思考十《Introduction to Linear Algebra》 Date No.
edition 3

前言：学习数学最重要的principle是抓其本质，~~化繁为简~~"化繁为简"。

一、向量和运算

1. 向量的本质

在我学基础代数学的时候，向量代数的本质被归结为"空间本质的线性化"。话没错，但"空间"的意思容易被误解。比如，我们从二维、三维空间推向量运算，可以得到正确的（见下面线性性的说明），却不利于多维的理解！

$$\left\{\begin{array}{l}\text{加法}\\ \text{倍积}\\ \text{内积}\end{array}\right\} \text{满足线性性 [1]，交换率}$$

[1] 倍积：$k\cdot\vec{a}$。假设函数 $f(x)$ 满足线性性，则需满足以下两个条件

(i) $f(a+b) = f(a) + f(b)$　(ii) $f(ka) = kf(a)$

如此一来则 $f(k_1 a + k_2 b) = k_1 f(a) + k_2 f(b)$

检验 假设运算对象为向量时，倍积运算是否满足线性性：即 $f(\vec{a}) = k\vec{a}$ 时，容易得到

(i) $k(\vec{a}+\vec{b}) = k\vec{a} + k\vec{b}$，$k\cdot k_1\vec{a} = k\cdot k_1(\vec{a})$

∴ 向量的倍积运算满足线性性，即有

$$k(c_1\vec{a} + c_2\vec{b}) = \cancel{k c_1\vec{a} + k c_2\vec{b}}\ c_1 k\vec{a} + c_2 k\vec{b}$$

内积同样如此。$\cancel{f}$ $f(\vec{x}) = \vec{a}\cdot\vec{x}$

则 (i) $\vec{a}(\vec{x}_1 + \vec{x}_2) = \vec{a}\vec{x}_1 + \vec{a}\vec{x}_2$　(ii) $k(\vec{a}\vec{x}) = k\vec{a}\cdot\vec{x}$

我们有 $\vec{a}\cdot(c_1\vec{x}_1 + c_2\vec{x}_2) = c_1\vec{a}\vec{x}_1 + c_2\vec{a}\vec{x}_2$

二维、三维几何 $\xrightarrow{\text{推导}}$ 向量运算　三种运算确实有很好的几何意义，如 $\vec{a}\cdot\vec{b} = \frac{1}{2}(|\vec{a}+\vec{b}|^2 - |\vec{a}|^2 - |\vec{b}|^2)$

（$|\vec{a}+\vec{b}|^2 = |\vec{a}|^2 + |\vec{b}|^2 + 2\vec{a}\vec{b}$，$\vec{a}\vec{b} = |\vec{a}||\vec{b}|\cos\theta$）　但！
斜边　　下一页

① 线性性的本质，一种运算，首先可以提~~…~~分离，其次可以提常数

Date · · No.

知道了点积的本质之后，我们再看下二维情况下的一些概念在多维空间中是否有类似的意义（当然是由点积定义的量）

Length $= \|\vec{v}\| = \sqrt{\vec{v}\cdot\vec{v}}$ (or called "Norm")

Unit vector $\vec{u}$，则 $\sqrt{\vec{u}\cdot\vec{u}} = 1 \Rightarrow \vec{u}\cdot\vec{u} = 1$

angle between two vector：$\cos\theta = \dfrac{\vec{v}\cdot\vec{w}}{\|\vec{v}\|\,\|\vec{w}\|}$

↳ 它们在多维空间中有什么意义呢？

么歌影间 ←

  对美：2个向量，夹角越小，相似度越高 $\dfrac{D\cdot Q}{\|D\|\,\|Q\|}$

D&Q 相同位置为1个，不同位置为1 ↓ ⇒ $\cos\theta$ ↑

运算

关于线性性的再一步说明： 矩阵也满足线性性

  为了使矩阵运算也有线性性，我们有了[1]矩阵的

  加法、倍积，乘法

  符合线性方程组理论 ← $(m\times n,\ n\times 1)$ 运算的定义和性质见前文（计算）

  $(m\times n,\ n\times k)$ 推导见文后附

∴有性质

$$C(A+B) = CA + CB$$

$$C(kA) = kCA$$

$$C(k_1A + k_2B) = k_1CA + k_2CB$$

另外，根据矩阵相乘的结合律

$(AB)C = A(BC)$

$= ABC$ ②

① 具体如何得到，见前文

② 从②可以得到 $A^pA^q = A^{p+q}$

BLOCK

Date No.

Matrix × Matrix 的两种看法之一

这种 $A[b_1, b_2, \cdots b_n] = [Ab_1, Ab_2, \cdots Ab_n]$ 的看法

同样也说明 $I\underset{\text{vector}}{b}=b$ 成立，$I\underset{\text{matrix}}{B}=B$ 也成立

更深入一步，我们可以得到分块矩阵算法：

$$AB = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} = \begin{bmatrix} A_{11}B_{11} + A_{12}B_{21} & \cdots \\ A_{21}B_{11} + A_{22}B_{21} & \cdots \end{bmatrix}$$

证明：先将B分列

$$\begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \Rightarrow \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}, \text{ 根据 } AB = A[B_1, B_2] = [AB_1, AB_2] \quad \cdots ①$$

$$\quad\; B_1 \quad B_2$$

接下来算 AB_1，将A分行

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \Rightarrow \begin{array}{c} \begin{bmatrix} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{bmatrix} \end{array}\begin{array}{l} A_1 \\ A_2 \end{array}, \text{ 根据 } AB_1 = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} B_1 = \begin{bmatrix} A_1B_1 \\ A_2B_1 \end{bmatrix} \quad \cdots ②$$

其中 $A_1B_1 = A_{11}B_{11} + A_{12}B_{21}$ (~~[illegible]~~)

牢记！

这就是矩阵的 $\boxed{\text{分块相乘}}$

证明这个的 ← 《Matrix computation》 P21.

