PCA 这里是利用**垂直距离** (perpendicular offsets) 来拟合直线，也叫垂直最小二乘 (Perpendicular Least Squares)。

7.4 SVD 的几何理解

(orthogonal) x (diagonal) x (orthogonal) = (rotation) x (stretching) x (rotation)

好几个性质：A 的范数 norm 等。。。详见书本 7.4

7.4.1 伪逆 $\mathbf{A}^+$

对于前面提到的 SVD，有任意矩阵 $\boldsymbol{A}_{m\times n}$ 乘上它的行空间正交基向量 $\boldsymbol{v}_i$ 得到列空间中的 $\sigma_i\boldsymbol{u}_i$。

若是当它的逆矩阵 $\boldsymbol{A}^{-1}$ 存在，则上述定义式子和性质则要反过来，即从 $A\boldsymbol{v}=\sigma\boldsymbol{u}$ 变成 $A^{-1}\boldsymbol{u}=\boldsymbol{v}/\sigma$，且 A^{-1} 的奇异值为 $1/\sigma$，向量 $\boldsymbol{u}'s$ 在 $\boldsymbol{A}^{-1}$ 的行空间中，$\boldsymbol{v}'s$ 在 $\boldsymbol{A}^{-1}$ 的列空间中。

然而**不需要“逆矩阵 $\boldsymbol{A}^{-1}$ 存在”**这个前提，也存在一个矩阵 $\boldsymbol{A}^+_{n\times m}$ 能够满足 $A^+\boldsymbol{u}_i=\boldsymbol{v}_i/\sigma$。

$\boldsymbol{A}_{m\times n}$ 的伪逆矩阵：

$$A^+=V\Sigma^+U^{\mathrm{T}}=\begin{bmatrix}\boldsymbol{v}_1\cdots\boldsymbol{v}_r\cdots\boldsymbol{v}_n\end{bmatrix}_{n\times n}\begin{bmatrix}\sigma_1^{-1}&&&\\&\ddots&&\\&&\sigma_r^{-1}&\\&&&\end{bmatrix}_{n\times m}\begin{bmatrix}\boldsymbol{u}_1\cdots\boldsymbol{u}_r\cdots\boldsymbol{u}_m\end{bmatrix}^{\mathrm{T}}_{m\times m}$$

如果 $\boldsymbol{A}^{-1}$ 存在的话，则 $\boldsymbol{A}^{-1}=\boldsymbol{A}^+$，即 $m=n=r$，相当于将原来的 $U\Sigma V^{\mathrm{T}}$ 转换成 $V\Sigma^{-1}U^{\mathrm{T}}$。只有当矩阵 $\boldsymbol{A}$ 的秩 $r<m$ 或 $r<n$ 时，即它没有逆矩阵时才需要伪逆 $\boldsymbol{A}^+$，且 $\boldsymbol{A}^+$ 有相同的秩 r。

$$A^+\boldsymbol{u}_i=\frac{1}{\sigma_i}\boldsymbol{v}_i\quad for\quad i\le r\quad and\quad A^+\boldsymbol{u}_i=\mathbf{0}\quad for\quad i>r$$

将先前的矩阵乘的结果写出来就是上面的式子。

AA^+ 和 A^+A 的行列空间互换了。

$$\Sigma^+\Sigma = \begin{bmatrix} \boldsymbol{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

AA^+ and A^+A 是投影矩阵，看 Fig 7.6

- AA^+ 把向量投影到 A 的列空间
- A^+A 把向量投影到 A 的行空间

7.4.2 列向量相关的最小二乘

在 chapter4.3 中，矩阵 A 的列向量是线性无关的，但是如果改成线性相关，那么利用等式 $A^{\mathrm{T}}A\boldsymbol{x} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$ 求最小二乘解的时候，$\boldsymbol{x}$ 就会有无穷多个解，无法给出最好的解。

可以利用伪逆来给出最短的解：$\boldsymbol{x}^+ = A^+\boldsymbol{b}$(这里最短指的是向量的模最小)。

8 线性变换

变换 (Transformation) T 对空间 $\boldsymbol{V}$ 的每个向量 $\boldsymbol{v}$ 赋予一个输出 $T(\boldsymbol{v})$。

