

18.06 线性代数 Linear Algebra

第三单元 正定矩阵以及应用

Unit 3 Positive Definite Matrices & Applications

第 25 讲 对称矩阵和正定性

第 26 讲 复矩阵；快速傅里叶变换

第 27 讲 正定矩阵和最小值

第 28 讲 相似矩阵和若尔当标准型

第 29 讲 奇异值分解

第 30 讲 线性变换及对应矩阵

第 31 讲 基变换和图像压缩

第 32 讲 复习三

第 33 讲 左右逆和伪逆矩阵

第 34 讲 总复习

第 25 讲 对称矩阵和正定性

Symmetric matrices and positive definiteness

对称矩阵是最重要的矩阵之一，其特征值为实数并且拥有一套正交特征向量。正定矩阵的性质则更好。

对称矩阵 Symmetric matrices

包含特殊性质的矩阵，例如 Markov 矩阵，其特征值和特征向量往往拥有一定特性。对称矩阵 $A=A^T$ 的特征值为实数，具有完全正交的特征向量。这里的“有”，是指可以选出一套完全正交的特征向量（例如在重特征值条件下，可能存在一个平面内向量都可以作为特征向量）。

如果 A 具有 n 个线性无关的特征向量，可以对角化得到 $A=SMS^{-1}$ 。而对于对称矩阵， $A=Q\Lambda Q^{-1}=Q\Lambda Q^T$ ，其中 Q 为正交矩阵，列向量为标准正交基，这个公式本身还显示了矩阵的对称性。矩阵能够进行这种分解，在数学上称为“谱定理”（spectral theorem），将特征值视为“谱”，在物理上称之为“主轴定理”。

实特征值 Real eigenvalues

矩阵 A 具有特征值 λ 和特征向量 x ，则有 $Ax=\lambda x$ 。若为 λ 复数，则其共轭复数满足 $A\bar{x}=\bar{\lambda}\bar{x}$ 。两侧取转置，则有：

$$\bar{x}^T A^T = \bar{x}^T \bar{\lambda} ,$$

两侧同乘 x ，则：

$$\bar{x}^T A^T x = \bar{x}^T \bar{\lambda} x ,$$

因为矩阵为对称矩阵，所以有：

$$\bar{x}^T \lambda x = \bar{x}^T \bar{\lambda} x .$$

因此若有 $\bar{x}^T x \neq 0$ 则 $\lambda = \bar{\lambda}$ ，即 λ 为实数。

因为 \bar{x} 为 x 的共轭，因此两者点积 $\bar{x}^T x$ 得到的是 x 中的每一分量与其共轭复数的乘积之和，即复向量 x 的模，当且仅当 $x=0$ 时，其模为 0。

$$\bar{x}^T x = [\bar{x}_1 \quad \bar{x}_2 \quad \bar{x}_3 \quad \bar{x}_4] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = |x_1|^2 + |x_2|^2 + |x_3|^2 + |x_4|^2$$

因此得证 λ 为实数。

对于复矩阵 A ，若有 $A^T = A$ 则上面的推导全部成立，因此共轭对称复矩阵 A 的特征值为实数。

对于对称矩阵， $A=Q\Lambda Q^{-1}=Q\Lambda Q^T$ ，可以写作：

$$A = Q\Lambda Q^T = [\mathbf{q}_1 \quad \mathbf{q}_2 \quad \cdots \quad \mathbf{q}_n] \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \begin{bmatrix} \mathbf{q}_1^T \\ \mathbf{q}_2^T \\ \vdots \\ \mathbf{q}_n^T \end{bmatrix}$$

$$= \lambda_1 \mathbf{q}_1 \mathbf{q}_1^T + \lambda_2 \mathbf{q}_2 \mathbf{q}_2^T + \cdots + \lambda_n \mathbf{q}_n \mathbf{q}_n^T$$

矩阵是 $\mathbf{q}_k \mathbf{q}_k^T$ ($= \mathbf{q}_k \mathbf{q}_k^T / (\mathbf{q}_k^T \mathbf{q}_k)$) 是朝向向量 \mathbf{q}_k 的投影矩阵, 所以每一个对称矩阵都是正交投影矩阵的线性组合。这是理解谱定理的另一种方法。

当确认矩阵特征值为实数后, 下一个要考虑的问题就是它是正还是负数, 因为这影响着微分方程中体系的稳定与否。但是对于大型矩阵, 通过计算 $|\mathbf{A} - \lambda \mathbf{I}| = 0$ 得到特征值进行判定难以实现, 即使用 MATLAB 求解, 结果也不一定可靠, 但 MATLAB 可以得到矩阵的主元, 而对称阵的主元中正负数的个数与特征值相同, 即正主元的数目等于正特征值的数目。

对于对称阵主元与特征值符号相匹配这件事情, 通常是用合同矩阵的惯性定理加以证明。若存在可逆矩阵 \mathbf{C} 满足 $\mathbf{A} = \mathbf{C} \mathbf{B} \mathbf{C}^T$, 就称 \mathbf{A} 与 \mathbf{B} 为合同矩阵, 惯性定理是指矩阵的特征值符号经过合同变换后不发生变化。对称阵 \mathbf{A} 经过消元操作可得 $\mathbf{A} = \mathbf{L} \mathbf{D} \mathbf{L}^T$, 而经过对角化处理得到 $\mathbf{A} = \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^T$, 两者比较可知 $\mathbf{D} = (\mathbf{L}^{-1} \mathbf{Q}) \mathbf{\Lambda} (\mathbf{L}^{-1} \mathbf{Q})^T$, 因此两个对角阵为合同矩阵, 主元和特征值的符号匹配。

GS 在书里给了一个接近同伦理论的证明:

对于矩阵 $\begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & \\ & -8 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix}$, 在下三角阵转变为 $\begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, 的过程中, 矩阵的变化为 $\begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 \\ 0 & -8 \end{bmatrix}$, 矩阵的特征值也从 4, -2 变为 1, -8

(即原矩阵的主元)。这个过程中如果发生符号变化, 则总有一个瞬间特征值穿过 0 点, 那个瞬间矩阵应该是奇异阵, 但是相乘的三个矩阵一直是非奇异阵, 因此不可能有这样一个时刻存在, 所以特征值不会发生符号变化, 即和主元符号保持一致。

矩阵 $\mathbf{A} + b\mathbf{I}$ 的特征值比矩阵 \mathbf{A} 的特征值大 b , 可以通过 $\mathbf{A} + b\mathbf{I}$ 的主元来了解矩阵 \mathbf{A} 的特征值与 b 的大小关系, 因此利用这个性质可以估计特征值的状态。

正定矩阵 Positive definite matrices

正定矩阵 \mathbf{A} 是特征值都为正数的对称矩阵。它的主元也均为正数。

例如矩阵 $\mathbf{A} = \begin{bmatrix} 5 & 2 \\ 2 & 3 \end{bmatrix}$ 。主元为 5 和 $(\det \mathbf{A})/5 = 11/5$ 。主元都为正数, 且本身为对称矩阵, 因此 \mathbf{A} 是正定矩阵。计算其特征值为: $|\mathbf{A} - \lambda \mathbf{I}| = \lambda^2 - 8\lambda + 11 = 0$, $\lambda = 4 \pm \sqrt{5}$ 。

因为主元和特征值都是正的, 可以知道正定矩阵的行列式也是正的。但是反之, 并不成立。矩阵 $\begin{bmatrix} -1 & 0 \\ 0 & -3 \end{bmatrix}$ 的行列式也是正的, 但不是正定矩阵。

若将行列式作为正定的判据，则要求 n 阶矩阵左上角的所有 $k \times k$ ($1 \leq k \leq n$) 子行列式 (subdeterminant) 数值均为正，矩阵才能确定为正定矩阵。

本讲的内容将之前教授的主元、行列式和特征值的概念结合在了一起，对于正定矩阵这些都是正的，当完全掌握了它们的性质后会推广到非对称矩阵，甚至非方阵。

第 26 讲 复矩阵；快速傅里叶变换

Complex matrices; fast Fourier transform

实矩阵也可能有复特征值，因此无法避免在矩阵运算中碰到复数，本讲学习处理复数矩阵和复向量。

最重要的复矩阵是傅里叶矩阵，它用于傅里叶变换。而对于大数据处理快速傅里叶变换（FFT）显得更为重要，它将傅立叶变换的矩阵乘法中运算的次数从 n^2 次降至 $n \log_2 n$ 次。

复向量 Complex vectors

对于给定的复向量 $\mathbf{z} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} \in \mathbb{C}^n$ ，其元素中有复数，因此 $\mathbf{z}^T \mathbf{z}$ 无法给出向量的长

度。例如： $\begin{bmatrix} 1 & i \end{bmatrix} \begin{bmatrix} 1 \\ i \end{bmatrix} = 0$ 。则定义 $|\mathbf{z}|^2 = \bar{\mathbf{z}}^T \mathbf{z} = |z_1|^2 + |z_2|^2 + \cdots + |z_n|^2$ 为向量长度。因此

向量 $\begin{bmatrix} 1 \\ i \end{bmatrix}$ 的长度就是 $\begin{bmatrix} 1 & -i \end{bmatrix} \begin{bmatrix} 1 \\ i \end{bmatrix} = 2$ 。记 $|\mathbf{z}|^2 = \bar{\mathbf{z}}^T \mathbf{z} = \mathbf{z}^H \mathbf{z}$ 。H 来自于 “Hermite”。

与之相似，内积的定义也变为 $\mathbf{y}^H \mathbf{x} = \bar{\mathbf{y}}^T \mathbf{x} = \bar{y}_1 x_1 + \bar{y}_2 x_2 + \cdots + \bar{y}_n x_n$ 。

复矩阵 Complex matrices

上一讲中讲到了对于复矩阵 \mathbf{A} ，若有 $\bar{\mathbf{A}}^T = \mathbf{A}$ 则复矩阵 \mathbf{A} 的特征值为实数。这种复矩阵被称为埃尔米特矩阵（Hermitian matrixes，又译作“厄米特矩阵”或者“厄米矩阵”）。转置共轭记作 $\mathbf{A}^H = \bar{\mathbf{A}}^T$ 。

例如矩阵 $\begin{bmatrix} 2 & 3+i \\ 3-i & 5 \end{bmatrix}$ 为埃尔米特矩阵。它具有实数特征值和正交的特征向量。

由性质可知埃尔米特矩阵对角线均为实数。

此处向量标准正交的意思是 $\bar{\mathbf{q}}_j^T \mathbf{q}_k = \begin{cases} 0 & j \neq k \\ 1 & j = k \end{cases}$ 。用 n 个标准正交的复向量作为

列向量可以构造一个矩阵 \mathbf{Q} ，则有 $\mathbf{Q}^T \mathbf{Q} = \mathbf{I} = \mathbf{Q}^H \mathbf{Q}$ 。这个复空间的正交矩阵称为酉矩阵（unitary matrix）。