除了

$$[A_{11} \; A_{12}] \begin{bmatrix} B_{11} \\ B_{21} \end{bmatrix} = A_{11}B_{11} + A_{12}B_{21} \quad ③$$

还有

$$\begin{bmatrix} A_1 \\ \vdots \\ A_q \end{bmatrix} [B_1 \cdots B_r] = C \qquad C_{\alpha\beta} = A_\alpha B_\beta \quad ④$$

①② 这是矩阵相乘的两种看法 [ ] [| |] 一次[ ]

③ 矩阵相乘第三种看法 [| |] 一 [ ]

Elimination = FACTORIZATION : A=LU

重新考虑 A→上三角矩阵 U 的 Elimination 过程是 EA，即一个 Elimination matrix 与 A 相乘。那么如何通过逆操作实现 $L \xrightarrow{how} A$？

$$\left.\begin{aligned} EA &= U \quad 则 \\ E^{-1}U &= A \qquad 即\ A = E^{-1}U \end{aligned}\right\}\ 这过程 可以分解$$

之前将 A 化为上三角矩阵 U，用 Elimination matrix E

则 $A = E^{-1}U$，~~记作 A = LU~~

以上是只需一个 E^{-1}，即只需作一次 Elimination 情况下。一般来说的化简过程需要不只一次 Elimination

$E_n \cdots E_2 E_1 A = U$ 则 $A = (E_1^{-1} E_2^{-1} \cdots E_n^{-1})U$

记作 $A = LU$

（注 E_i^{-1} 也是对角线都是 1 的下三角矩阵
同样 L 也是）

例子

$$A = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix} \longrightarrow U: \begin{bmatrix} 2 & 1 & 0 \\ 0 & \frac{3}{2} & 1 \\ 0 & 0 & \frac{4}{3} \end{bmatrix}$$

$R_2 - \frac{1}{2}R_1$

$R_3 - \frac{2}{3}R_2$

$$L: \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 1 & 0 \\ 0 & \frac{2}{3} & 1 \end{bmatrix} \qquad A = LU$$

L与U的形式（无 Row Exchange 情况下）

1) 当 A 的一行以 0 开头时，L 的这行也是（无需 subtract）
当 A 的一列以 0 开头时，U 的这列也是（不用消元）
（中间的会在 Elimination 过程中出现）

2) L 的对角线肯定都是 1，U 则不然。如何把 U 也变成 1 呢？
这里就要用一种特殊矩阵：对角矩阵

Date · · No.

TRASPOSES 和 PERMUTATIONS

$(A^T)_{ij} = A_{ji}$ 转置

性质简例：下三角 转成上三角（但 inverse 仍是下三角）

$(AB)^T = B^TA^T$ i

$(A^{-1})^T = (A^T)^{-1}$ ii

i) 证明 $(AB)^T = B^TA^T$，首先看 $(Ax)^T = x^TA^T$

我们知道 Ax 是 $\sum_{i=1}^{n} A(:,i)\,x_i$

x^TA^T 是 $\sum_{i=1}^{n} x_i\,A^T(i,:)$] 一样的乘法

得到 $(Ax)^T = x^TA^T$ 之后，将B拆分到拆分很容易证明 $(AB)^T$ 的了

我们发现，矩阵的乘法三种拆分是证明很多东西的利器！

$AB = [Ax_1 \quad Ax_2 \quad \dots]$ 转置后 $\begin{bmatrix} x_1^TA^T \\ x_2^TA^T \\ \vdots \end{bmatrix} = B^TA^T$

将B拆分到拆分　　正好是 B^T 按行拆分后的乘法

ii) $A^{-1}A = I \Rightarrow A^T(A^{-1})^T = I^T = I$

$\therefore (A^T)^{-1} = (A^{-1})^T$ 直观意义：转和逆 同时出现的 先后顺序无关

转置到底意味着什么？ $(Ax)^Ty = x^T(A^Ty)$

谈道在这一本书 那很难用...... 无法...

Symmetric Matrix：$A^T = A$（在转置下定义的）

根据 $A^{-1} = (A^T)^{-1} = (A^{-1})^T$

∴ 对称矩阵的逆也是对称的 ⋮⋮

对任一矩阵 R，R^TR 一定是对称矩阵

R^TR：$(R^TR)^T = R^T(R^T)^T = R^TR$

RR^T 也一样

三、向量空间与子空间

R^n a real vector space is a set of "vectors"① together with rules for vector addition and multiplication by real numbers.

⇓ 二种运算闭包

① 如何理解，向量空间R^n下之向 vector 都是实向量……（进一步探讨这里先停止）

Subspace（子空间）：对线性组合闭包，即若$\vec v$，$\vec w$在子空间中

i. $\vec v+\vec w$ 在子空间中

ii. $c\vec v$ 在子空间中，c为任意scalar

对于一个R^3，它可能有以下四类子空间

(L) Any line through (0,0,0)　　(P) Any plane through (0,0,0)

(Z) single vector (0,0,0)　　(R^3) the whole space

注. part of the line or plane 即不满足条件！

对$Ax=b$的另一种角度观察：

~~若x任意，~~ b可以看成A的列的线性组合

对一个矩阵A，我们将它的列向量所表示的向量空间称作 column space，记作 C(A)

这样，解除之后可以理解成为判断b是不是在C(A)上了。这是另一种区别/看察系数的方式。

若A满秩，任何 则必定可以

若A不满，{ 若b可以被表示（elimination之后一个为0）

　　　　　若b不可以被表示（　　　　都不为0）

$A\vec x=\vec 0$ 的解所构成的空间是 nullspace，记作 N(A)

为什么是一个空间呢？倘若x，y满足 $Ax=0$，$Ay=0$ 则 x，y满足~~线性组合~~子空间的条件 即：

$\vec x+\vec y$ 也是它的解

$c\vec x$ 也是它的解

∴ $A\vec x=\vec 0$ 的解构成一个子空间。

辛 ← 本章教育人以子空间定义来说明!!!

很显然，~~仅有~~ 比如 子空间 $S = x_2\begin{bmatrix}-2\\1\\0\end{bmatrix} + x_3\begin{bmatrix}3\\0\\1\end{bmatrix}$ 为例

它不能表示所有的三维向量，如 向量 $\begin{bmatrix}0\\1\\1\end{bmatrix}$

满足 2,3 component 之后，x_2, x_3 也就定了，不能再满足 first component 了。

但已知它能满足一个子空间，问：任意 2 个 3维向量 $\begin{bmatrix}a\\1\\0\end{bmatrix}$ $\begin{bmatrix}b\\0\\1\end{bmatrix}$ 是否都能形成一个子空间？（的线性组合）（三维）

答：可以，这 2 个向量可以构成一个 $\vec{A}x = \vec{0}$ 的方程，

即 $x - ay - bz = 0$，$\vec{A}$ 为 $[1, -a, -b]$

这样 $\begin{bmatrix}a\\1\\0\end{bmatrix}$ $\begin{bmatrix}b\\0\\1\end{bmatrix}$ 的线性组合构成的子空间恰好是

$\vec{A}x = \vec{0}$ 的 nullspace。（A 与 nullspace matrix（指特解构成的 matrix）有如下关系 $A = \begin{bmatrix}I & F\\0 & 0\end{bmatrix}$，$N = \begin{bmatrix}-F\\I\end{bmatrix}$）

而且来说，这个问题可以表示为如下

~~当~~ 当 $m \le n$ 时，m 个 $\mathbb{R}^n$ 向量的秩为 k ($k \le m$)，则 这 m 个向量 ~~可以构成~~ 的线性组合可以构成一个 k 维的子空间

$$\begin{bmatrix} I_k & 0 \\ \begin{smallmatrix}\times & \times & \times\\ \times & \times & \times \\ \times & \times & \times\end{smallmatrix} & 0 \end{bmatrix}$$

→ 行向 $\vec{A}_{m \times n}^{T}$... n, m

Rank 的一个小性质

两个向量的外积形式，~~此~~ uv^T，如

$\begin{bmatrix}1\\2\\3\end{bmatrix}[1, 3, 10] = \begin{bmatrix}1 & 3 & 10\\2 & 6 & 20\\3 & 9 & 30\end{bmatrix}$ 相乘的秩肯定为 1. …①

~~矩阵秩~~

重要性质：矩阵行秩=列秩！

① 矩阵有两种证明，都能得到行、列都是倍数关系

Date . . No.

