INTRODUCTION TO LINEAR ALGEBRA

一. 对matrix的理解

- 把 matrix multiplying 看成 column form

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = x\begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \end{pmatrix} + y\begin{pmatrix} a_{12} \\ a_{22} \\ a_{32} \end{pmatrix} + z\begin{pmatrix} a_{13} \\ a_{23} \\ a_{33} \end{pmatrix} = A\vec{v} = \vec{b}$$

然后, 可以通过matrix实现每一行与每一行之间的运算

通过追踪这种行之间的计算便可以求出inverse

因 … $EA\vec{v} = (EA)\vec{v}$ … 原理

在求inverse的过程中. $A^{-1}\underbrace{[A\ I]}_{n\times(2n)\text{ matrix}} = \underbrace{[I\ A^{-1}]}_{n\times(2n)\text{ matrix}}$

- 有时 我们想 factorise matrix:

$A = LDU$, where $L = \begin{pmatrix} 1 & 0 & 0 \\ a_{21} & 1 & 0 \\ a_{31} & a_{32} & 1 \end{pmatrix}$ $D = \begin{pmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{pmatrix}$ $U = \begin{pmatrix} 1 & b_{12} & b_{13} \\ 0 & 1 & b_{23} \\ 0 & 0 & 1 \end{pmatrix}$

L的推导用以下propertity: 在$L^{-1}A$中, L^{-1}的l_{ij} means 在A的第i行增加第j行 l_{ij}次

$L^{-1}A = DU \Rightarrow A = LDU$

其用处为, 在计算机计算$A\vec{x} = \vec{b}$时

让 $L\vec{c} = \vec{b}$, 得$DU\vec{x} = \vec{c}$. 此为两个pivot的形式, 因此可以快速求解

- transposes &

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

$(AB)^T = B^T A^T$

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

A^T is invertible exactly when A is invertible

if $A = A^T$ is factoriseed into LDU with no row exchange, then $U = L^T$

if D has elements only on its diagnose, then $D = D^T$

- permutation:

a permutation matrix P has the rows of identity I in any order

there are n! permutation matrices of order n

in this way, $P^{-1} = P^T$, since $PP^T = PP^{-1} = I$

证明: P本质上是一堆在I上的行变换

我们发现有些时候需要row exchange 去factorise找pivot, 所以

$PA = LDU$ 因为 row 互换的P无法被改变

二. Vector space and subspace (m行n列)

- a space satisfies: ① zero vector
  ② closure of addition and scalar multiple

- the column space consists of all linear combinations of columns
  the combination are all possible vectors of $A\vec{x}$ they fill the column space $C(A)$, it is a subset of $\mathbb{R}^m$, where m is number of rows
  the system $A\vec{x}=\vec{b}$ is solvable if and only if $\vec{b}$ is in $C(A)$
  本质上就是以每一column为基底向量所扩张成的空间

- nullspace of A consists of all solutions to $A\vec{x}=\vec{0}$. these vectors $\vec{x}$ are in $\mathbb{R}^n$. the nullspace containing all solutions of $A\vec{x}=\vec{0}$ is $N(A)$, where n is the number of columns ($A\vec{x}=0$, $\vec{x}\in\mathbb{R}^n$)
  this is a subspace since for $\vec{x},\vec{y}\in N(A)$, $\vec{x}+\vec{y}$, $c\vec{x}\in N(A)$ since $A(\vec{x}+\vec{y})=\vec{0}+\vec{0}$, $A(c\vec{x})=cA\vec{x}=\vec{0}$
  a general solution for $A\vec{x}=\vec{0}$ is $\vec{x}=\vec{0}$. and $N(A)$ is the combinations of special solutions
  在我们解special solution的时候：经过elimination，寻找pivot

etc.