傅里叶变换 Fourier transform

傅里叶级数是将周期函数或者信号变换为不同频率的三角函数的和函数。

$$f(x) = a_0 + a_1 \cos x + b_1 \sin x + a_2 \cos 2x + b_2 \sin 2x + \cdots$$

在电子工程或者计算机科学中 $n \times n$ 矩阵的行和列从第 0 行和第 0 列开始计数，最后到第 $n-1$ 行和第 $n-1$ 列。我们在讨论傅里叶矩阵的时候遵从这种习惯。

$$\mathbf{F}_n = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega & \omega^2 & & \omega^{(n-1)} \\ 1 & \omega^2 & \omega^4 & & \omega^{2(n-1)} \\ \vdots & & & \ddots & \\ 1 & \omega^{n-1} & \omega^{2(n-1)} & & \omega^{(n-1)^2} \end{bmatrix}$$

$(\mathbf{F}_n)_{jk} = \omega^{jk}$ ，傅里叶矩阵为对称矩阵 $\mathbf{F}_n = \mathbf{F}_n^T$ 。矩阵中的 $\omega^n = 1, \omega = \exp(i2\pi/n)$ 。矩阵的列向量正交。 ω 的方次分布在复平面的单位圆上，只是幅角不同。当 $n=4$ 时， $\omega^4 = 1, \omega = \exp(i2\pi/4) = i$ 。

$$\mathbf{F}_4 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & i^2 & i^3 \\ 1 & i^2 & i^4 & i^6 \\ 1 & i^3 & i^6 & i^9 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{bmatrix}$$

从矩阵可以得到一个四点（离散的）傅里叶变换，它的逆矩阵就是反傅里叶变换。因为傅里叶矩阵列向量正交，所以其逆矩阵很容易计算。实际上这个矩阵可以分解成一系列稀疏矩阵，并且它们的逆矩阵都很容易得到。

计算可知列向量的模不是 1，矩阵除以 2 之后，向量标准正交： $\frac{1}{4} \mathbf{F}_4^H \mathbf{F}_4 = \mathbf{I}$ 。它的逆矩阵就是共轭转置。

快速傅里叶变换 Fast Fourier transform

64 阶傅里叶矩阵 \mathbf{F}_{64} 中的 ω_{64} 与 32 阶傅里叶矩阵 \mathbf{F}_{32} 的元素 ω_{32} 相比，幅角是其一半， $(\omega_{64})^2 = \omega_{32}$ 。可以从分块矩阵运算找到两者的联系：

$$[\mathbf{F}_{64}] = \begin{bmatrix} \mathbf{I} & \mathbf{D} \\ \mathbf{I} & -\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{F}_{32} & \mathbf{0} \\ \mathbf{0} & \mathbf{F}_{32} \end{bmatrix} [\mathbf{P}]$$

其中 \mathbf{P} 是置换矩阵，而 \mathbf{D} 为对角矩阵：

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & \cdots & 0 & 0 \\ 0 & 0 & 1 & 0 & & & 0 & 0 \\ & & \vdots & & & & & \\ 0 & 0 & 0 & 0 & \cdots & \cdots & 1 & 0 \\ 0 & 1 & 0 & 0 & \cdots & \cdots & 0 & 0 \\ 0 & 0 & 0 & 1 & \cdots & \cdots & 0 & 0 \\ & & \vdots & & & & & \\ 0 & 0 & 0 & 0 & \cdots & \cdots & 0 & 1 \end{bmatrix}, \mathbf{D} = \begin{bmatrix} 1 & & & & & & & \\ & \omega & & & & & & \\ & & \omega^2 & & & & & \\ & & & \ddots & & & & \\ & & & & \omega^{31} & & & \end{bmatrix}$$

\mathbf{P} 的效果是使得所乘的向量 \mathbf{x} 中序数为奇数的分量如 x_1, x_3, x_5 等提到前面，而偶数分量 x_2, x_4 等放到后面。

计算 64 阶傅里叶变换（傅里叶矩阵乘以向量）的计算量是 64^2 ，而等式右侧的计算量是 2×32^2 （两个 32 阶的傅里叶矩阵）再加上一些修正项，修正项主要来自

于与对角矩阵 \mathbf{D} 的乘法，大约为 32 次。继续对 \mathbf{F}_{32} 进行分解，计算的运算量再一次下降变为 $2 \times (2 \times 16^2 + 16) + 32$ 。连续进行拆分，傅里叶矩阵的尺寸变化依次为 64、32、16、8、4、2、1，经过 $\log_2 64$ 次分解，最后仅剩修正项的运算， $\log_2 64 \times 32$ 次。对于 n 阶矩阵，可将 n^2 次计算降至 $(n/2) \log_2 n$ 。例如对于 1024 阶矩阵，运算量从 1024×1024 降至 5×1024 。

第 27 讲 正定矩阵和最小值

Positive definite matrices and minima

本讲学习正定矩阵，这部分内容将本课程之前的知识点：主元、行列式、特征值以及方程的稳定性融为一体。本讲介绍如何判定一个矩阵是否正定矩阵，以及当一个矩阵是正定矩阵时，其内涵和矩阵操作的效果有何特别之处。此外还有正定矩阵与几何的关系：椭圆和正定有关，双曲线与正定无关。

正定矩阵 Positive definite matrices

给定一个 2×2 矩阵，有四个途径判定矩阵是否正定矩阵：

- 1) 特征值： $\lambda_1 > 0, \lambda_2 > 0$
- 2) 行列式（所有子行列式）： $a > 0, ac - b^2 > 0$
- 3) 主元： $a > 0, (ac - b^2)/a > 0$
- 4) 表达式 $\mathbf{x}^T \mathbf{A} \mathbf{x} > 0$ ($\mathbf{x} = 0$ 除外)。通常这就是正定的定义，而前三条是用来验证正定性的条件。

给定矩阵 $\mathbf{A} = \begin{bmatrix} 2 & 6 \\ 6 & y \end{bmatrix}$ ，从判据可知矩阵为正定阵的条件是 $2y - 36 > 0$ ，即 $y > 18$ 。

矩阵 $\begin{bmatrix} 2 & 6 \\ 6 & 18 \end{bmatrix}$ 正好处在判定为正定矩阵的临界点上，称之为半正定 (positive semidefinite) 矩阵，它具有一个特征值 0，是奇异矩阵，只有一个主元，行列式为 0。半正定矩阵特征值大于等于 0。

再观察 $\mathbf{x}^T \mathbf{A} \mathbf{x}$ 判据：

$$\begin{aligned} \mathbf{x}^T \mathbf{A} \mathbf{x} &= \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 2 & 6 \\ 6 & 18 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= 2x_1^2 + 12x_1x_2 + 18x_2^2 \end{aligned}$$

之前讨论得都是线性方程 $\mathbf{A} \mathbf{x}$ ，现在引入 \mathbf{x}^T ，变成二次，如果对于任意 x, y ，这种二次型 (quadratic form) $ax^2 + 2bxy + cy^2$ 均大于零，则矩阵为正定矩阵。

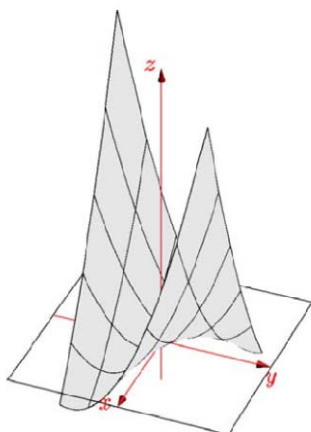
在本例的半正定矩阵中，当 $x_1 = 3, x_2 = -1$ 时 $2x_1^2 + 12x_1x_2 + 18x_2^2 = 2(x_1 + 3x_2)^2 = 0$ 。

如果将矩阵变为 $\begin{bmatrix} 2 & 6 \\ 6 & 7 \end{bmatrix}$ ，二次型为 $f(x, y) = 2x^2 + 12xy + 7y^2$ ，从几何图像上看没有最小值点，在 origin 处有一鞍点。鞍点在某个方向上看是极大值点，在另一方向上是极小值点，实际上最佳观测角度是特征向量的方向。

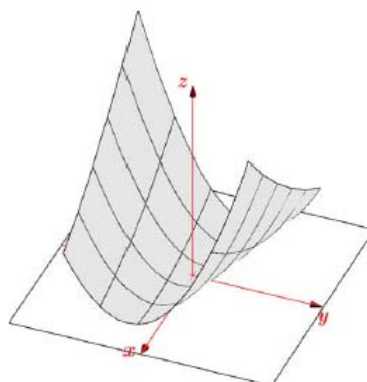
如果将矩阵变为 $\begin{bmatrix} 2 & 6 \\ 6 & 20 \end{bmatrix}$ ，主元为正；特征值之积为行列式的值 4，和为矩阵的迹 22，因此特征值为正；子行列式均为正。矩阵为正定矩阵。

二次型 $f(x, y) = 2x^2 + 12xy + 20y^2$ ，其图像最小值点为原点，一阶偏导数为 0，

二阶偏导数为正。

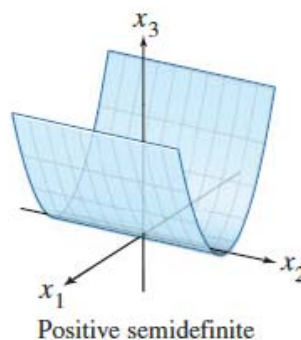
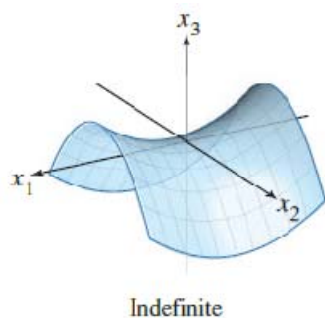
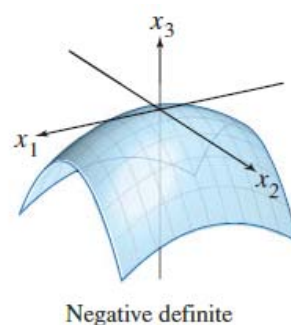
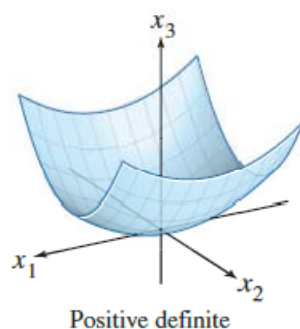


$f(x, y) = 2x^2 + 12xy + 7y^2$ 的图像



$f(x, y) = 2x^2 + 12xy + 20y^2$ 的图像

当年上课时候老师讲双曲面的鞍点,画的图不好,有个篮球队的哥们理解不了,于是下了课跑去问老师,老师指着他的热身篮球裤说,就你这种裤腿两侧能撕开的篮球裤,全部打开,拎起来,裤裆地方就是鞍点.....算了,反正 GS 画图真心不行,看看 Lay 的吧:



微积分中判定最小值点的判据:一阶导数等于零 $\frac{du}{dx} = 0$, 二阶导数为正 $\frac{d^2u}{dx^2} > 0$ 。

线性代数中判据为二阶导数矩阵正定。

对于二次型我们可以用配方的办法来验证其是否具有最小值:

$$f(x, y) = 2x^2 + 12xy + 20y^2 = 2(x + 3y)^2 + 2y^2$$

配方使得 x^2 的系数和交叉项 xy 的系数配合形成完全平方的形式,这个时候用到的 y^2 的系数正好是 18,即判定正定的临界点。如果实际的系数 d 大于 18,则还