★ 对于上述情况，此m个R^n向量的column space由k个相互独立的vector构成，是R^n一个dimension为k的子空间，这k个相互独立的vector [| | ... |] 构成的sequence(通常矩阵表示)叫作 ~~set~~ basis 空间的基，

每一个vector以唯一base

四个空间与$A_{m\times n}$相关的空间

1. Row space $C(A^T)$，R^n的子空间
2. Column space $C(A)$，R^m的子空间
3. null space $N(A)$，R^n的子空间
3. left nullspace $N(A^T)$，R^m的子空间 $A^Ty=0$ 即 $y^TA=0$

m [ ] n个R^m的向量

2个beautiful line (都 = r)

a. Row space 和 Column space 有相同的 dimension
b. nullspace dimension 为 n−r，即 Rank(1+3) = n
left null space dimension 为 m−r，即 Rm(2+4) = m

A的column space维数与R的column space维数相等

∵ $C(A)$ dim r = $C(A^T)$ dim r，$C(R)$ dim r = $C(R^T)$ dim r

又∵ $C(A^T)$ dim r = $C(R^T)$ dim r 行行等变换

∴ $C(A)$ dim r = $C(R)$ dim r

注意：只有维数相等，basis不同的！

$A\begin{bmatrix}2&3\\4&6\end{bmatrix}$ $R\begin{bmatrix}2&3\\0&0\end{bmatrix}$ (当然Row space都相同)

得到A与R的4个space的维数都相等。

a b 证明：a直出，b以R来证明，就证明了A。间接

于是就有了A的4个space之间良好的关系

~~正交~~ orthogonality

四. 正交 (ORTHOGONALITY)

空间level上的正交：$\vec{v}\cdot\vec{w}=0$ or $\vec{v}^T w=0$, for all $\vec{v}$ in V and all $\vec{w}$ in W，则 V 与 W 正交

1. A 的 Row space 与 N 正交. $Ax=0$，故 N 中的任一 vector 都与 A 的任一横向量点积为 0。故 $C(A^T)$ 与 $N(A)$ 正交

2. 不仅如此！它们不仅 Orthogonality 而且 orthogonal complement！即正交互补空间。因为维数相加刚好是 n：$(r+(n-r))$ (注. 互补的意义在于补充2)

正交互补表示：$V^\perp$ 表示 V 的正交互补空间

故 $N(A)^\perp = C(A^T)$

$C(A^T)^\perp = N(A)$

补充1：空间正交，则两间唯一相交的 vector 必定是 $\vec{0}$。
证：若 $\vec{a}$ 相交于两，则 $\vec{a}\cdot\vec{a}$ 为两空间横向量点积为 0，故 $\vec{a}=\vec{0}$.

∴ 墙和地面不是正交空间！

补充2：$A\vec{x}=0$，则 $\vec{x}$ 与 A 的 Row space 正交的横向量，即与每一个 Row 的点积都为 0，则它在 $N(A)$ 中，因为满足 $A\vec{a}=0$，则 $\vec{x}$ 必是 $N(A)$ 中的某一个向量

同理 同样与 $N(A)$ 正交的向量必在 $C(A^T)$ 中。可证，若存在不是这样的向量 $\vec{x}$，则我们可以将其作为一个 row 加入 A 中，由于 $\vec{x}$ 不在 $C(A^T)$ 中，则 row space 的 dim 会变成 $r+1 \neq r$

必定能加入其中

Fundamental Theorem of Linear Algebra (part 2)

$N(A)^\perp = C(A^T)$ Nullspace 和 Row space 是 R^n 中的正交互补空间

$N(A^T)^\perp = C(A)$ Left Nullspace 和 Column space 是 R^m 中的正交互补空间

Date · · No.

定理: n个线性无关的向量能作为 R^n 的基 (basis)

$C(A^T)$ 中的 r 个基 vector，加上 $N(A)$ 中的 (n-r) 个 vector，若这 n 个都是线性无关的，故可作为 R^n 的基

$\therefore \begin{bmatrix} I & F \\ -F & I \end{bmatrix}_{n\times n}$ 是个满秩矩阵 → 记牢，当然可以分解。

→ 已证明，但是意义何在？

R^n 中的任意向量 $\vec{x} = \vec{x}_r + \vec{x}_n$

$A\vec{x}_n = 0$

① 所有行向量被A乘，$A\vec{x}_r = \vec{b}$ ⇒ 都在 $C(A)$ 中

② 所有列空间向量都来自 $A\vec{x}_r$

③ 又因为由来是唯一的 ⇒ 都来自 $C(A^T)$

$C(A^T)$ 主空间 $C(A)$ 值空间 且一一对应!

证明：

你们现在了这个P，其秩为1。我们猜 $(I-P)$ 也是一个 projection matrix，它所作用应该是把 b 投影到 a 的正交互补空间，其投影为 e

$(I-P)b = b - Pb = e$

那么 $I-P$ 怎么证明呢，证明它能将任意向量投影到 a 的正交互补空间里呢？

已知 对任意向量 b，都能投影为 e

e 在 a 的正交互补空间里！

补充说明： $P^2=P$ 因为投影之后，第二次无意义

$$\frac{aa^T}{\boxed{a^Ta}} \cdot \frac{aa^T}{a^Ta} = \frac{aa^T}{a^Ta}.$$

2) 投影到 $C(A)$

problem: given $a_1, a_2, \dots, a_n$ in R^m, assume these a's are linearly independent. Find the combination $\hat{x}_1 a_1 + \hat{x}_2 a_2 + \dots + \hat{x}_n a_n$ that is closest to a given b.