$$U=\begin{pmatrix} P & x & x & x & x & x \\ 0 & P & x & x & x & x \\ 0 & 0 & 0 & 0 & P & x \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$U\cdot\vec{x}$, $\vec{x}\in\mathbb{R}^6$
pivot variable: a_1, a_2, a_5
free variable: a_3, a_4, a_6
try: $(a_1, a_2, 1, 0, a_5, 0)$ ……

pivot是那个同column下边全为0，同row以左全0的不为0的P
在计算时，对应column的variable为pivot variable，其余为free variable
考虑特解时让free variables依次为1计算即可

suppose $A\vec{x}=\vec{0}$ has more unknown than equations ($n>m$: more columns that rows), then there are non-zero solutions because there must be free columns without pivots
the rank of A is the number of its pivots / number of independent rows / the dimension of the column space
the pivot columns are not combinations of earlier columns and we can then write pivot columns into 1 in the pivot and 0 in everywhere else

- 如果 $A\vec{x}=\vec{b}$，$\vec{b}\neq\vec{0}$，同理转化成 $R\vec{x}=\vec{d}$，$\vec{d}\neq\vec{0}$
  则令让R中的free variable均为0，得到一个特殊解 ($\vec{x}_p$)
  则通解为 $\vec{x}=\vec{x}_p+\vec{x}_n$，其中 $A\vec{x}_n=\vec{0}$
- 对于rank为r的 $m\times n$ (row x column) matrix

$r=m, r=n$	invertible, so $A\vec{x}=\vec{b}$ has 1 solution	$R=I$
$r=m, r<n$	$A\vec{x}=\vec{b}$ has ∞ solutions	$R=(I\ F)$
$r<m, r=n$	$A\vec{x}=\vec{b}$ has 0 or 1 solution	$R=\begin{pmatrix} I \\ 0 \end{pmatrix}$
$r<m, r<n$	$A\vec{x}=\vec{b}$ has 0 or ∞ solutions	$R=\begin{pmatrix} I & F \\ 0 & 0 \end{pmatrix}$ $N=\begin{pmatrix} -F \\ I \end{pmatrix}$

其中R是一个经过eliminate后的有pivot column的matrix，$RN=0$

- row space of a matrix is the subspace of $\mathbb{R}^n$ spanned by the rows. the row space of A is $C(A^T)$ it is the column space of A^T
  注意：column space 是 $\mathbb{R}^m$ 的subspace，但 row space 是 $\mathbb{R}^n$ 的subspace
- assume $\vec{v}\in\mathbb{R}^n$ and $\vec{v}_i$ are all basis vectors, $1\leq i\leq n$
  there is only one way to write $\vec{v}$ as the combination of basis vectors
  prove by contradiction: if $\vec{v}=\sum_{i=1}^{n}a_i\vec{v}_i=\sum_{i=1}^{n}b_i\vec{v}_i$ and not all $a_i=b_i$.
  so $\vec{0}=\sum_{i=1}^{n}(a_i-b_i)\vec{v}_i$, not all $a_i-b_i=0$
  since basis vectors are linearly independent
  impossible.

  if $\vec{v}_1\cdots\vec{v}_m$ and $\vec{w}_1\cdots\vec{w}_n$ are both bases for the same vector space, then $m=n$
  prove by contradiction: if $n>m$, then $\vec{w}_i$ must be a combination of $\vec{v}_1\cdots\vec{v}_m$

  $$W=(\vec{w}_1\,\vec{w}_2\cdots\vec{w}_n)=(\vec{v}_1\,\vec{v}_2\cdots\vec{v}_m)\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}=VA$$

  and the shape of matrix A is $m\times n$,
  so $A\vec{x}=\vec{0}$ has a non-zero solution
  since basis vectors are linearly independent
  impossible

[后期的note：当matrix放左边，与column有关；放右边与row有关；space本身与matrix有关，nullspace与$A\vec{x}=0$的$\vec{x}$有关]
column space与nullspace一起提出，反之亦然
row space与nullspace只是正好orthogonal，反之亦然]