剩余 $(d-18)y^2$ 二次型在原点之外一定大于零 若小于 18 则二次型可以小于等于 0。

对于 $f(x, y) = 2x^2 + 12xy + 20y^2 = 2(x + 3y)^2 + 2y^2$ ，其几何图像为碗型的曲面，如果我们用 $f=1$ 的截面横截曲面，得到的就是 $2(x + 3y)^2 + 2y^2 = 1$ 的椭圆曲线。而对于双曲面进行切割就得到双曲线。

配方法其实就是消元：

$$\begin{bmatrix} 2x^2 & 6xy \\ 6xy & 20y^2 \end{bmatrix}, \begin{bmatrix} 2 & 6 \\ 6 & 20 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 2 & 6 \\ 0 & 2 \end{bmatrix}. \quad f(x, y) = 2(x + 3y)^2 + 2y^2$$

主元就是平方项系数， L 矩阵中的行操作数 l_{21} 就是配方项内 y 的系数。因此这就是为什么主元为正则矩阵为正定矩阵，因为主元是每一个完全平方项的系数。本例中二次型表达式的配方说明了二维的情形，而线代的理论可以将之推广到 n 维。

二阶导数的矩阵记为 $\begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix}$ ，矩阵对称代表交叉二阶偏导数与求导顺序无关 $f_{xy} = f_{yx}$ 。在微积分中我们学到的判据 $f_{xx}f_{yy} > f_{xy}^2$ ，和二阶矩阵判定正定是等价的，并且线代可以推广到 n 维。

3 阶矩阵 $A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$ ，它是正定矩阵。计算子行列式得到 $|2| = 2$ ， $\begin{vmatrix} 2 & -1 \\ -1 & 2 \end{vmatrix} = 3$ ， $\begin{vmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{vmatrix} = 4$ 。主元是 2，3/2，4/3。特征值是 $2 - \sqrt{2}$ ，2， $2 + \sqrt{2}$ 。

这是 GS 最爱的矩阵之一，可以用来把二阶微分方程变成离散问题，因为它每一行都是差分方程 $f_{n+1} - 2f_n + f_{n-1}$ 。

其二次型为 $\mathbf{x}^T A \mathbf{x} = 2x_1^2 + 2x_2^2 + 2x_3^2 - 2x_1x_2 - 2x_2x_3$ 。

这是一个四维的图像，三个维度 x_1, x_2, x_3 ，还有函数 f ，如果用 $f=1$ 切割图像，则得到 $2x_1^2 + 2x_2^2 + 2x_3^2 - 2x_1x_2 - 2x_2x_3 = 1$ 。这是一个椭球体，三个特征值不同，因此椭球的三个长轴长度不同。三个轴的方向就是特征向量的方向，轴长度就是特征值，矩阵的分解 $A = Q\Lambda Q^T$ 很好的说明了这件事，这就是所谓的“主轴定理”。

对于三条判据可以判定正定： $\mathbf{x}^T A \mathbf{x} > 0$ ($\mathbf{x} = 0$ 除外)。已经分析了主元要大于零的原因，因为它是配方后的参数，只有都大于零才能保证正定。以下对于判据 1 和 2 做简要说明。

对称矩阵 A ，其正交的特征向量可以张成整个空间，因此任意向量 \mathbf{x} 均可表示成特征向量的线性组合 $\mathbf{x} = c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \dots + c_n\mathbf{x}_n$ ，

代入得 $\mathbf{x}^T A \mathbf{x} = c_1^2\lambda_1 + c_2^2\lambda_2 + \dots + c_n^2\lambda_n$ ，当特征值都大于零且 $\mathbf{x} \neq 0$ 时，才能保证 $\mathbf{x}^T A \mathbf{x} > 0$ 。因此条件 1 与正定性等价。

记 \mathbf{A}_k 为矩阵 \mathbf{A} 左上角 k 阶方块，取特殊向量 $\mathbf{x} = \begin{bmatrix} x_k \\ 0 \end{bmatrix}$ ，即后 $n-k$ 个元素为 0，则有 $\mathbf{x}^T \mathbf{A} \mathbf{x} = \begin{bmatrix} x_k & 0 \end{bmatrix} \begin{bmatrix} \mathbf{A}_k & * \\ * & * \end{bmatrix} \begin{bmatrix} x_k \\ 0 \end{bmatrix} = \mathbf{x}_k^T \mathbf{A}_k \mathbf{x}_k$ 。若矩阵 \mathbf{A} 满足正定性，则所有 \mathbf{A}_k 均满足正定性。已证明正定性等价于特征值均为正，而矩阵行列式等于特征值之积，因此可知子行列式均大于零。反之亦成立，两命题等价。

第 28 讲 相似矩阵和若尔当标准型

Similar matrices and Jordan form

本讲介绍相似矩阵，这些内容以及奇异值分解是线性代数最核心的概念。

正定矩阵 $A^T A$

若矩阵 A 满足对任意向量 $x \neq 0$ 均有 $x^T A x > 0$ ，则称矩阵为正定矩阵，可以通过特征值、主元和行列式的办法来判断矩阵的正定性。

正定矩阵来自于最小二乘问题。有大量的实际问题用到了长方形矩阵，而最小二乘问题中用到了长方形矩阵的积 $A^T A$ ，它是正定矩阵。

正定矩阵 A 是对称矩阵，它的逆矩阵 A^{-1} 也是正定矩阵，逆矩阵的特征值是原矩阵的倒数，因此也都是正数。若矩阵 A 和 B 都是正定矩阵，则 $A+B$ 也是正定矩阵： $x^T A x > 0$ ， $x^T B x > 0$ ，则有 $x^T (A+B) x > 0$ 。

如果 A 是一个 $m \times n$ 长方形矩阵，则 $A^T A$ 是对称方阵。通过讨论 $x^T (A^T A) x$ 的正负可以确认它是正定矩阵： $x^T (A^T A) x = (Ax)^T (Ax) = |Ax|^2 \geq 0$ 。当且仅当 $Ax=0$ 时，表达式为 0。当矩阵 A 的各列线性无关时，即矩阵为列满秩 $r=n$ ， A 的零空间只有零向量，即此条件下仅有零向量，满足 $x^T (A^T A) x = 0$ 。因此矩阵列满秩时， $A^T A$ 是正定矩阵。正定矩阵将之前的知识点串联起来。

相似矩阵 Similar matrices

A 和 B 均是 $n \times n$ 方阵，若存在可逆矩阵 M ，使得 $B = M^{-1} A M$ ，则 A 和 B 为相似矩阵。

特征值互不相同 Distinct eigenvalues

若矩阵 A 具有 n 个线性无关的特征向量，可以对角化得到 $S^{-1} A S = \Lambda$ ，则 A 相似于 Λ ，这里的 M 是特征向量矩阵 S 。如果将 M 取其它可逆矩阵，可以得到和 A 相似的另一矩阵 B ，实际上这样可以定义一类矩阵， Λ 是其中最简洁的一个。

$$\text{例：} A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \text{ 则 } \Lambda = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}, \text{ 而取另一 } M, \text{ 则有}$$
$$B = \begin{bmatrix} 1 & -4 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 4 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} -2 & -15 \\ 1 & 6 \end{bmatrix}.$$

相似矩阵最重要的特性是：**相似矩阵具有相同的特征值**。事实上，所有特征值为 3 和 1 的二阶矩阵都是 A 的相似矩阵。

证明矩阵 A 的相似矩阵 $B = M^{-1} A M$ ，具有和矩阵 A 相同的特征值 λ ：矩阵 A 具有的特征值 λ ，即存在特征向量 x 满足 $Ax = \lambda x$ 。则有：

$$\begin{aligned} A M M^{-1} x &= \lambda x \\ M^{-1} A M M^{-1} x &= \lambda M^{-1} x \end{aligned}$$

$$BM^{-1}x = \lambda M^{-1}x$$

即矩阵具有特征值 λ ，且特征向量为 $M^{-1}x$ 。

因此，相似矩阵具有相同的特征值，并且线性无关的特征向量的个数相同，但是特征向量往往不同。如果矩阵 A 的特征值互不相等 $\lambda_1 \neq \lambda_2 \neq \dots \neq \lambda_n$ ，而与另一个矩阵 B 的特征值完全相同 $\lambda_1 = \lambda_1', \lambda_2 = \lambda_2', \dots, \lambda_n = \lambda_n'$ ，则它们与相同的对角矩阵 Λ 相似。

重特征值 Repeated eigenvalues

如果矩阵有重特征值，则可能无法进行对角化。

例：二阶矩阵有重特征值 $\lambda_1 = \lambda_2 = 4$ 。

第一类： $\begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$ 只与自己相似， $M^{-1} \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix} M = 4M^{-1}IM = \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$ 。这个系列的相似矩阵仅包含其自身。

第二类包含其它所有的重特征值为4的矩阵：其中最简洁的是 $\begin{bmatrix} 4 & 1 \\ 0 & 4 \end{bmatrix}$ ，元素1的位置换上其它数值仍然是相似矩阵。这个最优形式称为若尔当（Jordan form）标准型。有了这个理论，就可以处理不可对角化的矩阵，完成近似的“对角化”转化为若尔当标准型进行处理。

与 $\begin{bmatrix} 4 & 1 \\ 0 & 4 \end{bmatrix}$ 相似的矩阵，迹为8，行列式为16，因此我们可以构造出很多相似矩阵： $\begin{bmatrix} 5 & 1 \\ -1 & 3 \end{bmatrix}$ ， $\begin{bmatrix} 4 & 0 \\ 17 & 4 \end{bmatrix}$ ， $\begin{bmatrix} a & * \\ * & 8-a \end{bmatrix}$它们都不能对角化（因为若可以对角化则按照特征值可知结果为 $4I$ ，而 $4I$ 只与自己相似）。

若尔当标准型 Jordan form

更复杂的情况，一个四阶矩阵具有重特征值0， $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = 0$ 。

$$A = \left[\begin{array}{cccc|c} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array} \right] \text{ 它的秩为 } 2, \text{ 因此其零空间的维数为 } 4-2=2, \text{ 而零空间的向}$$

量就是矩阵的特征向量，满足 $Ax=0x$ ，所以矩阵 A 只有两个特征向量。若尔当指出上对角线每增加一个1，矩阵就减掉一个特征向量，本例中特征向量数为 $4-2=2$ 。

$$\text{矩阵 } B = \left[\begin{array}{cccc|c} 0 & 1 & 7 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array} \right] \text{ 与矩阵 } A = \left[\begin{array}{cccc|c} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array} \right] \text{ 为相似矩阵。}$$

但矩阵 $C = \left[\begin{array}{cc|cc} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{array} \right]$ 与 $A = \left[\begin{array}{ccc|c} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \end{array} \right]$ 并不是相似矩阵，两者具有不同

的若尔当块。若尔当块形如 $J_i = \begin{bmatrix} \lambda_i & 1 & 0 & \cdots & 0 \\ 0 & \lambda_i & 1 & \ddots & \vdots \\ 0 & 0 & \ddots & \ddots & 0 \\ \vdots & & \ddots & \ddots & 1 \\ 0 & 0 & \cdots & 0 & \lambda_i \end{bmatrix}$ ，对角线上为重特征值 λ_i ，上