$n=1$ 即为投影到线上面问题，我们再找 $n>1$ 时的 $\hat{x}$

$\hat{x} = [\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n]^T$

同理，这样的 $A\hat{x}$ 只有一条，也可到方程求解它

$$A^T(b - A\hat{x}) = 0 \quad ①$$ 即每一个 a_i 都与 $b-A\hat{x}$ 点积为 0

$\therefore \quad A^TA\hat{x} = A^Tb$ 若 A^TA 可逆则 $\hat{x} = (A^TA)^{-1}A^Tb$

$$p = A(A^TA)^{-1}A^Tb$$

$$P = A(A^TA)^{-1}A^T$$

与一维时的区别

① 与 column 空间正交的，肯定在 A 的 left null space 中，即 $N(A^T)$，由点积就是 $A^T(b-A\hat{x})=0$

LEAST SQUARES APPROXIMATIONS 最小二乘估计

strang书中将 $b-A\hat{x}$ 称作"error"是有原因的。在 ~~第~~ ~~给~~ 给定A，b 求b在C(A)上的投影p，投影矩阵P的时候，A通常是 $m\times n$ $(m>n)$ 的情况，即A长成下面这个样子

这样的 $Ax=b$，~~通~~ 常常是无解的，即b在C(A)之外（不能用A的列向量线性表示）

但是，我们仍想找一个x使得Ax与b最接近，~~怎~~ 怎么找？找它的投影~~是~~就好…… 书上说……

书上将 $b-A\hat{x}$ 看作error，这个最小 则

$A^T(b-A\hat{x})=0$ 即 ~~$A^TA\hat{x}=A^Tb$~~

即 $A^TA\hat{x}=A^Tb$ （可以看作 $A\hat{x}=b$ 两边同用 A^T 乘）

这个解叫做 least square solution.

这个解有什么用呢？除了求投影

Application: Find a straight line to ~~the~~ m points.

eg find the closet line to the points (0,6), (1,0), (2,0)

解：假设这条线为 $b=C+Dt$ 则方程组

$$C+D\cdot 0=6 \qquad C+D\cdot 1=0 \qquad C+D\cdot 2=0$$

$$\begin{bmatrix}1&0\\1&1\\1&2\end{bmatrix}\begin{bmatrix}C\\D\end{bmatrix}=\begin{bmatrix}6\\0\\0\end{bmatrix}$$

~~根据代数方程组理论~~，b显然不在A的列空间内。

$Ax=b$ 不可解！①

用投影的方法求 $\hat{x}=(A^TA)^{-1}A^Tb$

是不是很神奇！当m=100的时候就是找到一条最好的直线了！

① 可以将列向量，b排成一排，用行 elimination，可以发现秩为3

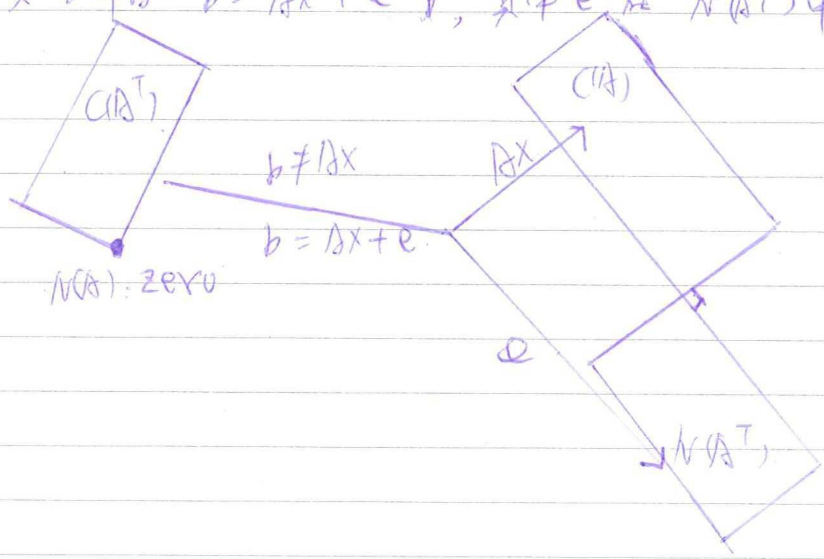
② ~~可以~~ 说明b不在列空间内；也可以简单地看 ~~[A:b]~~ [A:b] 不能消到 ~~秩~~ 秩为2

a. 与 Fundamental Theorem of Linear Algebra 的对象 A 不同，最小二乘（背后是 projection 方法）的对象是 A 的 nullspace 实际上只有一个点 zero!

依据：A 的 column vector 之间是 independent，这样 $(A^TA)^{-1}$ 才有意义

b. 在 Fundamental Theorem of Linear Algebra 中，我们把重心放在 Ax 上，而非 b 上，研究的都是有解的情况。如，任一 $x=x_r+x_n$ ~~从而每一个能够以 A 的 column vector 线性表示的 b~~，都是来自于 A 的行空间 $C(A^T)$ 的一个向量。

而在这里，我们研究的是不在 C(A) 中的向量 b，我们将其分解为 $b = Ax + e$，其中 e 在 $N(A^T)$ 中



扩展：从一维函数到多维 比如可以求匹配曲线 $y = C + Dt_1 + Et_2$

这样 A 的 independent column vector 只是从 2 个变成了 3 个。

格拉姆-施密特正交化

ORTHOGONAL BASES AND GRAM-SCHMIDT

1) 正交能简化运算，除了前面描述线性代数的本质。
2) 如何构造正交向量

确实，我们已经能解决非正交基上的一些问题(如投影)

(保持这原理可花了不少时间，而且我们见到的各个 space 一般都是非正交基)

但是，正交基能带来很多好的运算性质！

定义：1) orthonormal，即 given $q_1, \dots, q_n$ if

$$q_i^T q_j = \begin{cases} 0 & \text{when } i \neq j \ (\text{orthogonal vectors}) \\ 1 & \text{when } i = j \ (\text{unit vectors}: \|q_i\| = 1) \end{cases}$$

2) 由 orthonormal ~~vector~~ columns 组成的矩阵叫做 Q 来表示

$$Q = \begin{bmatrix} & \\ & \end{bmatrix}_{m \times n} \quad m \geq n$$

性质 $Q^T Q = I$ ① $\Leftrightarrow$ orthonormal

$$Q^T Q = \begin{bmatrix} - & q_1^T & - \\ - & q_2^T & - \\ & \vdots & \\ - & q_n^T & - \end{bmatrix}_{n \times m} \begin{bmatrix} | & | & & | \\ q_1 & q_2 & \cdots & q_n \\ | & | & & | \end{bmatrix}_{m \times n} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = I$$

~~注意~~

根据在 Projection 一节中证明的定理，Q^TQ ~~为对角阵~~ 的秩为 n ①

!!! 突然是显而易见的么...