8.1 线性变换定义

如果变换 T 满足以下条件：

(a) $T(\boldsymbol{v}+\boldsymbol{w})=T(\boldsymbol{v})+T(\boldsymbol{w})$

(b) $T(c\boldsymbol{v})=cT(\boldsymbol{v}) \quad \text{for all } c$

则称该变换是线性的。

若 $\boldsymbol{v}=\mathbf{0}$，则输出 $T(\boldsymbol{v})=\mathbf{0}$。

将 (a) 和 (b) 结合起来，线性变换：$T(c\boldsymbol{v}+d\boldsymbol{w})=cT(\boldsymbol{v})+dT(\boldsymbol{w})$

- 线性变换可以作用在向量空间、矩阵空间以及函数空间上；
- 矩阵乘法是线性变换；
- 两个线性变换的乘积 ST 依然是线性的 $(ST)(\boldsymbol{v})=S(T(\boldsymbol{v}))$；

线性加位移变换 $T(\boldsymbol{v})=A\boldsymbol{v}+\boldsymbol{u}_0$ 称为仿射变换 (affline transformation)，虽然通过这个变换直线还是保持直线，但是该变换不是线性的。

如果输出包含了平方、乘积、长度 (v_1^2 or v_1v_2 or $\|\boldsymbol{v}\|$)，那么变换 T 不是线性的。

旋转 (Rotation) 也是一种变换 T，比如将每个输入向量旋转指定角度，旋转变换是线性的。

从 $\boldsymbol{u}=c_1\boldsymbol{v}_1+c_2\boldsymbol{v}_2+\cdots+c_n\boldsymbol{v}_n \Rightarrow T(\boldsymbol{u})=c_1T(\boldsymbol{v}_1)+c_2T(\boldsymbol{v}_2)+\cdots+c_nT(\boldsymbol{v}_n)$ 知，对于向量空间而言，所有向量都可以表示为基向量的线性组合。因此，理解一个线性变换，只需要了解基向量是怎么变换的：输入空间的一组基向量 $v_1,\ldots,v_n$ 的对应线性变换 T 的输出

$$T(\boldsymbol{v}_1),T(\boldsymbol{v}_2),T(\boldsymbol{v}_3)\ldots T(\boldsymbol{v}_n)$$

知道这些就可以确定任何向量 $\boldsymbol{u}$ 的线性变换 $T(\boldsymbol{u})$。因为向量空间内的向量 u 是基向量 v's 的线性组合。而由线性变换的两条性质易得：**任意的向量的线性变换都可以用基向量的线性变换的结果进行表示**。

(在笛卡尔坐标系下，坐标就是 x,y——这实际上是取坐标轴上的一组单位向量作为基向量的结果。推广来看: $\boldsymbol{u}=c_1\boldsymbol{v}_1+c_2\boldsymbol{v}_2+\cdots+c_n\boldsymbol{v}_n$ 将 u 表示成基向量 v's 的线性组合: 这里对应的系数即为一组 "坐标"。这说明坐标来自一组基。因此，u 的坐标是一组数字，数字的个数表示 u 由多少个基向量组成。)

Example 4：微分可以看成是一种变换 T，并且能求出对应的变换矩阵 A;

Example 5：积分也可以看成一种变换 T，并且对应的变换矩阵是 A 的伪逆 A^+;

给出结论:**所有线性变换都有对应的矩阵表示!** (详见 /home/yzy/GitProject/MIT-Linear-Algebra-Notes/[31] 线性变换及对应矩阵/线性代数 31.pdf 第 2.3 节)

8.2 线性变换对应的矩阵表示

假设有输入向量 $\boldsymbol{v}_{n\times1}$，经过线性变换 T 得到输出向量 $T(\boldsymbol{v})=\boldsymbol{u}_{m\times1}$，那么说明该变换对应的矩阵是 $\boldsymbol{A}_{m\times n}$。

在对应的向量空间 $\boldsymbol{V},\boldsymbol{W}$ 分别选择的**两组基底决定了矩阵 $\boldsymbol{A}$**。

但是空间中有多个基底，所以相同的一个线性变换 T 能够被不同的矩阵所表示，所以线性代数的一个主题就是选出对应线性变换 T 能够给出最好的矩阵表示 (对角矩阵) 的基底。