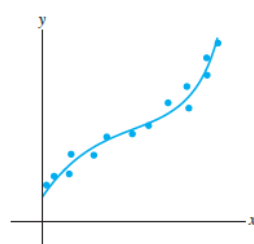
对角线为 1，其它位置的元素均为 0，每个若尔当块只有 1 个特征向量。若干个若尔当块可以拼成一个若尔当矩阵。

若尔当矩阵： $J = \begin{bmatrix} J_1 & 0 & \cdots & 0 \\ 0 & J_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & J_d \end{bmatrix}$

两个矩阵具有相同的特征值和特征向量个数，但是其若尔当块的尺寸不同，两者也并不是相似矩阵。如前述矩阵 A 与 B 并不相似。

若尔当理论：任意 n 阶矩阵 A 都与一个若尔当矩阵 J 相似。若尔当矩阵中的每一个若尔当块对应一个特征向量。若矩阵具有 n 个不同的特征向量 则可以对角化，此时其若尔当标准型 J 就是对角矩阵 Λ 。若出现重特征值，则特征向量个数变少。

说到了 $A^T A$ 和最小二乘问题就要解释一下，GS 举得曲线拟合的例子，都是线性公式 $y=ax+b$ ，但实际上最小二乘法也处理非线性方程，因为这里所谓的非线性是对 x 而言，而只要对于所求的参数是线性方程就可以。比如下面的例子中 x 的方幂组成的矩阵 X 只是一个系数矩阵，对于所求的参数 β 这仍是个线性方程组。



Observation vector: $y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$, Design matrix: $X = \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 \\ 1 & x_2 & x_2^2 & x_2^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 & x_n^3 \end{bmatrix}$, Parameter vector: $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}$, Residual vector: $\epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$

第 29 讲 奇异值分解

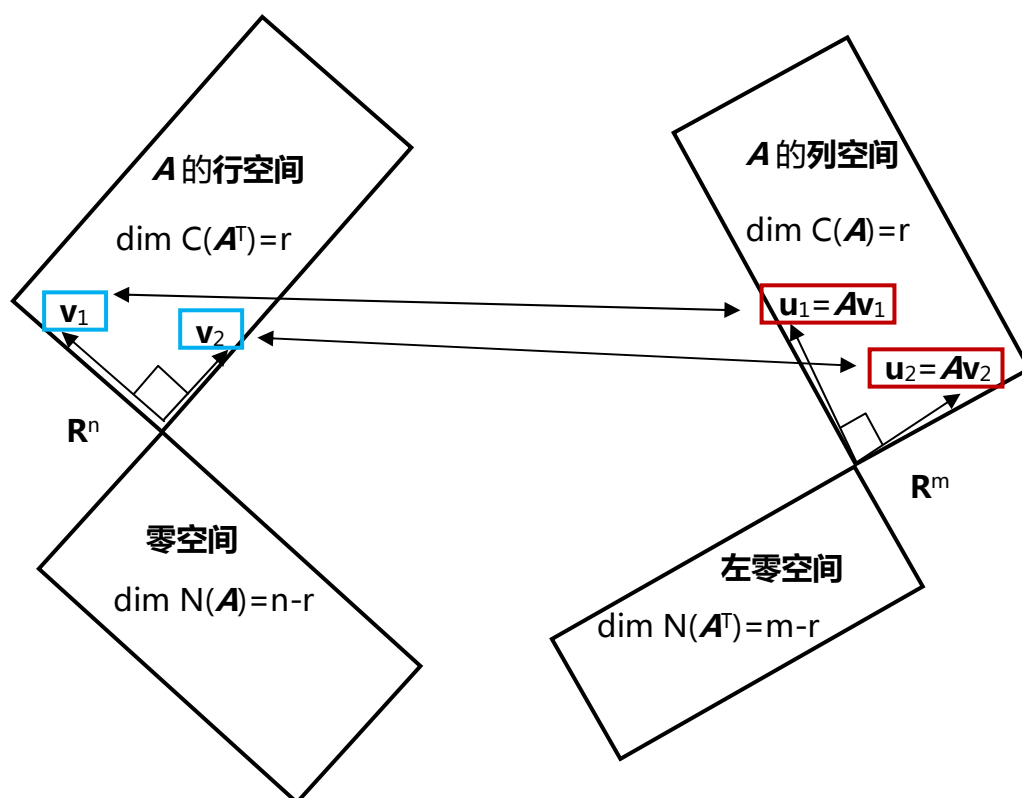
Singular value decomposition

本讲介绍奇异值分解 (Singular value decomposition), 简称 SVD。这是矩阵最终也是最好的分解, 任意矩阵可分解为 $A = U\Sigma V^T$, 分解结果为正交矩阵 U , 对角阵 Σ 和正交矩阵 V 。

如果矩阵 A 是正定矩阵, 它的奇异值分解就是 $A = Q\Lambda Q^T$, 一个正交矩阵 Q 可以满足分解, 而不需要两个。而对于可对角化的矩阵有 $A = SAS^{-1}$, 但特征向量矩阵 S 并不是正交矩阵, 而 SVD 中的 U 和 V 都是正交矩阵。

如何实现

可以将矩阵 A 视为一种线性变换操作, 将其行空间中的一个向量 v_1 变为其列空间中的向量 $u_1 = Av_1$ 。奇异值分解就是要在行空间中寻找一组正交基, 将其通过矩阵 A 线性变换生成列空间中的一组正交基 $Av_i = \sigma_i u_i$ 。



找出矩阵 A 行空间中的正交基很容易, Gram-Schmidt 正交化过程就可以做到, 但是随便的一组正交基经过矩阵 A 变换得到的向量并不一定正交, 因此满足此要求的行空间的正交基非常特殊。而矩阵 A 零空间的向量所对应的是矩阵 Σ 对角线上的 0 元素, 因此很容易处理。

用矩阵数学语言描述这一过程

问题的核心就是找到行空间中一组特殊的正交基：

$$\begin{aligned} A[\mathbf{v}_1 \quad \mathbf{v}_2 \quad \cdots \quad \mathbf{v}_r] &= [\sigma_1 \mathbf{u}_1 \quad \sigma_2 \mathbf{u}_2 \quad \cdots \quad \sigma_r \mathbf{u}_r] \\ &= [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdots \quad \mathbf{u}_r] \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r \end{bmatrix} \end{aligned}$$

如果加入零空间的部分，等式就变为 $A\mathbf{V} = \mathbf{U}\Sigma$ 。其中零空间对应的正交基 $\mathbf{v}_{r+1}, \dots, \mathbf{v}_n$ ，经过线性变换得到 $A\mathbf{v} = 0$ ，对应 Σ 矩阵中对角线最后的特征值 $\sigma_{r+1} = \sigma_{r+2} = \dots = 0$ 。在等式 $A\mathbf{V} = \mathbf{U}\Sigma$ 两侧右乘 \mathbf{V}^{-1} 得到 $A = \mathbf{U}\Sigma\mathbf{V}^{-1} = \mathbf{U}\Sigma\mathbf{V}^T$ 。

现在的问题就是怎么找到符合要求的向量 \mathbf{v}_i 和 \mathbf{u}_i 。

为了得到这两个正交矩阵，考虑首先解决其中的一个，在等式 $A = \mathbf{U}\Sigma\mathbf{V}^T$ 两侧分别乘上等式 $A^T = \mathbf{V}\Sigma^T\mathbf{U}^T$ 两侧的项：

$$\begin{aligned} A^T A &= \mathbf{V}\Sigma^T\mathbf{U}^T\mathbf{U}\Sigma\mathbf{V}^T \\ &= \mathbf{V}\Sigma^T\Sigma\mathbf{V}^T \\ &= \mathbf{V} \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_r^2 \end{bmatrix} \mathbf{V}^T \end{aligned}$$

上式其实是正定矩阵 $A^T A$ 的正交分解， \mathbf{v}_i 就是矩阵 $A^T A$ 的特征向量， σ_i^2 就是矩阵 $A^T A$ 的特征值，奇异值 σ_i 要取正平方根。用同样的办法也可以求得 \mathbf{U} ，它的列向量就是矩阵 AA^T 的特征向量。

例 1：矩阵 $A = \begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix}$ ，求其 SVD 分解。

矩阵为可逆矩阵，秩为 2，则需要在行空间中求得 $\mathbf{v}_1, \mathbf{v}_2$ ，列空间中求得 $\mathbf{u}_1, \mathbf{u}_2$ ，以及伸缩因子 σ_1, σ_2 。

计算得 $A^T A = \begin{bmatrix} 4 & -3 \\ 4 & 3 \end{bmatrix} \begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix} = \begin{bmatrix} 25 & 7 \\ 7 & 25 \end{bmatrix}$ ，它的特征向量为 $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ 和 $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$ 。标准化得到 $\mathbf{v}_1 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$ ， $\mathbf{v}_2 = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$ ，求得 $\sigma_1^2 = 32$ ， $\sigma_2^2 = 18$ 。

求 \mathbf{U} 的过程可以利用矩阵 AA^T 。

$AA^T = \begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix} \begin{bmatrix} 4 & -3 \\ 4 & 3 \end{bmatrix} = \begin{bmatrix} 32 & 0 \\ 0 & 18 \end{bmatrix}$ ，求得 $\sigma_1^2 = 32$ ， $\sigma_2^2 = 18$ ，它的特征向量为 $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ 和 $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ 。

但是我们直接将 $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ 作为矩阵 U 代入，会产生错误 $U\Sigma V^T = \begin{bmatrix} 4 & 4 \\ 3 & -3 \end{bmatrix}$ 。这是因为确定特征向量的过程中，特征向量反向仍然符合要求，通过现在的方法无法确认向量的符号，但是一旦我们确认 v 的方向之后， u 的方向也就随之确定，将 v 代入 $Av = U\Sigma$ 计算 u 可以避免这种问题。 u 和 v 之间的符号联系在进行 AA^T 的计算时被切断了，而用 $Av = U\Sigma$ 可以避免此问题。

例 2：奇异阵 $A = \begin{bmatrix} 4 & 3 \\ 8 & 6 \end{bmatrix}$ ，求其 SVD 分解。

矩阵的秩为 1，行空间和列空间都是 1 维的。行空间和零空间可以找到一组正交基转换得到列空间和左零空间的一组正交基。

很容易确定 $v_1 = \begin{bmatrix} 0.8 \\ 0.6 \end{bmatrix}$ ， $v_2 = \begin{bmatrix} 0.6 \\ -0.8 \end{bmatrix}$ ， $u_1 = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ ， $u_2 = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 \\ -1 \end{bmatrix}$ 。

$A^T A = \begin{bmatrix} 4 & 8 \\ 3 & 6 \end{bmatrix} \begin{bmatrix} 4 & 3 \\ 8 & 6 \end{bmatrix} = \begin{bmatrix} 80 & 60 \\ 60 & 45 \end{bmatrix}$ ，秩 1 矩阵，很容易求得 $\sigma_1^2 = 125$ ， $\sigma_2^2 = 0$ 。