因它不仅仅是对称矩阵，它还是个 I ... ② —— $I_{n \times n}$

不过幸说，性质与前面证明的不冲突！

另，由于 $Q^TQ = I$

$\therefore$ 当 Q 为方阵 $n \times n$ 时，$Q^T = Q^{-1}$：transpose = inverse

列向量 orthonormal 的 matrix 称为 orthogonal matrix

Date · · No.

格拉姆—斯密特正交化

给三个线性无关的向量 a, b, c 将其化为 orthonormal 关系。

1. 化为 ~~orthonogal~~ orthogonal　　2. 每个 vector/‖vector‖

$A = a$　　$B = b - A\frac{A^Tb}{A^TA}$ (error)

$C = c - [A\ B]\frac{[A\ B]^Tc}{[A\ B]^T[A\ B]}$

$= c - A\frac{A^Tc}{A^TA} - B\frac{B^Tc}{B^TB}$

D (if exists) ......

$q_1 = \frac{A}{\|A\|}$

$q_2 = \frac{B}{\|B\|}$

$q_3 = \frac{C}{\|C\|}$

$\boxed{\text{QR分解}}$

一个矩阵A，经过 Gram-Schmidt 正交化之后，从 [a b c] 变到 $Q = [q_1\ q_2\ q_3]$。从A到Q的过程，由于 ~~[illegible]~~ a, b, c 能被 q_i 表示，肯定有第三个 matrix R，使得 A 与 Q 联系起来。这就是 Algebra 中第二个矩阵分解算法：QR分解！

在之前计算 b在 orthonormal 的向量组成的空间上投影时，我们算得

$$\hat{x} = Q^Tb = \begin{bmatrix} - q_1^T - \\ - q_2^T - \\ \vdots \\ - q_n^T - \end{bmatrix} b = \begin{bmatrix} q_1^Tb \\ q_2^Tb \\ \vdots \\ q_n^Tb \end{bmatrix}$$

$$p = Q\hat{x} = \begin{bmatrix} | & | & & | \\ q_1 & q_2 & \cdots & q_n \\ | & | & & | \end{bmatrix}\begin{bmatrix} q_1^Tb \\ q_2^Tb \\ \vdots \\ q_n^Tb \end{bmatrix} = q_1(q_1^Tb) + q_2(q_2^Tb) + \cdots + q_n(q_n^Tb)$$

~~[illegible]~~ 当Q为方阵时，即 ~~[illegible]~~ b的投影是其本身时

(~~[illegible]~~，~~[illegible] b在Q空间内，b = p = Q\hat{x}~~)

$b = q_1(q_1^Tb) + q_2(q_2^Tb) + \cdots + q_n(q_n^Tb)$

这是向量分解的基础！

五、DETERMINANTS 行列式

行列式能拿来干吗？

1. 与 inverse 的关系

i) 与 inverse 一样，仅是方阵才有 $|A|$

ii) $|A| \neq 0$，A 才有 A^{-1}

iii) A^{-1} 的 $\det A^{-1} = 1/\det A$，联系 ii) $\det A^{-1}$ 无意义，若 $\det A = 0$

这样当 $\det A \neq 0$ 时 $Ax=b$ 就能直接求出为 $x = A^{-1}b$

2.

3. 使得 $A-\lambda I$ 为 singular，$\det(A-\lambda I)=0$ 的 λ 是 A 的特征值！

定义：The product of the pivots is the determinants.

eg) $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ $\det A = a(d - \frac{c}{a}b) = ad - bc$

关于 $n>2$ 的 det，我们不打算给出它的公式（见后有时间说明，太复杂了）。我们通过给出 det 的性质来求 det 的值。

性质：

1. $\det I = 1$ 符合定义

2. 斜对称性（reversal）：A 行互换 A'，$\det A' = -\det A$

（这样可以求任意 permutation matrix 了，只看数数换了几次行）

3. det 是 A 的行向量的多线性函数

$$\det \begin{vmatrix} a_1 \\ \vdots \\ \lambda a_i + u a_i^* \\ \vdots \\ a_m \end{vmatrix} = \lambda \det \begin{vmatrix} a_1 \\ \vdots \\ a_i \\ \vdots \\ a_m \end{vmatrix} + u \det \begin{vmatrix} a_1 \\ \vdots \\ a_i^* \\ \vdots \\ a_m \end{vmatrix}$$

对每一行都是线性性，（超过 2 个向量称为多线性性了）

故 det 是行的斜对称、多线性性函数

Date · · No.

计算 determinant. PERMUTATIONS & COFACTORS

引入一个矩阵

$$\begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{bmatrix}$$

方法1. $\det P \det A = \det L \det U$ $(\det L = 1)$

很容易得到 $\det A = 5$ (实际上 A若为 $n\times n$, $\det A = n+1$)

注: 关于 permutation, 根据 det 的性质, 它就是 I 经过行互换后得到的, 换奇数次为-1, 偶数次为1.

方法2. 以 det 的前三个性质为依据, 以 2×2 matrix 为例, 得到行列式的公式:

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = \begin{vmatrix} a & 0 \\ c & d \end{vmatrix} + \begin{vmatrix} 0 & b \\ c & d \end{vmatrix}$$ (行的线性性)

$$= \underline{\begin{vmatrix} a & 0 \\ c & 0 \end{vmatrix}} + \begin{vmatrix} a & 0 \\ 0 & d \end{vmatrix} + \begin{vmatrix} 0 & b \\ c & 0 \end{vmatrix} + \underline{\begin{vmatrix} 0 & b \\ 0 & d \end{vmatrix}}$$

同一列有之间是线性相关的, 则 $\det = 0$

$$= \begin{vmatrix} a & 0 \\ 0 & d \end{vmatrix} + \begin{vmatrix} 0 & b \\ c & 0 \end{vmatrix} = \cancel{\det}\ ad - bc$$

推广至 $n\times n$ 矩阵, 可以发现 det 值由以下规则构成:

1) 每行都出现一个元素构成一个乘积, 这些元素两两之间不能在同一列, 故共有 $n!$ 个乘积

2) 一个乘积的正负由它对应的 permutation 矩阵决定, 如

$$\begin{bmatrix} 0 & b \\ c & 0 \end{bmatrix} = \underline{\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}} \begin{bmatrix} b & 0 \\ 0 & c \end{bmatrix},$$

$\therefore \det A = \sum (\det P)\, a_{1\alpha} a_{2\beta} \cdots a_{n\omega}$ = BIG FORMULA.

Date . . No.