- the row space and colum space have the same dimension r
$N(A)$ and $N(A^T)$ have dimensions $n-r$ and $m-r$, to make up full n and m [证明见前页表格，以及fixed/free variables]
- an elimination matrix takes A to R, the invertible matrix E is the product of the elementary matrices that reduce A to R

$$EA=R \quad A=E^{-1}R \quad A^TE^T=R^T \quad A^T=R^T(E^{-1})^T$$

(1) A has the same row space as R, same dimension r and same basis
proof: every row of A is the combination of the rows of R
every row of R is a combination of the rows of A
(2) the column space of A has dimension of r.
proof: E是满秩，整个space跟着transform
R
(3) A has the same nullspace as R, same dimension and same basis
proof: the elimination doesn't change the solution, $A\vec{x}=R\vec{x}=\vec{0}$
(4) the left nullspace of A (nullspace of A^T) has the dimension as the left nullspace of R.

- for a matrix with rank of 1, we can write it as $A=\vec{u}\vec{v}^T$, where $\vec{u}$ is the first column and $\vec{v}^T$ is the first row
in this way, the nullspace is the plane perpendicular to $\vec{v}$, since $A\vec{x}=\vec{u}\vec{v}^T\vec{x}=\vec{0}$, $\vec{v}^T\vec{x}=0$

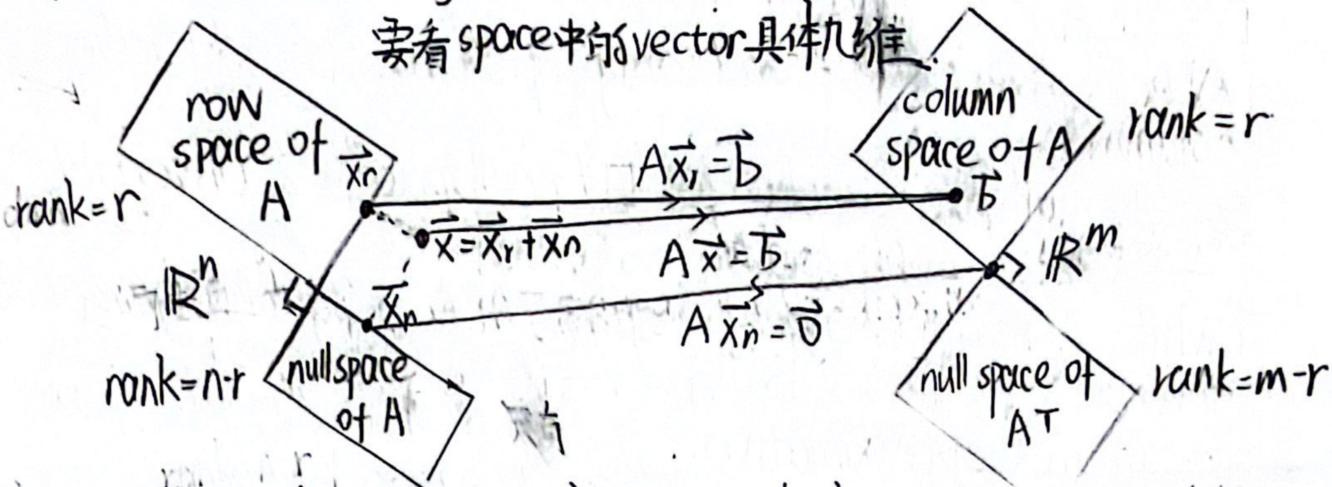
三 Orthogonality 正交性

- the row space is perpendicular to the nullspace, every row of A is perpendicular to every solution of $A\vec{x}=\vec{0}$ [考虑向量化看待base]
the column space is perpendicular to the nullspace of A^T.
- two subspace V and W of a vector space are orthogonal if every vector $\vec{v}$ in V is perpendicular to every vector $\vec{w}$ in

[注意：单纯垂直不一定满足条件. $\vec{v}^T\vec{v}\neq 0$ etc.

- the orthogonal complement of a subspace V contains every vector that is perpendicular to V. this orthogonal subspace is denoted by $V^{\perp}$.