矩阵的 SVD 分解为： $A = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 & 2 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} \sqrt{125} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0.8 & 0.6 \\ 0.6 & -0.8 \end{bmatrix}$

做奇异值分解就是在矩阵的四个子空间中找到合适的基：

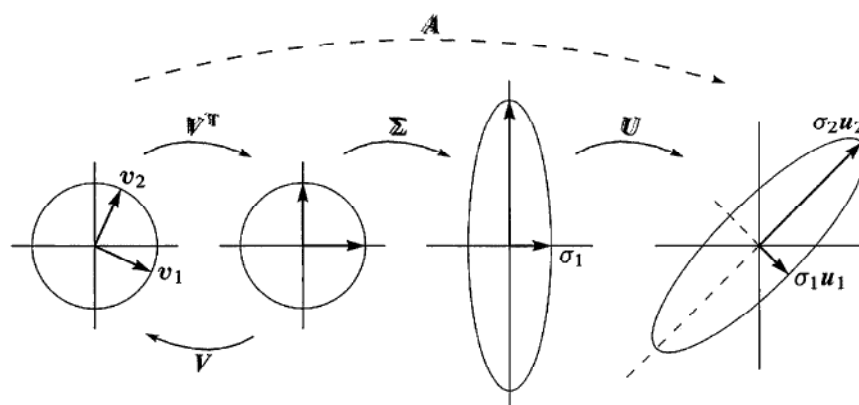
v_1, v_2, \dots, v_r 为行空间的标准正交基。

u_1, u_2, \dots, u_r 为列空间的标准正交基。

$v_{r+1}, v_{r+2}, \dots, v_n$ 为零空间的标准正交基。

$u_{r+1}, u_{r+2}, \dots, u_m$ 为左零空间的标准正交基。

奇异值分解在最小二乘法问题中有重要应用，因为在实际问题中常碰到矩阵 A 不是列满秩的状态，因此 $A^T A$ 不可逆，无法用之前的方法求最优解。即使是列满秩的情况当矩阵是超大型矩阵时， $A^T A$ 的计算量太大，用奇异值分解的办法会降低计算量。



图为 GS 给出的二阶方阵 SVD 的几何意义。

第 30 讲 线性变换及对应矩阵

Linear transformations and their matrices

线性变换 Linear transformations

本讲介绍线性变换。在传统的线代课程中，线性变换会出现在的比矩阵还要早的位置，实际上可以在脱离坐标和具体数值的情况下讨论线性变换。但是面对科学计算的课题，还是要回到坐标上来。每个线性变换都对应于一个矩阵，矩阵概念的背后就是线性变换的概念。

无论是“线性变换”还是“矩阵”，对于初学者来说都是陌生而且突兀的概念。矩阵看起来直观，但是其运算规则总让人摸不着头脑；线性变换的概念显得抽象一些，但其实也可以描述得很朴素。GS 是从矩阵的概念出发，往求解线性方程组的方向引导大家，他最成功的地方是引入了列图像以及矩阵乘法的行操作、列操作，将矩阵运算变得不那么“没头没脑”。而之前推荐过的《线性代数应该这样学》是从线性变换的角度出发，用朴素的描述在更高的抽象层面讨论，很多原理与法则就变得比较简洁。个人认为两者都优于从行列式出发进入线性代数的路线。在《理解矩阵》中，作者写到矩阵的背后就是线性变换，相似矩阵描述的实际上是同一种线性变换。GS 说理解线性变换的本质就是找到它背后的矩阵。这实际上是一个意思，只是出发点和看问题的角度不同。当具体的矩阵和抽象的线性变换在大脑中合二为一的时候，才是对线代最本质的理解。

例：投影 Projection

抛开矩阵，从线性变换的概念来讨论“投影”。通过线性变换使得平面内的一个向量变为平面内的另一个向量，这种变换关系通常称之为“映射”（mapping）。

$$T: \mathbf{R}^2 \rightarrow \mathbf{R}^2$$

$T(\mathbf{v})$ 就像一个函数，给一个输入，进行线性变换后得到一个输出。比如将二维平面的向量投影到一条直线上。

线性代数只讨论线性变换，而线性变换符合如下规则：对于任意向量 \mathbf{v} ， \mathbf{w} 和标量 c ，

$$T(\mathbf{v} + \mathbf{w}) = T(\mathbf{v}) + T(\mathbf{w}) ; T(c\mathbf{v}) = c T(\mathbf{v})$$

将两者结合就得到： $T(c\mathbf{v} + d\mathbf{w}) = c T(\mathbf{v}) + d T(\mathbf{w})$ 。

反例：平面平移 Shift whole plane

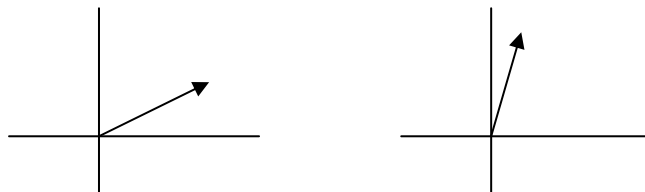
沿着某方向 \mathbf{v}_0 平移一个平面。这就不是一个线性变换。两条法则都违反。更简单的验证方法是线性运算规则的特例 $T(\mathbf{0}) = \mathbf{0}$ ，“平移”不符合这个规则特例。

反例：求长度 $T(\mathbf{v}) = \|\mathbf{v}\|$

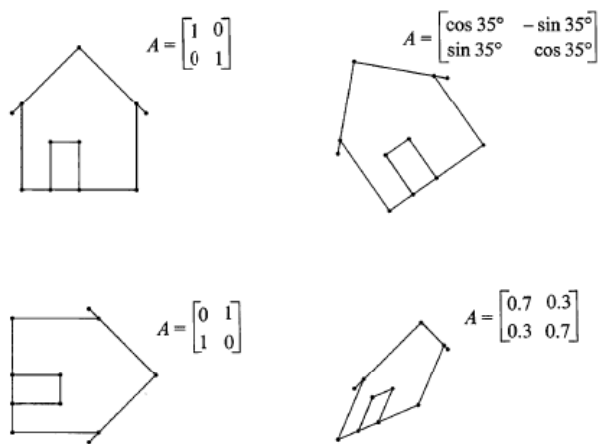
这个线性变换，输入一个三维向量，得到一个数值，或者说一维向量 $T: \mathbf{R}^3 \rightarrow \mathbf{R}^1$ 。这个变换可以满足 $T(\mathbf{0}) = \mathbf{0}$ ，但是数乘负数就不符合规则。

例：旋转 45 度 Rotation by 45°

这个变换 $T: \mathbf{R}^2 \rightarrow \mathbf{R}^2$ ，是一个线性变换，数乘和加法均符合。



对二维平面的图像做线性变换的操作示意图。



例：矩阵

$T(\mathbf{v}) = \mathbf{A}\mathbf{v}$ ，矩阵乘法是一个线性变换。 $\mathbf{A}(\mathbf{v} + \mathbf{w}) = \mathbf{A}(\mathbf{v}) + \mathbf{A}(\mathbf{w})$ ； $\mathbf{A}(c\mathbf{v}) = c\mathbf{A}(\mathbf{v})$ 。整个平面可以通过矩阵的乘法完成变换。

例如，对上图的房子图像施加矩阵 $\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ ，则输出的是上下颠倒的房子。

理解线性变换的本质就是确定它背后的矩阵。

例：

对某一线性变换 $T: \mathbf{R}^3 \rightarrow \mathbf{R}^2$ ，输入一个三维向量而输出是一个二维向量。变成矩阵的形式 $T(\mathbf{v}) = \mathbf{A}\mathbf{v}$ ，则矩阵 \mathbf{A} 是一个 2×3 矩阵。

描述线性变换 Describing $T(\mathbf{v})$

在平面内，如果我们已经了解两个线性无关的向量 \mathbf{v}_1 和 \mathbf{v}_2 经过线性变换的结果 $T(\mathbf{v}_1)$ 和 $T(\mathbf{v}_2)$ ，我们实际上可以通过其线性组合，了解平面内所有的向量线性变换的结果。因此如果我们想了解线性变换对整个输入空间的影响，只需要确定它的一组基 $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ 线性变换的结果。

$$\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \dots + c_n\mathbf{v}_n$$

$$T(\mathbf{v}) = c_1 T(\mathbf{v}_1) + c_2 T(\mathbf{v}_2) + \dots + c_n T(\mathbf{v}_n)$$

线性变换与坐标无关，而矩阵是与坐标有关的。选定一组基，则对于一个向量而言 c_1, c_2 等等就是一组坐标值，给定了将向量表示为基向量线性组合的唯一的表达式。因此可以说坐标源自于一组基， c_1, c_2, \dots, c_n 就是向量的一组坐标值。通常给出空间的坐标是标准坐标，即一组标准基。例如，

$$\mathbf{v} = \begin{bmatrix} 3 \\ 2 \\ 4 \end{bmatrix} = 3 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + 2 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + 4 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

如果用矩阵 \mathbf{A} 来表示线性变换 $T: \mathbf{R}^n \rightarrow \mathbf{R}^m$ 。则需要两组基，即输入空间的一组基和输出空间的一组基，来分别确认输入向量与输出向量的坐标。设 $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ 为输入空间的基，而 $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_m$ 为输出空间的基，用矩阵来表示线性变换就是将向量的坐标乘以矩阵得到它在输出空间的坐标。

例：投影

将二维平面内的向量投影到一条直线，选择输入空间的基向量为 $\mathbf{v}_1, \mathbf{v}_2$ ，其中 \mathbf{v}_1 是沿着投影方向的向量， \mathbf{v}_2 是垂直于投影方向的向量，而输出空间的基选择为 $\mathbf{w}_1 = \mathbf{v}_1, \mathbf{w}_2 = \mathbf{v}_2$ 。对于输入空间中任意向量 \mathbf{v} 有 $\mathbf{v} = c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2$ ，输出为 $T(\mathbf{v}) = c_1 \mathbf{v}_1$ 。

因此这个线性变换的矩阵就是 $\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ ，输入 $\begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$ 得到 $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} c_1 \\ 0 \end{bmatrix}$ 。

这组例子中实际上我们以投影矩阵的特征向量为基，因此得到的矩阵 \mathbf{A} 是对角阵。对角线上就是投影矩阵的特征值 1 和 0。

而如果我们以标准坐标为基，即 $\mathbf{w}_1 = \mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ ， $\mathbf{w}_2 = \mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ 。对应的投影矩阵

就是之前学到的投影矩阵 $\mathbf{P} = \frac{\mathbf{a}\mathbf{a}^T}{\mathbf{a}^T\mathbf{a}}$ 对于投影到斜率 45 度的直线， $\mathbf{P} = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}$ 。

如何确定矩阵 \mathbf{A} Rule to find \mathbf{A}

矩阵 \mathbf{A} 的列实际上是描述输入原空间的基向量得到的列空间线性组合的系数：

$$T(\mathbf{v}_1) = a_{11}\mathbf{w}_1 + a_{21}\mathbf{w}_2 + \dots + a_{m1}\mathbf{w}_m,$$

$$T(\mathbf{v}_2) = a_{12}\mathbf{w}_1 + a_{22}\mathbf{w}_2 + \dots + a_{m2}\mathbf{w}_m,$$

.....