六、EIGENVALUES AND EIGENVECTORS 特征值/向量

除了线性方程 $Ax=b$（被称作稳定状态问题 steady state problem），还有一类方程 $du/dt=Au$ 而u是随时间变化的向量，称作 动态问题，dynamic problem

⇒ 需要 eigenvalue 来求解

另外，eigenvalue, eigenvector 在求 matrix power，A^k 上很有用

特征向量是什么？

绝大多数向量被A乘后（$A\vec{x}$）会改变方向，例外就是A的特征向量，即 $A\vec{x}=\lambda\vec{x}$

λ就是特征值，

表明这个向量是变长还是变短（$|\lambda|>1$？），有没有变方向（λ负？）

很自然而我们就有2个问题：

1. 给定一个A，它的特征值有几个，特征向量有几个，如何求？
2. A的特征向量形成的子空间维数是多少？

究竟如何解决这两个问题是理解 eigenvector 与 eigenvalue 的关键

定理1. λ是A的特征值 ⇔ $\det(A-\lambda I)=0$ （characteristic polynomial 特征多项式）

⇒ ∵λ是A的特征值 则 $(A-\lambda I)\vec{x}=0$ 有非零解，故 $\det(A-\lambda I)=0$

（否则若只有零解，则 $\vec{x}=0$，不符合定义[1]）

⇐ $\det(A-\lambda I)=0$，则 $(A-\lambda I)\vec{x}=0$ 有非零解，故λ为A的特征值

注[1]：特征向量不能是0，用0作为特征向量不具备任何意义，而特征值可以是0 ⇔ $\det A=0$

定理1 解决了 给定一个A有几个λ的问题。答案是有n个解，因为 $\det(A-\lambda I)=0$ ~~得到~~，根据det的公式，λ最高次为n

若能表示则是很棒的事情，很好求解了：

$\vec{x} = a\vec{x_1} + b\vec{x_2}$，$\vec{x_1},\vec{x_2}$ 是 $A_{2\times2}$ 的特征向量，线性无关

则 $A\vec{x} = A(a\vec{x_1}+b\vec{x_2}) = a\lambda_1\vec{x_1} + b\lambda_2\vec{x_2}$

∴ 若A的 eigenvectors 可作为 $\mathbb{R}^n$ 的基，即它们之间线性无关，会是很好的性质！我们把这等价于对角化的充分必要条件。

在继续讨论这个之前，先讲清一个点：初等变换改变λ. e.g.

$A=\begin{bmatrix}1&1\\1&1\end{bmatrix}$ $\quad \begin{vmatrix}1-\lambda&1\\1&1-\lambda\end{vmatrix} = \lambda(\lambda-2)$

而 $U=\begin{bmatrix}1&1\\0&0\end{bmatrix}$ $\quad \begin{vmatrix}1-\lambda&1\\0&-\lambda\end{vmatrix} = \lambda^2-\lambda-1$ ← λ不同啊.

…… ①

∴ λ阵的情况与 Rank(A) 无关？

$$
\begin{array}{l}
A \;:\; \Lambda \text{（λ在对角线上形成的对角矩阵）}\\
\;\downarrow \boxed{E,P}\\
U \quad \Lambda' \neq \Lambda
\end{array}
$$

唯一有关联的是，当 det A = 0 时，即A可逆时，0是A的特征值！

∵若我们想通过研究A，得到λ的特征，再得到 $\vec{x}$ 的特征，能否撑起整个 $\mathbb{R}^n$ 空间，是行不通的……②

$A \not\longrightarrow$ λ集 $\longrightarrow$ $\vec{x}$集

我们只能反推，若A有n个不同的λ，则 eigenvectors 能撑起 $\mathbb{R}^n$！

①②待深入研究，反正只是猜想。

— 比较 A 与 Λ，很显然，它们有相同的特征值，那么 eigenvectors 呢？

来求下 λ_i 对应的 eigenvectors：

$(\Lambda - \lambda_i I)\vec{x_i} = 0$

分析 $\begin{bmatrix} \lambda_1-\lambda_i & & & & \\ & \ddots & & & \\ & & \lambda_i-\lambda_i & & \\ & & & \ddots & \\ & & & & \lambda_n-\lambda_i \end{bmatrix} \vec{x_i} = 0 \Rightarrow \vec{x_i} = k \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix}$ …… 第 i 行

∴ Λ 的特征向量为 I 的列向量，与 A 不同。

	eigenvectors matrix	eigenvalue
Λ	I	Λ
A	S	Λ

— S 的列向量可以随意 scalar，仍旧满足 $S^{-1}AS = \Lambda$

$[x_1, x_2 \cdots x_n]^{-1} A [x_1\ x_2 \cdots x_n] = \Lambda$

$[ax_1, bx_2, \cdots, cx_n]^{-1} A [ax_1\ bx_2 \cdots cx_n] = \Lambda$

↳ 仍是 A 的特征向量 $\in N(A-\lambda I)$

极端情况就是 $A = I$ 时：所有向量都是 A 的特征向量…

— 除了 S，还有没有其它矩阵能对角化 A？

假如存在这样的矩阵 S'，$AS' = S'\Lambda'$ 则取 S' 的列 s_i'

$As_i' = \lambda_i s_i'$ 显然定出 s_i' 与 λ_i 只能是 A 的 eigenvectors 与 values.

故 $S' = S$

只有 S 能对角化 A

用matrix power来计算 $u_{k+1}=Au_k$，即 $u_k=A^k u_0$

关键想法：u_0用A的eigenvectors表示。

步骤如下：

1. 找到A的eigenvalues和eigenvectors（需要independent）
2. u_0表示成S的线性组合，$u_0=c_1\vec{x}_1+\cdots+c_n\vec{x}_n$
3. $u_k=A^k u_0=c_1\lambda_1^k\vec{x}_1+\cdots+c_n\lambda_n^k\vec{x}_n$

用矩阵的语言描述如下

2. $u_0=\begin{bmatrix} | & | & & | \\ x_1 & x_2 & \cdots & x_n \\ | & | & & | \end{bmatrix}\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}$，即 $u_0=S\vec{c}$

3. $u_k=A^k u_0=S\Lambda^k S^{-1}S\vec{c}=S\Lambda^k\vec{c}$

$S\Lambda^k c=c_1\lambda_1^k\vec{x}_1+c_2\lambda_2^k\vec{x}_2+\cdots+c_n\lambda_n^k\vec{x}_n$

归纳：即将u_k表示成A^k的eigenvectors线性组合

一个优点是，不管矩阵有无重复特征值，都能判断矩阵是否可对角化的条件：

1. 我们已证明 ~~有~~ 不同λ之间的eigenvectors必定线性无关，因而有，要是矩阵有n个不同λ，则S就表示整个$\mathbb{R}^n$，即有n个eigenvectors independent

2. 同一个λ若有≥2个eigenvector，则由于这些向量$\in N(A-\lambda I)$，它们之间线性无关

∴综上，一个矩阵可对角化的充分必要条件是：每个特征值对应的特征向量线性无关的最大个数等于该特征值的重数。（见下一页）

也等价于有n个线性无关向量特征。

Date · · No.