- 要看space中的vector具体几维



N(A) is the orthogonal complement of the row space $C(A^T)$
$N(A^T)$ is the orthogonal complement of the column space $C(A)$
every vector $\vec{b}$ in the column space comes from one and only one vector in row space
prove by contradiction: if $A\vec{x}_r = A\vec{x}_r' \neq \vec{0}$,
$A(\vec{x}_r - \vec{x}_r') = \vec{0}$, $\vec{x}_r - \vec{x}_r' \in N(A^T)$
impossible

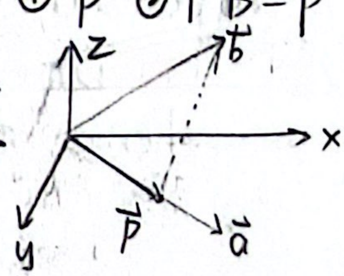
- projection的本质是把一个向量project到一个space上，为了描述space我们用basis

1) to line: 设一线过原点，方向为$\vec{a}$，在这条线上，我们想找一点$\vec{p}$，使$\vec{p}$, $\vec{b}$(任一点)距离最短

设$\vec{p}=\hat{x}\vec{a}$，其中$\hat{x}$为scalar，我们想找：① $\vec{p}$ ② $P\vec{b}=\vec{p}$

for ①: error $\vec{e}=\vec{b}-\hat{x}\vec{a}$

$\vec{a}\cdot(\vec{b}-\hat{x}\vec{a})=0$, $\hat{x}=\frac{\vec{a}\cdot\vec{b}}{\vec{a}\cdot\vec{a}}=\frac{\vec{a}^T\vec{b}}{\vec{a}^T\vec{a}}$



for ②: $\vec{p}=P\vec{b}$

$\frac{\vec{a}\vec{a}^T\vec{b}}{\vec{a}^T\vec{a}}=P\vec{b}$ $\quad P=\frac{\vec{a}\vec{a}^T}{\vec{a}^T\vec{a}}$

2) onto space: 设matrix A 的column均为space basis，我们想在这个subspace上找一点$\vec{p}$，使$\vec{p}$, $\vec{b}$(任一点)距离最短

设$\vec{p}=A\hat{x}$，其中$\hat{x}$为vector，我们想找 ① $\vec{p}$ ② $P\vec{b}=\vec{p}$

for ①: error $\vec{e}=\vec{b}-A\hat{x}$

因为orthogonal, column space 的basis与$\vec{e}$点乘得0

$A^T\vec{e}=\vec{0}$, $A^T(\vec{b}-A\hat{x})=\vec{0}$, $A^TA\hat{x}=A^T\vec{b}$

$\hat{x}=(A^TA)^{-1}A^T\vec{b}$, $\vec{p}=A(A^TA)^{-1}A^T\vec{b}$

for ①: $P=A(A^TA)^{-1}A^T$

[注意: A不一定为square matrix]

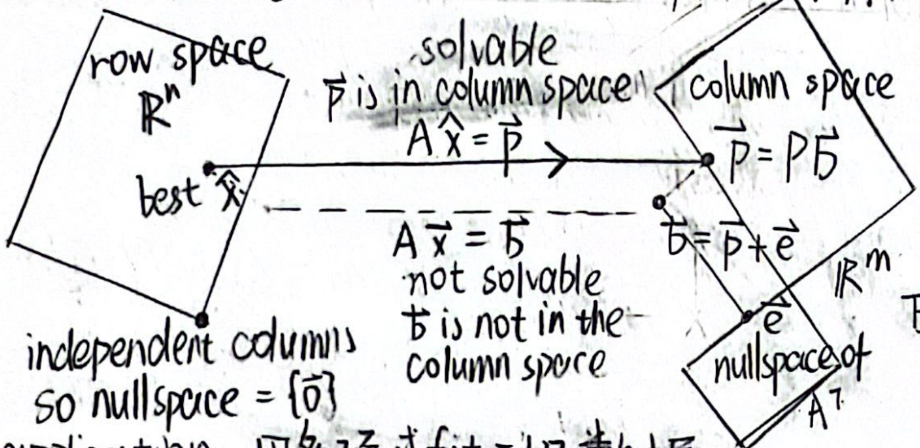
[A^TA is invertible if and only if A has linearly independent columns