当输入空间和输出空间有同样的基底组合 (即是**同一个空间**)：

- 如果输入基底和输出基底相同，线性变换 $T(\boldsymbol{v})=\boldsymbol{v}$ 的变换矩阵便是 $\boldsymbol{I}$;
- 如果输入基底和输出基底不同，那么变换矩阵就是对应的**基变换矩阵**! (见 Sec. 基变换:Example 2)

因为 T 是线性变换，所以有：

$$\boldsymbol{v}=\begin{bmatrix}\boldsymbol{v}_1,\cdots,\boldsymbol{v}_n\end{bmatrix}\begin{bmatrix}c_1\\ \vdots\\ c_n\end{bmatrix},\quad T(\boldsymbol{v})=\begin{bmatrix}T(\boldsymbol{v}_1),\cdots,T(\boldsymbol{v}_n)\end{bmatrix}\begin{bmatrix}c_1\\ \vdots\\ c_n\end{bmatrix}$$

其中 $\begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$ 是向量线性组合的系数。

8.2.1 基变换

假设输入空间和输出空间的基底 basis 分别是矩阵 $\boldsymbol{V}$ 和 $\boldsymbol{W}$ 的列向量，则对应线性变换 $T = I$ 的基变换矩阵 $\boldsymbol{B}$ 是

$$\boldsymbol{W}\boldsymbol{B} = \boldsymbol{V} \Rightarrow \boldsymbol{B} = \boldsymbol{W}^{-1}\boldsymbol{V}$$

假设基向量分别是 $\boldsymbol{v}'s, \boldsymbol{w}'s$：

$$\begin{matrix} \boldsymbol{u} = c_1\boldsymbol{v}_1 + \cdots + c_n\boldsymbol{v}_n \\ \boldsymbol{u} = d_1\boldsymbol{w}_1 + \cdots + d_n\boldsymbol{w}_n \end{matrix} \Rightarrow \begin{bmatrix} \boldsymbol{v}_1 \ldots \boldsymbol{v}_n \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} \boldsymbol{w}_1 & \cdots & \boldsymbol{w}_n \end{bmatrix} \begin{bmatrix} d_1 \\ \vdots \\ d_n \end{bmatrix} \Rightarrow \boldsymbol{V}\boldsymbol{c} = \boldsymbol{W}\boldsymbol{d}$$

原向量 $\boldsymbol{u}$ 在输出空间的基向量中对应的系数 $\boldsymbol{d}$：$\boldsymbol{d} = \boldsymbol{W}^{-1}\boldsymbol{V}\boldsymbol{c}$

所以若输入空间是常见的标准空间，即 $\boldsymbol{V} = \boldsymbol{I}$，当将其转变为一个不同的输出空间的基向量 $\boldsymbol{W}$ 时，用于基变换的矩阵不是 $\boldsymbol{W}$ 而是 $\boldsymbol{B} = \boldsymbol{W}^{-1}$

8.2.2 构建对应线性变换的矩阵

现假设一个线性变换 T 将空间 $\boldsymbol{V}_{dim=n}$ 转换为空间 $\boldsymbol{W}_{dim=m}$，分别从 $\boldsymbol{V}, \boldsymbol{V}$ 中选择一组基底 $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$, $\{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_m\}$，则对应该线性变换 T 的矩阵为:$\boldsymbol{A}_{m\times n}$。

构建规则：

将线性变换 T 应用到 $\boldsymbol{v}_j$ 上，则 $T(\boldsymbol{v}_j) = a_{1j}\boldsymbol{w}_1 + \cdots + a_{mj}\boldsymbol{w}_m$，即输出向量是 $\boldsymbol{W}$ 的基底的线性组合。此时，系数 $a_{1j}, \cdots, a_{mj}$ 便是矩阵 $\boldsymbol{A}$ 的第 j 列！

相当于：

$$\boldsymbol{u} = \begin{bmatrix} \boldsymbol{v}_1, \cdots, \boldsymbol{v}_n \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} \xRightarrow{T} T(\boldsymbol{u}) = \begin{bmatrix} T(\boldsymbol{v}_1), \cdots, T(\boldsymbol{v}_n) \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$$

$$= \begin{bmatrix} \boldsymbol{w}_1, \cdots, \boldsymbol{w}_m \end{bmatrix} \begin{bmatrix} \boldsymbol{a}_1, \cdots, \boldsymbol{a}_n \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} = \boldsymbol{W}_{m\times m}\boldsymbol{A}_{m\times n}\boldsymbol{c}$$