symmetric matrix 对称矩阵

symmetric matrix is the most important matrix in the world

— Strang

Projection matrix:

1) 若投影到 line 上 $P=\frac{aa^T}{a^Ta}$

$a=(a_1 \cdots a_n)^T$ 则 ~~令~~ 若 $A=aa^T=\begin{bmatrix}a_1\\ \vdots \\ a_n\end{bmatrix}[a_1 \cdots a_n]$

则 $A_{ij}=a_ia_j$

$A_{ji}=a_ja_i=A_{ij}$

故 $P=P^T$，P 为对称矩阵

且有 eigenvalue $\begin{cases}0 & Px=0 \quad (\text{当 } x \text{ 与 } a \text{ 垂直}) \\ 1 & Px=x \quad (\text{当 } x \text{ 与 } a \text{ 重合})\end{cases}$

2) 若投影到一个子空间，则

$P=\frac{AA^T}{A^TA}=A(A^TA)^{-1}A^T$

$P^T=A((A^TA)^{-1})^TA^T=A((A^TA)^T)^{-1}A^T=A(A^TA)^{-1}A^T=P$

∴ P 为对称矩阵

对称矩阵的美好性质：

Spectral theorem: Every symmetric matrix has the factorization $A=Q\Lambda Q^T$, with real eigenvalues in Λ and orthonormal eigenvectors in Q:

$A=Q\Lambda Q^{-1}=Q\Lambda Q^T$, with $Q^{-1}=Q^T$ 即 $Q^TQ=I$

证明分两步：

1. 有 n 个不同 eigenvalues 时，不同 eigenvector independent (上一节已证) 且 orthogonal (这一节证)

2. 即使 eigenvalue 有重根，仍旧有 n 个 orthonormal eigenvectors

(不同 λ 的 eigenvectors 之间正交，同一 λ 的 eigenvectors 个数=重数，且正交?!

Date · · No.

当从 一般矩阵(可对角化条件满足) ⟶ 对称矩阵

$A = S\Lambda S^{-1}$ $\qquad\qquad$ $A = Q\Lambda Q^T$

有什么优势呢，作为对称矩阵？

$$A = Q\Lambda Q^T = \begin{bmatrix} | & | & & | \\ x_1 & x_2 & \cdots & x_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \begin{bmatrix} - x_1^T - \\ - x_2^T - \\ \vdots \\ - x_n^T - \end{bmatrix}$$

$$= \lambda_1 x_1 x_1^T + \lambda_2 x_2 x_2^T + \cdots + \lambda_n x_n x_n^T$$ 此即为谱分解！

其中每一个 $x_i x_i^T$ 都可作为一个 projection matrix (投影到 x_i 这条线上的 P)。

故 $A = \lambda_1 P_1 + \lambda_2 P_2 + \cdots + \lambda_n P_n$

矩阵本质 分解成 n 个秩为1的 矩阵

在 SVD 分解中也有重要应用

对称矩阵属于同一λ的 eigenvectors 之间的 independent 向量是可以 gram-schmidt 正交化的，正交化之后还是这个λ的特征向量(不同λ之间正交化后就不再属于任一λ的特征向量了)

故 symmetric matrix 可对角化的证明化为其任一λ：几何重数 = 代数重数

当 LDL^T 与 I 靠近时，即 $3\to 2\to 1\to 0$ 时，LDL^T 的特征值从一开始的 4, -2 变到 1, -8（即 pivots）。

	$A=LDL^T$	$A=IDI^T$
λ	4, -2	1 -8
Pivots	1, -8	1 -8

这个过程中，eigenvalue 的值变了，但符号不变（无特征值变 0，可以通过其 $\lambda=0$，$\det A=0$，可以从 LDL^T 上看出是不可能的）

即对称矩阵 主对角线外的元素 $\longrightarrow 0$ 的过程中，符号不变，也说明了特征值与 pivots 的符号相同

证毕：

∴ 正定阵的 pivots 都是正的，即 $A=LDL^T$，D的元素都为正

~~正定矩阵：~~ $x^TAx>0$ ~~x是A的~~ eigenvectors
for all nonzero vector x

证：$Ax=\lambda x$，~~若A是~~

$x^TAx=\lambda x^Tx>0$ 由于 $x^Tx>0$ 故 $\lambda>0$

~~证毕~~

$x^TAx=(x\ y)\begin{bmatrix}a&b\\b&c\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}=ax^2+2bxy+cy^2>0$

正定矩阵的逆、根

定理4. 1) 根据定理1，正定阵 $\det A>0$，必定 nonsingular

2) 其逆阵 $A^{-1}=(Q\Lambda Q^T)^{-1}=Q\Lambda^{-1}Q^T$ （$Q^TQ=I$）

若 $\Lambda=\mathrm{diag}(\lambda_1,\dots,\lambda_n)$ 则 $\Lambda^{-1}=\mathrm{diag}(\frac{1}{\lambda_1},\dots,\frac{1}{\lambda_n})$

故 A^{-1} 也为正定阵

定义.

$$A^{\frac12}=\sum_{i=1}^{n}\sqrt{\lambda_i}\,e_ie_i^T=Q\Lambda^{\frac12}Q^T$$，也是正定阵

叫根号矩阵，$A^{\frac12}A^{\frac12}=A$

性质：$(A^{\frac12})^{-1}=Q\Lambda^{-\frac12}Q^T$，$\Lambda^{-\frac12}=\mathrm{diag}(\frac{1}{\sqrt{\lambda_1}},\dots,\frac{1}{\sqrt{\lambda_n}})$

写成 $A^{-\frac12}$

$A^{\frac12}A^{-\frac12}=A^{-\frac12}A^{\frac12}=I$

$A^{-\frac12}A^{-\frac12}=A^{-1}$

Date · · No.