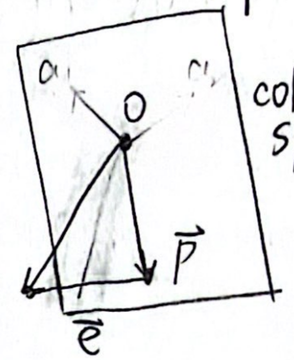
proof: nullspace unchanges. 因为 nullspace $=\mathbb{Z}$(invertible), 所以必须 independent columns]

[while A has independent columns, A^TA is square, symmetric, invertible]

- least squares approximation

when $A\vec{x}=\vec{b}$ has no solution find $A^TA\hat{x}=A^T\vec{b}$ $(=A^T\vec{p})$





application: 用多项式fit一组数据

(1) 如用$C+Dt$来fit数据, 则

$$\begin{cases} C+Dt_1=b_1 \\ C+Dt_2=b_2 \\ \vdots \\ C+Dt_m=b_m \end{cases} \Rightarrow \begin{pmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \\ 1 & t_m \end{pmatrix}\begin{pmatrix} C \\ D \end{pmatrix}=\begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} \Rightarrow A\vec{x}=\vec{b}$$

then solve $A^TA\hat{x}=A^T\vec{b}$, $\hat{x}=\begin{pmatrix} C \\ D \end{pmatrix}$

即 $$\begin{pmatrix} m & \Sigma t_i \\ \Sigma t_i & \Sigma t_i^2 \end{pmatrix}\begin{pmatrix} C \\ D \end{pmatrix}=\begin{pmatrix} \Sigma b_i \\ \Sigma t_i b_i \end{pmatrix}$$

then errors are $e_i=b_i-C-Dt_i$

如果能用orthogonal columns表示matrix会方便计算

let $\hat{t}=\dfrac{t_1+t_2+\cdots+t_m}{m}$, $T_i=t_i-\hat{t}$, thus $\Sigma T_i=0$

这样, 得 $$\begin{pmatrix} m & 0 \\ 0 & \Sigma T_i^2 \end{pmatrix}\begin{pmatrix} C \\ D \end{pmatrix}=\begin{pmatrix} \Sigma b_i \\ \Sigma T_i b_i \end{pmatrix}$$

(2) 如果用parabola去fit

$$\begin{cases} C+Dt_1+Et_1^2=b_1 \\ C+Dt_2+Et_2^2=b_2 \\ \vdots \\ C+Dt_m+Et_m^2=b_m \end{cases} = \begin{pmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ & \vdots & \\ 1 & t_m & t_m^2 \end{pmatrix} \begin{pmatrix} C \\ D \\ E \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

thus $e_i = b_i - C - Dt_i - Et_i^2$

- the vectors $\vec{q}_1, \vec{q}_2, \dots, \vec{q}_n$ are orthonormal if

$$\vec{q}_i^T \vec{q}_j = \begin{cases} 0 & \text{if } i \neq j \text{ (orthogonal vectors)} \\ 1 & \text{if } i = j \text{ (unit vectors } |\vec{q}_i| = 1) \end{cases}$$

a matrix with orthonormal columns is assigned Q

matrix Q is easy to work with since $Q^TQ = I$

when Q is square, $QQ^T = I$ as well, so $Q^T = Q^{-1}$ transpose = inverse

every permutation matrix is an orthonormal matrix

orthonormal matrix leaves lengths unchanged. $|Q\vec{x}| = |\vec{x}|$

Q also preserves dot products: $(Q\vec{x})^T(Q\vec{y}) = \vec{x}^TQ^TQ\vec{y} = \vec{x}^T\vec{y}$