这样矩阵 \mathbf{A} 就满足 “ $\mathbf{A}[\text{输入向量的坐标}] = [\text{输出空间的坐标}]$ ”。

例如 $\mathbf{A} \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix}$ ，所得结果就是输出空间的坐标，也就是输出空间的基进行

线性组合所需要的系数。

介绍一个特别的线性变换——求导， $T = \frac{d}{dx}$ 。

输入： $c_1 + c_2x + c_3x^2$ 基： $1, x, x^2$

输出： $c_2 + 2c_3x$ 。 基： $1, x$

这是一个 $T: \mathbf{R}^3 \rightarrow \mathbf{R}^2$ 的线性变换。

矩阵 A 满足 $A \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} c_2 \\ 2c_3 \end{bmatrix}$ ，可求得矩阵 $A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$ 。

更普遍的来讲，矩阵的逆矩阵就是线性变换的逆变换，矩阵的乘积就是线性变换的乘积，矩阵乘法源自于线性变换。

因为 GS 举了个 $T(v) = Av$ 的例子，所以有人可能晕了，不理解矩阵到底乘以坐标值还是向量本身。简单说明一下，始终记住“ A [输入向量的坐标]=[输出空间的坐标]”，是坐标值！坐标值！坐标值！在 $T(v) = Av$ 的例子中使用的是标准基，因此对 v 做线性变换 Av 所对应的矩阵就是 A 本身。

而如果定义另一个线性变换 $T(v)$ ，对 $v = c_1v_1 + c_2v_2 + \dots + c_nv_n$ 进行线性变换，

用矩阵 A 描述线性变换，则矩阵 A 要满足乘以坐标值 $\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}$ 得到输出向量的坐标值

$\begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_m \end{bmatrix}$ ，用这个坐标值对输出空间的基 w_1, w_2, \dots, w_m 进行线性组合，从而得到输出

向量 $w = d_1w_1 + d_2w_2 + \dots + d_mw_m$ 。

如果我们对输入空间的一个基向量 v_1 做线性变换 $T(v_1)$ ，则矩阵 A 乘以的就是

坐标值 $\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$ 。得到的输出坐标是 $A \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix}$ ，当我们知道线

性变换结果 w 的时候，就可以用它的坐标值确认矩阵 A 第一列的分量。

第 31 讲 基变换和图像压缩

Change of basis; image compression

本讲介绍基变换。选择合适的基向量会给计算制造便利。基变换的一个重要应用就是压缩，图像、影像、音频和其它一些数据都会因为基变换而得到更高效的压缩储存。本讲的主题仍旧是线性变换和矩阵的关联。

图像压缩 Compression of images

本讲涉及的压缩过程是有损压缩。例如一幅像素是 512×512 的静态黑白图像，图像用一个向量来表示，向量的分量 x_i 表示像素的灰度，变化范围 $0 \leq x_i \leq 255$ ，占 8bits。该向量属于 \mathbf{R}^n 空间， $n=(512)^2$ 。彩色图像描述每个点的像素需要三个数据，向量长度是黑白图像的 3 倍。

图像的标准压缩方式为 JPEG（联合图像专家组 Joint Photographic Experts Group）。图像压缩的本质就是基变换。

压缩前图像采用的基向量是标准基。但是在图像中离得很近的区域，颜色是非常接近的，比如教学视频中黑板的一个区域，这些区域像素的灰度值很接近，但是用标准基来存储并没有利用上这一特点，这就给了我们压缩的空间。

标准基就是 $\begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \dots, \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$ ，而显然对于灰度很接近的情况， $\begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$ 是一个很好的

基（代表低频信号，频率为 0，平滑。在图像压缩后经常会存在，因为其系数通常很大。括号里的话是 GS 后来又提到的），在一组基中有一个这样的向量能解决很大的问题，可以处理像素灰度接近一致的情况。但图像不是灰度完全一致的，因此接

下来的问题是跟它相配合的基要选择哪些。极端的情况包括选择 $\begin{bmatrix} +1 \\ -1 \\ +1 \\ -1 \\ \vdots \end{bmatrix}$ ，它可以给出

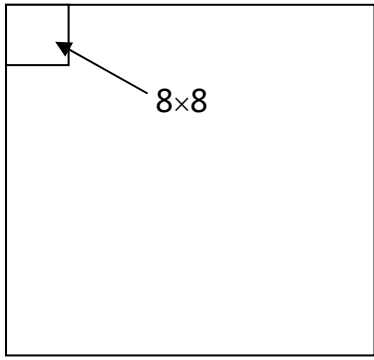
类似国际象棋盘那种黑白相间的状态（最高频信号，噪音、扰动.....在图像压缩后

很少存在），还有图像一半图像暗一半图像亮，可以选择 $\begin{bmatrix} \vdots \\ +1 \\ -1 \\ \vdots \\ -1 \end{bmatrix}$ 。

傅里叶基 Fourier basis

在 JPEG 中,将 512×512 的区域划分为 8×8 的区块进行处理,所用到的基是傅里叶基向量。

以 \mathbf{R}^8 中的基举例

$$\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ \omega \\ \omega^2 \\ \omega^3 \\ \omega^4 \\ \omega^5 \\ \omega^6 \\ \omega^7 \end{bmatrix}, \begin{bmatrix} 1 \\ \omega^2 \\ \omega^4 \\ \omega^6 \\ \omega^8 \\ \omega^{10} \\ \omega^{12} \\ \omega^{14} \end{bmatrix}, \dots, \begin{bmatrix} 1 \\ \omega^7 \\ \omega^{14} \\ \omega^{21} \\ \omega^{28} \\ \omega^{35} \\ \omega^{42} \\ \omega^{49} \end{bmatrix}$$


傅里叶基 (Fourier basis) 就是之前讲过的傅里叶矩阵的列向量,每个元素为复数的幂。在 8×8 区块中有 64 个系数,64 个基向量,在这个 64 维空间中做基变换。

$$\text{信号 } \mathbf{x} \xrightarrow{\text{无损}} \text{系数 } \mathbf{c} \xrightarrow{\text{有损}} \text{系数 } \hat{\mathbf{c}} \text{ (很多的 0)} \longrightarrow \hat{\mathbf{x}} = \sum \hat{c}_i \mathbf{v}_i$$

首先对输入的信号 \mathbf{x} ,从标准基变换为傅里叶基,得到系数 \mathbf{c} ,这一步是无损的过程。之后设置阈值进行压缩,超过阈值的认定为肉眼无法分辨的信号。在数学上表现为某些基向量的系数很小,这部分可以丢弃,随之得到新的系数 $\hat{\mathbf{c}}$ 。将新的系数赋值在傅里叶基上求和得到 $\hat{\mathbf{x}} = \sum \hat{c}_i \mathbf{v}_i$ 。此时求和项已经不是 64 项,可能只剩下两三项,这就是压缩。

视频文件可以视为图像的序列,一幅一幅进行图像压缩即可。但这样做没有利用好视频的性质,因为视频是连续的,一幅图像和下一幅图像非常接近,因此可以存储一幅基础图像,随后只存储下一幅图像对它的修正部分。

小波 Wavelets

下面介绍另一组和傅里叶竞争的基向量——小波。

以 \mathbf{R}^8 中的基举例

$$\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ -1 \\ -1 \\ -1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ -1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \dots, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ -1 \end{bmatrix}.$$

这个只是一个小波选择,还有很多种更多精细的选择,这一组基中有太多从 +1

跳转到-1 的变化。线性代数要做的基变换，就是将标准基下的向量 $\mathbf{p}=[p_1, p_2, p_3, \dots, p_8]^T$ 表示为小波基的线性组合，求出线性组合的参数 \mathbf{c} 满足 $\mathbf{p}=\mathbf{c}_1\mathbf{w}_1+\mathbf{c}_2\mathbf{w}_2+\dots+\mathbf{c}_8\mathbf{w}_8$ ，即 $\mathbf{p}=\mathbf{W}\mathbf{c}$ 。 \mathbf{W} 即以小波基向量为列向量的小波矩阵。因此有 $\mathbf{c}=\mathbf{W}^{-1}\mathbf{p}$ 。

好的基向量组要求：第一，可以快速求逆矩阵，例如快速傅里叶变换，这里也存在快速小波变换，因为小波矩阵列向量正交，因此可以转置得到逆矩阵；第二，要少量基向量就可以近似信号，可压缩的比例就比较高。

基变换 Change of basis

\mathbf{W} 的列向量是一组新的基向量。在旧基向量体系下的向量 \mathbf{x} 和新基向量下的 \mathbf{c} 的关系是： $\mathbf{x}=\mathbf{W}\mathbf{c}$ 。

变换矩阵

已知一个线性变换 $T: \mathbf{R}^8 \rightarrow \mathbf{R}^8$ 。当使用空间的一组基 $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_8$ 时，线性变换对应的矩阵为 \mathbf{A} ；当使用一组新的基 $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_8$ 时，线性变换对应的矩阵为矩阵 \mathbf{B} 。两个矩阵对应的是同一个线性变换，只是使用了不同的基向量。那么 \mathbf{A} 和 \mathbf{B} 为相似矩阵 $\mathbf{B}=\mathbf{M}^{-1}\mathbf{A}\mathbf{M}$ ， \mathbf{M} 是基变换矩阵。

复习一下线性变换的内容：

$$\begin{aligned}\mathbf{x} &= \mathbf{c}_1\mathbf{v}_1 + \mathbf{c}_2\mathbf{v}_2 + \dots + \mathbf{c}_8\mathbf{v}_8 \\ T(\mathbf{x}) &= \mathbf{c}_1 T(\mathbf{v}_1) + \mathbf{c}_2 T(\mathbf{v}_2) + \dots + \mathbf{c}_8 T(\mathbf{v}_8) \\ T(\mathbf{v}_1) &= a_{11}\mathbf{v}_1 + a_{21}\mathbf{v}_2 + \dots + a_{81}\mathbf{v}_8 \\ T(\mathbf{v}_2) &= a_{12}\mathbf{v}_1 + a_{22}\mathbf{v}_2 + \dots + a_{82}\mathbf{v}_8 \\ &\dots\end{aligned}$$

矩阵 \mathbf{A} 就是 a_{ij} 组成的矩阵。

如果我们使用的基向量就是特征向量 $T(\mathbf{v}_i)=\lambda_i\mathbf{v}_i$ ，矩阵 \mathbf{A} 就变成对角阵 $\mathbf{\Lambda}$ 。找出特征向量是压缩最理想的结果，但是找出图像的特征向量代价太大，因此我们找到代价小但是接近理想状态的基向量（例如小波基）进行基变换，完成压缩过程。

再强调一次，与矩阵 \mathbf{A} 做运算的是坐标值，运算结果也是坐标值。此处线性变换中输入和输出空间都是同一个空间，矩阵 \mathbf{A} 对应的是线性变换前后都使用一组基 $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_8$ 的情形，而矩阵 \mathbf{B} 对应的是线性变换前后都用基 $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_8$ 的情形。跟上一讲变换前用 \mathbf{v} 变换后用 \mathbf{w} 是不同的。