即 $T(\boldsymbol{V}) = \boldsymbol{W}\boldsymbol{A} \Rightarrow \boldsymbol{A} = \boldsymbol{W}^{-1}T(\boldsymbol{V})$，且 $\boldsymbol{W}\underbrace{\boldsymbol{A}\boldsymbol{c}} = \boldsymbol{W}\underbrace{\boldsymbol{d}}$，说明若是一个向量 $\boldsymbol{u}$ 找到了由基底 $\boldsymbol{v}'s$ 的线性组合所表示的系数 $\boldsymbol{c} = (c_1, \cdots, c_n)$，那么 $\boldsymbol{A}\boldsymbol{c}$ 便是向量 $\boldsymbol{u}$ 由进行 T 线性变换后 $T(\boldsymbol{u})$ 的基底 $\boldsymbol{w}'s$ 的线性组合所表示的系数！

（**注意**：因为从 $\boldsymbol{v}'s \Rightarrow T(\boldsymbol{v}'s)$，$T(\boldsymbol{v}'s)$ 不一定就是输出空间 $\boldsymbol{W}$ 的基底 $\boldsymbol{w}'s$。所以在得到 $T(\boldsymbol{v}'s)$ 后还要用输出空间的基底对其每个列向量进行线性表示，找到其中的系数作为 $\boldsymbol{A}$。）

如果输入空间和输出空间的基底分别是标准基底 (standard basis)：columns of $I_{n\times n}$ and $I_{m\times m}$，那么线性变换 $T(\boldsymbol{x}) = A\boldsymbol{x}$ 对应的矩阵表示为 A。线性变换使用标准基底得到的对应矩阵叫做标准矩阵 (standard matrix)。

线性变换 T, S 对应的矩阵为 A, B，则线性变换 TS 对应的矩阵为 AB！

8.2.3 选择最好的基底

当选择不同的基底时，线性变换 T 可以被不同的矩阵所表示。

基底一般有两种选择：1. 特征向量；2. 奇异向量。

Example 7：

- 输入空间与输出空间的基底都是 $(1,0),(0,1)$，所以变换矩阵才是 A。
- 将输入与输出空间的基底一起改为特征向量 $(1,-1),(1,1)$
- 则变换矩阵 A 就变成了由特征值组成的对角矩阵 Λ

假设以线性变换 T 是从 $\boldsymbol{R}^n \to \boldsymbol{R}^n$ 的，即输入空间和输出空间相同，则令输入向量的基底与输出向量的基底相同，且有如下选择：

- 一组基底 $\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n\}(\boldsymbol{V})$ 对应的变换矩阵为 A
- 另一组基底 $\{\boldsymbol{w}_1, \cdots, \boldsymbol{w}_n\}(\boldsymbol{W})$ 对应的变换矩阵为 A_{new}

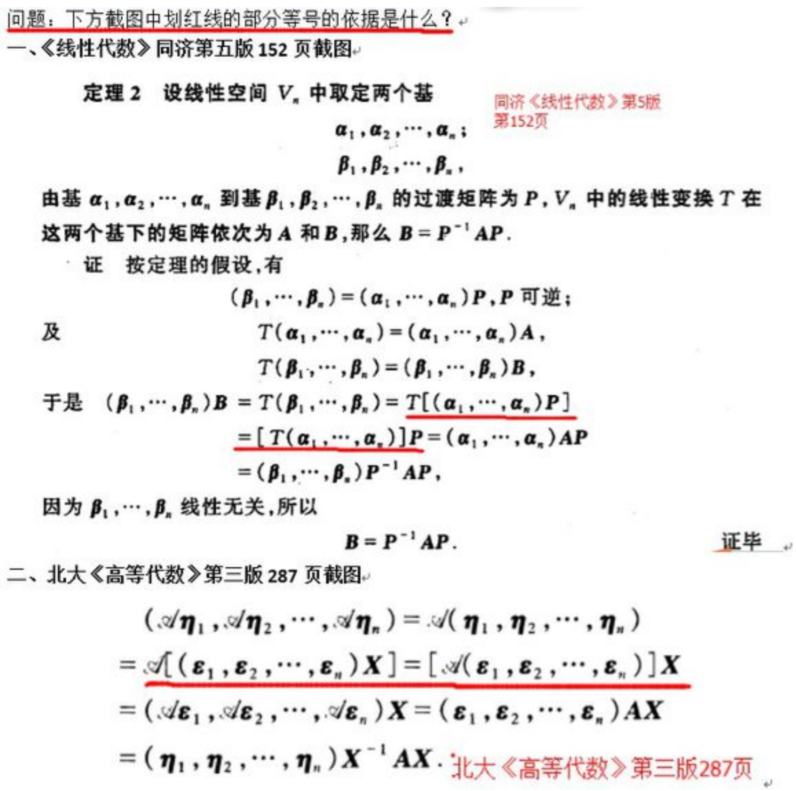
线性变换方式没有变化，只是在不同基下对应的坐标不同，线性变换矩阵不同。这里只是为了从线性变换矩阵的角度认识：不同基下，同一个线性变换对应的线性变换矩阵会有什么关系？