① $R(\text{null})$ 不变的理由，$B=M^{-1}AM$ ~~而有~~ M^{-1} 与 M 都是可逆的

引理：任意矩阵与可逆阵相乘不改变秩

即：若 P, Q 可逆，则 $r(A)=r(PA)=r(AQ)$

证明：可逆阵可表示成初等矩阵之乘积

$[A \ I] \xrightarrow[\text{变换}]{\text{初等}} [I \ A^{-1}]$，即

$(E_k \cdots E_2 E_1) PA = I$，即可逆阵 $\xrightarrow[\text{变换}]{\text{初等}} I$

$A = P^{-1} L I$，证毕

A与可逆阵之乘积即为A与一系列 ⟨ permutation matrix / Elimination matrix

而初等变换不改变矩阵之秩（不证）

阵之乘积，不改变A之秩

② $\vec{x}' = M^{-1}\vec{x}$（数量相当，下面证明 independent）

若 $\vec{x}$ 满足 $c_1\vec{x}_1 + c_2\vec{x}_2 = \vec{0}$ 只有 $c_1=c_2=0$ 这一个解

则 $c_1\vec{x}_1' + c_2\vec{x}_2' = c_1 M^{-1}\vec{x}_1 + c_2 M^{-1}\vec{x}_2$

$= M^{-1}(c_1\vec{x}_1 + c_2\vec{x}_2) = \vec{0}$

由于 M^{-1} 可逆满秩，故 $c_1\vec{x}_1 + c_2\vec{x}_2 = \vec{0}$，故 $c_1=c_2=0$

$\therefore M^{-1}\vec{x}_1$ 与 $M^{-1}\vec{x}_2$ independent

③ $N(A)$ 似与 $\lambda=0$ 有关。若A有 $\lambda=0$ ~~因~~ 则 $N(A)$ 由 $\lambda=0$ 时的 eigenvector 组成

即 $\det A=0$ 则A有特征值 $\lambda=0$，其特征向量组成 $N(A)$

当然B也有 $\lambda=0$，特征向量 $M^{-1}\vec{x}$ 组成 $N(B)$

所以问题归结为 $\vec{x}$ 形成的空间与 $M^{-1}\vec{x}$ 形成的空间是一样的吗？

当然不一样（举个例子就知道）

但 $r(N(A))$ 与 $r(N(B))$ 是一样的！都等于 $n-r$

Date · · No.

Singular Value Decomposition (奇异值分解)

首先定义:

The singular value decomposition (SVD) has orthogonal matrices U and V

$$AV = U\Sigma \text{, and then } A = U\Sigma V^{-1} = U\Sigma V^T$$

This is the new factorization of A: orthogonal times diagonal time orthogonal

若A为 $m\times n$ 矩阵, 则 V 为 $n\times n$, U为 $m\times m$

证明: A^TA 为 $n\times n$ 的对称阵, 它(可)分作如下分解

$$A^TA = Q\Lambda Q^{-1} \text{, Q 为 orthonormal matrix}$$

1) 假设其 eigenvectors 为 v_i, 则

$$A^TAv_i = \lambda_i v_i \quad , \quad v_i^TA^TAv_i = \lambda_i v_i^Tv_i = \lambda_i$$

$$\Rightarrow \|Av_i\| = \sigma_i \quad (若 \sigma^2 = \lambda_i) \quad \cdots ①$$

2) $AA^TAv_i = \lambda_i Av_i$, 令 $u_i = \frac{Av_i}{\sqrt{\lambda_i}}$, ~~则~~ $\sqrt{\lambda_i} = \sigma_i$

$$AA^Tu_i = \sigma_i^2 u_i \text{, ~~即 } u_i \text{ 是~~} \quad \cdots ②$$

由②得 u_i 是 AA^T 的特征向量, ~~特征值是~~ 其长度为 1

即:

- A^TA 的 orthonormal 特征向量为 v_i, 特征值为 σ_i^2
- AA^T 的 orthonormal 特征向量为 u_i, 特征值为 σ_i^2
- 其中 $u_i = Av_i/\sigma_i$

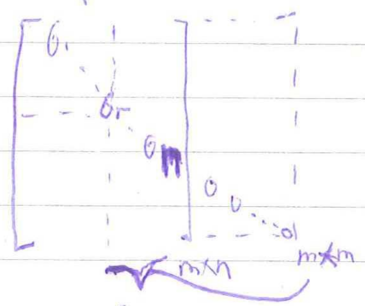
故 $Av_i = \sigma_i u_i$[①] v_i有 n个, u_i有 m个 (分别是 A^TA, AA^T 的 eigenvectors)

$$\therefore \underset{m\times n}{A}\ \underset{n\times n}{V} = \underset{m\times m}{U}\ \underset{m\times n}{\Sigma}$$

(以上证明包括了 $\sigma_i = 0$ 的情况)

① 类似 λ, σ 称 singular value 形式

Σ的形式如下



$\sigma_{r+1}\cdots\sigma_m$ 都是0

A^TA 求得的σ是实线部分
AA^T 求得的σ是虚线部分
它们的σ(非重复)的数量
都为r！

最后检查 A^TA，AA^T

$A^TAv_i=\sigma_i^2v_i$，包含两部分 $\begin{cases} A^TAv_i=\sigma_i^2v_i & (\sigma_i\neq 0) \quad ① \\ A^TAv_i=0 & (\sigma_i=0) \quad ② \end{cases}$

对于第②部分，事实上我们之前已经证明过：

$N(A)=N(A^TA)$

A^TA 有 n 个 orthonormal 的特征向量，但若 A 的秩仅为 r (r<n)，即 det A=0 时，A 有特征值 0 对应有 n−r 个特征向量 ∈ N(A)

→ 与此一致！满足 $A^TAv_i=0$

类似地，u_i 若是 ∈ $N(A^T)$ 则必 ∈ $N(AA^T)$

若 A 有秩为 r，则 A^T 也为 r，则 A^T 有特征值 0，对应的 m−r 个特征向量 ∈ $N(A^T)$，也属于 $N(AA^T)$ 故 $AA^Tu_i=0$

以上证明左页求法是对的

综上，我们给出 $AV=U\Sigma$ 的另一种非0形式表示：

$$A=\sum_{i=1}^{r(A)}\sigma_i u_i v_i^T$$

补充一些 SVD application

1. Euclidean norm of a matrix

$$\|A\|_2 = \max_{x\neq 0} \frac{\|Ax\|_2}{\|x\|_2} = \sigma_{max}$$ 又称谱范数

2. Euclidean condition number

$$\mathrm{cond}(A) = \sigma_{max}/\sigma_{min}$$

3. 求解线性方程或最小二乘问题

$$\left(\sum \sigma_i u_i v_i^T\right)x = b$$

$$x = \sum_{\sigma_i \neq 0} \frac{u_i^T b}{\sigma_i} v_i$$

特别适用于 rank deficient 问题 / ill-condition 问题. 因为小 σ 可以抛掉

making it less sensitive to perturbations in the data

4. Pseudoinverse of a matrix

将非零 σ 定义为 $\frac{1}{\sigma}$ ，0的保持不变

将 A 转置后 其它部分转置

$$A^+ = V\Sigma^+U^T$$

列满秩的情况 $(A^TA)^{-1}A^T$. 与它一致

实际上 $x = (A^TA)^{-1}A^Tb = V\Sigma^+U^Tb$

性质:

$$\left.\begin{aligned} (A^+)^+ &= A \\ AA^+A &= A \\ (AA^+)^T &= AA^+ \\ (A^+A)^T &= A^+A \end{aligned}\right\} \text{Moore-Penrose conditions}$$