第 32 讲 复习三

Exam 3 review

主要内容

- 特征值和特征向量 $A\mathbf{x}=\lambda\mathbf{x}$
- 微分方程 $d\mathbf{u}/dt=A\mathbf{u}$ 和矩阵指数形式 e^{At}
- 对称矩阵 $A=A^T$ 的特征值永远为实数，总有足够的特征向量实现矩阵的对角化 $A=Q\Lambda Q^T$ 。
- 正定矩阵
- 相似矩阵 $B=M^{-1}AM$ ，两矩阵具有相同的特征值，但是特征向量不同，矩阵的乘方形式很接近 $B^k=M^{-1}A^kM$ 。
- 奇异值分解 SVD， $A=U\Sigma V^T$ 。

例题：

1) 微分方程：

$$\frac{d\mathbf{u}}{dx} = A\mathbf{u} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \mathbf{u}$$

a) 求通解为 $\mathbf{u}(t) = c_1 e^{\lambda_1 t} x_1 + c_2 e^{\lambda_2 t} x_2 + c_3 e^{\lambda_3 t} x_3$

答：首先求处矩阵的特征值，矩阵为奇异矩阵，因此具有一个特征值 0。矩阵为反对称矩阵，因此特征值为纯虚数。

$$|A - \lambda I| = \begin{vmatrix} -\lambda & -1 & 0 \\ 1 & -\lambda & -1 \\ 0 & 1 & -\lambda \end{vmatrix} = -\lambda^3 - 2\lambda = 0$$

解得 $\lambda_1=0$ ， $\lambda_2=\sqrt{2}i$ ， $\lambda_3=-\sqrt{2}i$ 。

则通解形式为 $\mathbf{u}(t) = c_1 x_1 + c_2 e^{\sqrt{2}it} x_2 + c_3 e^{-\sqrt{2}it} x_3$ ，解既不发散也不收敛，纯虚数特征值使得解的分量始终在单位圆上循环。

b) 解为周期性函数，何时返回初值？（周期是多少）

解： $\sqrt{2}Ti = 2\pi i$ ，所以有 $T = \sqrt{2}\pi$ 。

c) 求出两个特征向量，证明其正交？

满足 $A^T A = A A^T$ 的矩阵具有正交的特征向量。对称和反对称矩阵都具有正交的特征向量。正交矩阵也符合此要求。

答：将特征值代入可以求得 $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$, $\mathbf{x}_2 = \begin{bmatrix} -1 \\ \sqrt{2}i \\ 1 \end{bmatrix}$, $\mathbf{x}_3 = \begin{bmatrix} 1 \\ \sqrt{2}i \\ -1 \end{bmatrix}$ 。可以用内积验证其正交性，要注意对复向量取内积要做共轭。

d) 微分方程的解写成矩阵指数形式为 $\mathbf{u}(t) = e^{At}\mathbf{u}(0)$ ，如何计算 e^{At} ？

答：若有 $\mathbf{A} = \mathbf{S}\mathbf{\Lambda}\mathbf{S}^{-1}$ 。则有 $e^{At} = \mathbf{S}e^{\mathbf{\Lambda}t}\mathbf{S}^{-1}$ 。

$$\text{其中 } e^{\mathbf{\Lambda}t} = \begin{bmatrix} e^{\lambda_1 t} & 0 & \dots & 0 \\ 0 & e^{\lambda_2 t} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \dots & 0 & e^{\lambda_n t} \end{bmatrix}。$$

2) 一个 3×3 矩阵 \mathbf{A} 具有特征值 $\lambda_1 = 0$ $\lambda_2 = c$ $\lambda_3 = 2$ 。对应的特征向量为 $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ ，

$$\mathbf{x}_2 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \mathbf{x}_3 = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}。$$

a) 求 c 使得矩阵可对角化。

答：矩阵具有足够多的线性无关的特征向量时，可以完成对角化。此处的特征向量不仅线性无关而且正交，因此任意 c 都可以满足矩阵可对角化。

b) 求 c 使得矩阵对称？

答：如果矩阵对称 $\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$ ，则其特征值为实数，因此 c 为实数时，此矩阵为对称矩阵。

c) 求 c 使得矩阵正定？

答：有一个特征值为 0，矩阵不可能正定。 c 为实数且 $c \geq 0$ 时，矩阵半正定。

d) 此矩阵可能是 Markov 矩阵么？

答：Markov 矩阵具有一个特征值 1，其它特征值小于 1。此矩阵具有特征值 2，因此不可能是 Markov 矩阵。

e) 矩阵 $\mathbf{P} = (1/2)\mathbf{A}$ 可能是投影矩阵么？

答：投影矩阵的特征值为 1 或者 0，所以当 $c = 0$ 或者 2 时，矩阵 $\mathbf{P} = (1/2)\mathbf{A}$ 是投影矩阵。

3) 对于矩阵 \mathbf{A} 可做奇异值分解 $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ ，已知 $\mathbf{\Sigma} = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$ ，并且矩阵 \mathbf{U} 和 \mathbf{V}

均为 2 列。

a) 对于矩阵 \mathbf{A} 我们知道些什么？

答：矩阵 \mathbf{A} 为二阶可逆矩阵。

b) 如果 $\Sigma = \begin{bmatrix} 3 & 0 \\ 0 & -5 \end{bmatrix}$ 呢？

答：不可能，因为奇异值分解要求奇异值大于 0。

c) 如果 $\Sigma = \begin{bmatrix} 3 & 0 \\ 0 & 0 \end{bmatrix}$ 呢？

答：矩阵为奇异阵，秩为 1，零空间的维数为 1。四个子空间由 \mathbf{U} 和 \mathbf{V} 的列向量张成。例如 \mathbf{v}_2 就在矩阵的零空间中。

4) 矩阵 \mathbf{A} 对称并且正交。

a) 我们对它的特征值知道什么？

答：对称矩阵的特征值为实数，正交矩阵的特征值的绝对值为 1。因此矩阵的特征值为 +1 或者 -1。

b) 矩阵 \mathbf{A} 是否为正定矩阵？

答：特征值可以是 -1，非正定。

c) 矩阵 \mathbf{A} 没有重特征值？

答：只能是 +1 或者 -1，因此 3 阶以上肯定有重特征值。

d) 矩阵 \mathbf{A} 可对角化么？

答：所有的对称矩阵和所有的正交矩阵都可对角化。

e) 矩阵 \mathbf{A} 可逆么？

答：正交矩阵都可逆。

f) 证明 $\mathbf{P} = (1/2)(\mathbf{A} + \mathbf{I})$ 是投影矩阵。

答： $\mathbf{P}^2 = (\frac{1}{2}(\mathbf{A} + \mathbf{I}))^2 = \frac{1}{4}(\mathbf{A}^2 + 2\mathbf{A} + \mathbf{I})$ 。其中 $\mathbf{A}^2 = \mathbf{A}^T \mathbf{A} = \mathbf{I}$ ，所以 $\mathbf{P}^2 = \frac{1}{2}(\mathbf{A} + \mathbf{I}) = \mathbf{P}$ 。因此它可以是投影矩阵。此外 $(1/2)(\mathbf{A} + \mathbf{I})$ 的特征值为 1 或者 0，并且是对称矩阵，这亦可证明它是投影矩阵。

第 33 讲 左右逆和伪逆

Left and right inverses; pseudoinverse

本节主要介绍左右逆矩阵和伪逆矩阵。

两侧逆矩阵 2-sided inverse

矩阵 A 的两侧逆矩阵 A^{-1} 满足 $AA^{-1}=I=A^{-1}A$ 。这就是我们通常说矩阵 A 的逆矩阵。此时 $r=m=n$ ， A 为满秩方阵。

左逆矩阵 Left inverse

矩阵 A 的列满秩 $r=n$ ，列向量线性无关。则矩阵的零空间 $N(A)$ 只有零向量，方程 $Ax=b$ 无解或者有唯一解。当矩阵 A 列满秩的时候，矩阵 $A^T A$ 为可逆矩阵，这是讨论最小二乘问题的核心。

矩阵 $A^T A$ 是 $n \times n$ 的对称矩阵且满秩，因此 $A^T A$ 是可逆矩阵，即 $(A^T A)^{-1} A^T A = I$ 。我们称 $A_{\text{left}}^{-1} = (A^T A)^{-1} A^T$ 为 A 的左逆矩阵。 A_{left}^{-1} 为 $n \times m$ 矩阵，而 A 为 $m \times n$ 矩阵， $A_{\text{left}}^{-1} A = I$ 是 $n \times n$ 矩阵。

右逆矩阵 Right inverse

矩阵 A 的行满秩 $r=m$ ，行向量线性无关。则矩阵的左零空间 $N(A^T)$ 只有零向量。 A 的零空间维数为 $n-m$ ，因此有 $n-m$ 个自由变量， n 大于 m 时，方程 $Ax=b$ 有无穷多解。

矩阵 A 的右逆矩阵为 $A_{\text{right}}^{-1} = A^T (AA^T)^{-1}$ 。

通常情况下右乘左逆矩阵得不到单位阵， $AA_{\text{left}}^{-1} = A(A^T A)^{-1} A^T = P$ ，这是列空间的投影矩阵。仅在 $m=n$ 的条件下， $AA_{\text{left}}^{-1} = I$ 。一个长方形矩阵 A 不可能有两侧逆，因为 A 和 A^T 总有一个零空间的维数不是 0。

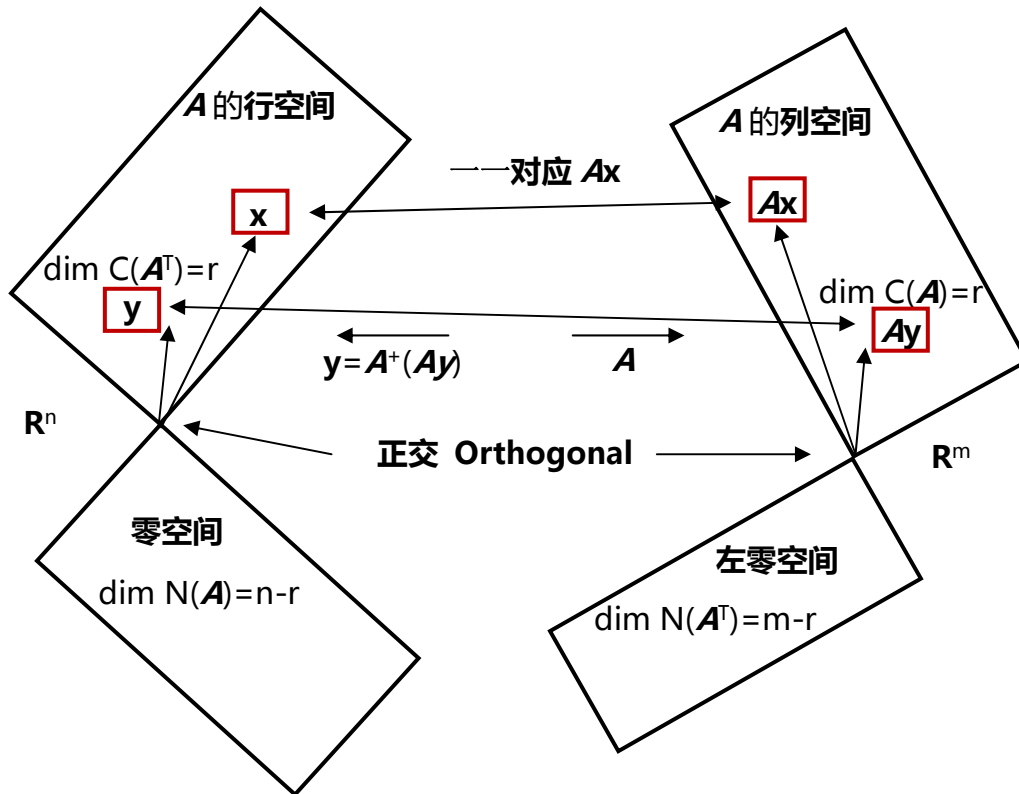
同样的，左乘右逆矩阵得到的是 $A_{\text{right}}^{-1} A = A^T (AA^T)^{-1} A$ ，这是向行空间投影的矩阵，投影矩阵在投影空间内表现的就像单位阵 I 一样。