从基变换一节中可知：一个向量可以写成多种基底的线性组合 (同一空间内) 且不同基底之间可以用过度矩阵来表示。~~利用这个条件来构造等式~~，找出上述关系。

假设 $\boldsymbol{R}^n$ 中有一向量 $\boldsymbol{u}$，且根据以上假设以及 $Sec\quad 8.2.2$ 有：

- $\boldsymbol{u} = \boldsymbol{V}\boldsymbol{c}, \quad T(\boldsymbol{u}) = \boldsymbol{V}\boldsymbol{A}\boldsymbol{c}$，其中 $\boldsymbol{c}$ 是向量 $\boldsymbol{u}$ 对应基底 $\boldsymbol{V}$ 下的线性组合系数也即坐标；
- $T(\boldsymbol{V}) = \boldsymbol{V}\boldsymbol{A}$；
- $\boldsymbol{u} = \boldsymbol{W}\boldsymbol{d}, \quad T(\boldsymbol{u}) = \boldsymbol{W}\boldsymbol{A}_{new}\boldsymbol{d}$，其中 $\boldsymbol{d}$ 是向量 $\boldsymbol{u}$ 对应基底 $\boldsymbol{W}$ 下的线性组合系数也即坐标；
- $T(\boldsymbol{W}) = \boldsymbol{W}\boldsymbol{A}_{new}$；
- 两个基底之间的过度矩阵为 $\boldsymbol{P}$(方阵且可逆)，且 $\boldsymbol{V}\boldsymbol{P} = \boldsymbol{W}$；

则根据上述条件构造等式：$\cancel{\boldsymbol{V}\boldsymbol{A}\boldsymbol{c} = \boldsymbol{W}\boldsymbol{A}_{new}\boldsymbol{d}}$，这种构造带有向量 $\boldsymbol{c}, \boldsymbol{d}$ 无法消去。正确的做法应该从基底的线性变换 T 出发，参考：



问题：下方截图中划红线的部分等号的依据是什么？

一、《线性代数》同济第五版 152 页截图

同济《线性代数》第5版第152页

定理 2 设线性空间 V_n 中取定两个基

$$\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n;$$
$$\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n,$$

由基 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n$ 到基 $\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n$ 的过渡矩阵为 P，V_n 中的线性变换 T 在这两个基下的矩阵依次为 A 和 B，那么 $B=P^{-1}AP$.

证 按定理的假设，有

$$(\boldsymbol{\beta}_1,\cdots,\boldsymbol{\beta}_n)=(\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_n)P,\ P\text{ 可逆};$$

及

$$T(\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_n)=(\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_n)A,$$
$$T(\boldsymbol{\beta}_1,\cdots,\boldsymbol{\beta}_n)=(\boldsymbol{\beta}_1,\cdots,\boldsymbol{\beta}_n)B,$$

于是

$$\begin{aligned}(\boldsymbol{\beta}_1,\cdots,\boldsymbol{\beta}_n)B &= T(\boldsymbol{\beta}_1,\cdots,\boldsymbol{\beta}_n)=T[(\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_n)P]\\ &=[T(\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_n)]P=(\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_n)AP\\ &=(\boldsymbol{\beta}_1,\cdots,\boldsymbol{\beta}_n)P^{-1}AP,\end{aligned}$$