伪逆矩阵 Pseudoinverse

可逆矩阵的零空间和左零空间只有零向量。列满秩的矩阵的零空间只有零向量，行满秩的矩阵的左零空间只有零向量。但对于不满秩的矩阵 ($r < n, r < m$)，其两个零空间均存在，这就使得它取不到逆矩阵。因为逆矩阵的作用可以看作是原矩阵的逆操作，但是矩阵 A 对其零空间中向量操作后变为 0 ，这时没有逆操作能够恢复这一过程，所以带有零空间就不能取逆矩阵。

观察不满秩矩阵 A 的四个子空间，其行空间和列空间的维数相等，均为 r 。在其行空间中的向量 x 经过矩阵 A 操作后，变为列空间中的向量 Ax 。而 x 与 Ax 为

一一对应的关系。如果将矩阵 A 限制在行空间和列空间上，它是个可逆矩阵，此时 A 的逆矩阵就是所谓的“伪逆矩阵 A^+ ”。



问题的关键就是一一对应，即对于行空间中的向量 $x \neq y$ ，其通过矩阵 A 映射到列空间得到的向量 $Ax \neq Ay$ 。

证明：若行空间中存在向量 y 与向量 x 在列空间中对应的向量相同，则有 $Ax - Ay = 0$ ，即 $A(x - y) = 0$ ，则 $x - y$ 为矩阵 A 零空间中的向量，但矩阵的行空间对线性运算封闭，因此 $x - y$ 为行空间中的向量，因两个子空间正交，所以有 $(x - y) = 0$ ，即 x 与 y 相同。因此行空间与列空间中的向量为一一对应。

统计学家非常需要伪逆矩阵，因为他们要做很多最小二乘法进行线性回归的问题。如果矩阵不满秩，则 $A^T A$ 不可逆，所以无法用之前的办法解决，此时要用到伪逆矩阵。

求伪逆矩阵 A^+ 的一个方法是利用奇异值分解 $A = U \Sigma V^T$ ，其中对角阵 Σ 两是由矩阵奇异值排列在对角线上构成的 $m \times n$ 矩阵，其秩为 r 。则其伪逆矩阵 Σ^+ 为 $n \times m$ 矩阵，矩阵的秩也为 r 。

$$\Sigma = \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & \end{bmatrix}, \Sigma^+ = \begin{bmatrix} 1/\sigma_1 & & & \\ & \ddots & & \\ & & 1/\sigma_r & \\ & & & \end{bmatrix}$$

$$\Sigma \Sigma^+ = \begin{bmatrix} 1 & & \\ & \ddots & \\ & & 1 \\ & & & 0 \end{bmatrix}, \Sigma^+ \Sigma = \begin{bmatrix} 1 & & \\ & \ddots & \\ & & 1 \\ & & & 0 \end{bmatrix}$$

矩阵 Σ 右乘伪逆矩阵得到 $m \times m$ 矩阵，矩阵对角线前 r 个元素为 1，其它元素均为 0，而左乘伪逆矩阵得到 $n \times n$ ，矩阵对角线前 r 个元素为 1，其它元素均为 0，这两个都是 Σ 的投影矩阵，一个投影到行空间，一个投影到列空间。

而矩阵 A 的伪逆矩阵为 $A^+ = V \Sigma^+ U^T$ 。

逆矩阵满足四个性质：

$$AA^+A = A$$

$$A^+AA^+ = A^+$$

$$AA^+ = (AA^+)^T$$

$$A^+A = (A^+A)^T$$

注意： $AA^+ = U \Sigma V^T V \Sigma^+ U^T = U \Sigma \Sigma^+ U^T$ 得到的并不是形如 $\Sigma \Sigma^+$ 这种对角线上只有 1 和 0 的对角阵，所得结果是 A 行空间的投影矩阵。

例如 $A = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ ，通过奇异值分解计算可以得到 $A^+ = \begin{bmatrix} 1/2 & 1/2 \end{bmatrix}$ 。而 $AA^+ = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}$ 。

第 34 讲 总复习

Course Review

本讲为线性代数课程总复习，复习的方法就是做往年试题。

1) 已知 $\mathbf{Ax} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ 无解， $\mathbf{Ax} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ 仅有一解。

a) 这个矩阵为 $m \times n$ 矩阵，秩为 r ，那么 m 是多少？ r ？

答： $m=3$ 。 $\mathbf{Ax} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ 无解说明 r 小于 m ， $\mathbf{Ax} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ 仅有一解说明矩阵零空间只

有零向量，所以 $r=n$ 。例如 $\mathbf{A} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ 或者 $\mathbf{A} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$ 。

b) 判断 $\det(\mathbf{A}^T \mathbf{A}) = \det(\mathbf{A} \mathbf{A}^T)$ 是否成立？

答：不成立。可以看下面两题的答案，分别是可逆和不可逆的矩阵。需要注意的是：只有当矩阵是方阵时，才有行列式的乘积等于乘积的行列式。

c) $\mathbf{A}^T \mathbf{A}$ 是否可逆？

答：可逆，矩阵 \mathbf{A} 列满秩 $r=n$ ，因此 $\mathbf{A}^T \mathbf{A}$ 可逆。

d) $\mathbf{A} \mathbf{A}^T$ 是否正定？

答：否，不满秩，不可能是正定矩阵。

e) $\mathbf{A}^T \mathbf{y} = \mathbf{c}$ ，证明对于任意右侧向量至少有一个解。

答：矩阵 \mathbf{A}^T 为 $n \times 3$ 矩阵，行满秩，矩阵 \mathbf{A}^T 的零空间维数为 $3-n$ ，方程有无穷多解。

2) 已知矩阵 $\mathbf{A} = [\mathbf{v}_1 \ \mathbf{v}_2 \ \mathbf{v}_3]$

a) 求解 $\mathbf{Ax} = \mathbf{v}_1 - \mathbf{v}_2 + \mathbf{v}_3$

答： $\mathbf{x} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$

b) 若 $\mathbf{v}_1 - \mathbf{v}_2 + \mathbf{v}_3 = \mathbf{0}$ ，则上一题的解是否不唯一？

答：列向量线性相关，零空间有非零向量， $\mathbf{Ax} = \mathbf{0}$ 解不唯一。

c) 若 \mathbf{v}_1 ， \mathbf{v}_2 和 \mathbf{v}_3 为标准正交，求 \mathbf{v}_1 ， \mathbf{v}_2 线性组合距离 \mathbf{v}_3 最近？

答：原点。

3) 已知 Markov 矩阵 $A = \begin{bmatrix} 0.2 & 0.4 & 0.3 \\ 0.4 & 0.2 & 0.3 \\ 0.4 & 0.4 & 0.4 \end{bmatrix}$ 。

a) 特征值？

答：列 1+列 2=2×列 3，所以矩阵为奇异阵，有一个特征值 $\lambda_1=0$ 。Markov 矩阵还有一个特征值 $\lambda_2=1$ ，从迹可知最后一个特征值为 $\lambda_3=-0.2$ 。

b) $u_k = A^k \begin{bmatrix} 0 \\ 10 \\ 0 \end{bmatrix}$ ，求 k 步之后的状态？当 k 趋近于无穷时的状态？

答： $u_k = c_1 \lambda_1^k x_1 + c_2 \lambda_2^k x_2 + c_3 \lambda_3^k x_3$ 。代入特征值可知，当 k 趋近于无穷时，只剩下 $c_2 x_2$ 。求解矩阵的特征向量 x_2 ，得到 $x_2 = \begin{bmatrix} 3 \\ 3 \\ 4 \end{bmatrix}$ 。 u_0 引入 10 个人 (Markov 矩阵可以当作人口流动模型来看)，并且最终的分配比例为 3:3:4，所以不用逐个求解特征向量和 c，可得 $u_k = \begin{bmatrix} 3 \\ 3 \\ 4 \end{bmatrix}$ 。

4) 求符合题目要求的 2×2 矩阵。

a) 投影到 $a = \begin{bmatrix} 4 \\ -3 \end{bmatrix}$ 所在直线的投影矩阵。

答： $P = \frac{aa^T}{a^T a} = \dots$

b) 矩阵具有特征值 0 和 3，对应的特征向量分别为 $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ 和 $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ 。

答： $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}^{-1} = \dots$

c) 矩阵 A 不能分解为 $B^T B$ 。

答：给出一个不对称矩阵就可以了。

d) 矩阵有正交的特征向量，但是非对称。

答：反对称矩阵或者在复数域上有正交特征向量的其它矩阵。

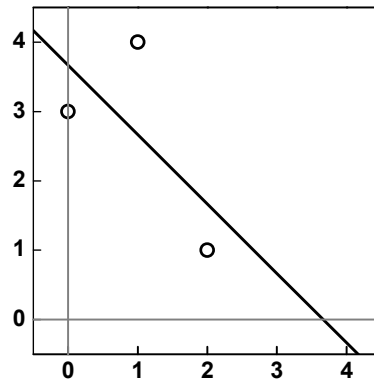
5) 已知 $\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} \hat{C} \\ \hat{D} \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \\ 1 \end{bmatrix}$ ， $\begin{bmatrix} \hat{C} \\ \hat{D} \end{bmatrix} = \begin{bmatrix} 11/3 \\ -1 \end{bmatrix}$ 。

a) 求投影 p 。

答：把 $\begin{bmatrix} \hat{C} \\ \hat{D} \end{bmatrix} = \begin{bmatrix} 11/3 \\ -1 \end{bmatrix}$ 代入得到 $\mathbf{p} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 11/3 \\ -1 \end{bmatrix} = \begin{bmatrix} 11/3 \\ 8/3 \\ 5/3 \end{bmatrix}$

b) 画图，三个点和拟合曲线。

答：



c) 找出另一个非零向量 \mathbf{b} ，使得 $\begin{bmatrix} \hat{C} \\ \hat{D} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ 。

答 求解的 \mathbf{b} 要与系数矩阵的列空间正交，才会使得最优解为 $\mathbf{0}$ 。解得 $\mathbf{b} = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$ 。