因为 $\boldsymbol{\beta}_1,\cdots,\boldsymbol{\beta}_n$ 线性无关，所以

$$B=P^{-1}AP.$$

证毕

二、北大《高等代数》第三版 287 页截图

$$\begin{aligned}&(\mathscr{A}\boldsymbol{\eta}_1,\mathscr{A}\boldsymbol{\eta}_2,\cdots,\mathscr{A}\boldsymbol{\eta}_n)=\mathscr{A}(\boldsymbol{\eta}_1,\boldsymbol{\eta}_2,\cdots,\boldsymbol{\eta}_n)\\ &=\mathscr{A}[(\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\cdots,\boldsymbol{\varepsilon}_n)\boldsymbol{X}]=[\mathscr{A}(\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\cdots,\boldsymbol{\varepsilon}_n)]\boldsymbol{X}\\ &=(\mathscr{A}\boldsymbol{\varepsilon}_1,\mathscr{A}\boldsymbol{\varepsilon}_2,\cdots,\mathscr{A}\boldsymbol{\varepsilon}_n)\boldsymbol{X}=(\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\cdots,\boldsymbol{\varepsilon}_n)\boldsymbol{AX}\\ &=(\boldsymbol{\eta}_1,\boldsymbol{\eta}_2,\cdots,\boldsymbol{\eta}_n)\boldsymbol{X}^{-1}\boldsymbol{AX}.\end{aligned}$$

北大《高等代数》第三版287页

以及https://www.bilibili.com/read/cv4496980/。

所以

$$T(\boldsymbol{W})=T(\boldsymbol{VP})=T(\boldsymbol{V}[\boldsymbol{p_1},\cdots,\boldsymbol{p_n}])=T([\boldsymbol{Vp_1},\cdots,\boldsymbol{Vp_n}])$$

利用线性变换的性质——只作用于向量，不作用于线性组合的系数！这里 $\boldsymbol{P}$ 的每个列向量都是 $\boldsymbol{V}$ 的列向量的组合系数，所以可以提出来！即

$$=T(\boldsymbol{V})\boldsymbol{P}=\boldsymbol{VAP}\Rightarrow\boldsymbol{WA}_{new}=\boldsymbol{VAP}=\boldsymbol{WP}^{-1}\boldsymbol{AP}\Rightarrow\underbrace{\boldsymbol{A}_{new}=\boldsymbol{P}^{-1}\boldsymbol{AP}}$$

结论:如果是同一种线性变换,在不同基下对应的线性变换矩阵为 A,A_{new}，则 A 与 A_{new} **相似**。

注意:书上在此处的描述实际上对应 $\boldsymbol{A}_{new}=\boldsymbol{P}^{-1}\boldsymbol{AP}=(\boldsymbol{V}^{-1}\boldsymbol{W})^{-1}\boldsymbol{A}(\boldsymbol{V}^{-1}\boldsymbol{W})$，因为书上指定输入基底 $\boldsymbol{V}=\boldsymbol{I}$，所以 $\boldsymbol{A}_{new}=\boldsymbol{W}^{-1}\boldsymbol{AW}\Rightarrow\boldsymbol{A}_{new}=\boldsymbol{B}^{-1}\boldsymbol{AB}(\boldsymbol{B}$ 对应 $\boldsymbol{W})$。

当输入空间 $\boldsymbol{V}$ 和输出空间 $\boldsymbol{W}$ 的维度不同时，对应线性变换 T 的矩阵

A 可能不是对称阵，也不是方阵。但是可以通过选择基底来产生对应线性变换的对角阵 (长方形对角阵)！

具体的做法：通过选择奇异向量作为基底，即利用 SVD——$U^{-1}AV = \Sigma_{m\times n}$，$\{\boldsymbol{v}_1,\cdots,\boldsymbol{v}_n\}$ 作为输入基底；$\{\boldsymbol{u}_1,\cdots,\boldsymbol{u}_m\}$ 作为输出基底。

在原基底下的变换矩阵与选择奇异向量作为基底的变换矩阵相比：$B_{\text{out}}^{-1}AB_{\text{in}} = U^{-1}AV = \Sigma$

总结：最佳的输入-输出基底是变换矩阵 A 的特征向量或奇异向量，$B^{-1}AB = \Lambda = \text{eigenvalues} \quad B_{\text{out}}^{-1}AB_{\text{in}} = \Sigma = \text{singular values}$