- when finding projection, if projection matrix is orthonormal:

$$Q\vec{x} = \vec{b},\ \hat{x} = Q^T\vec{b}\ (Q^TQ = I),\ P = QQ^T,\ \vec{p} = QQ^T\vec{b}$$

when Q is square, the subspace is the whole space, then the projection of $\vec{b}$ is just $\vec{b}$ itself. then every $\vec{b} = QQ^T\vec{b}$ is the sum of its components along the $\vec{q}_i$'s.

$$\vec{b} = \sum_{i=1}^{n} \vec{q}_i (\vec{q}_i^T \vec{b})$$

这体现了很多transformation的基本思想. break vectors or functions into perpendicular pieces

- the gram-schmidt process

前提 我们有independent vectors $\vec{a}, \vec{b}, \vec{c}$, 我们想要三个 orthonormal vectors $\vec{q}_1, \vec{q}_2, \vec{q}_3$

解法 (1) $\vec{m} = \vec{a}$, $\vec{q}_1 = \dfrac{\vec{m}}{|\vec{m}|}$

(2) $\vec{n} = \vec{b} - \dfrac{\vec{m}^T\vec{b}}{\vec{m}^T\vec{m}}\vec{m}$ 将$\vec{b}$ project至$\vec{m}$所在线上, 并求差

$\vec{q}_2 = \dfrac{\vec{n}}{|\vec{n}|}$

(3) $\vec{h} = \vec{c} - \dfrac{\vec{m}^T\vec{c}}{\vec{m}^T\vec{m}}\vec{m} - \dfrac{\vec{n}^T\vec{c}}{\vec{n}^T\vec{n}}\vec{n}$. 将$\vec{c}$ project至平面mon, 并求差

$\vec{q}_3 = \dfrac{\vec{h}}{|\vec{h}|}$

similarly, we can find orthonormal vectors groups of any number

如果 $A=(\vec{a}\ \vec{b}\ \vec{c})$，我们可以将其用 $Q=(\vec{q_1}\ \vec{q_2}\ \vec{q_3})$ 表示，

$$\underbrace{\underbrace{(\vec{a}\ \vec{b}}_{\text{co-plane}}\ \vec{c})}_{\text{co-space}}=(\vec{q_1}\ \vec{q_2}\ \vec{q_3})\begin{pmatrix}\vec{q_1}^T\vec{a} & \vec{q_1}^T\vec{b} & \vec{q_1}^T\vec{c}\\ 0 & \vec{q_2}^T\vec{b} & \vec{q_2}^T\vec{c}\\ 0 & 0 & \vec{q_3}^T\vec{c}\end{pmatrix}\quad \text{or } A=QR.$$

因为Q是orthogonal，$Q^T=Q^{-1}$，$R=Q^TA$.

你可能还记得在least-square approximation中，如果orthonormal columns，会便利计算

$$A^TA\hat{x}=A^T\vec{b}$$
$$\Downarrow$$
$$R^TQ^TQR\hat{x}=R^TQ^T\vec{b}\qquad \text{since } A=QR$$
$$\Downarrow$$
$$R^TR\hat{x}=R^TQ^T\vec{b}\qquad \text{since } Q^TQ=I$$
$$\Downarrow$$
$$R\hat{x}=Q^T\vec{b},\ \hat{x}=R^{-1}Q^T\vec{b}$$

四 determinants

(1) $\begin{vmatrix}ta & tb\\ c & d\end{vmatrix}=t\begin{vmatrix}a & b\\ c & d\end{vmatrix}$. $\begin{vmatrix}a+a' & b+b'\\ c & d\end{vmatrix}=\begin{vmatrix}a & b\\ c & d\end{vmatrix}+\begin{vmatrix}a' & b'\\ c & d\end{vmatrix}$.

(2) determinant changes signs when two rows are exchanged

推论：用(1)(2)可以得到将matrix任一row拆分并随后算det的可行性

(3) if two rows are equal, then $\det M=0$

proof: (2), $\det A=-\det A$

(4) subtracting a multiple of one row from another row leaves $\det A$ unchanged

proof (1),(3) $\begin{vmatrix}a & b\\ c-la & d-lb\end{vmatrix}=\begin{vmatrix}a & b\\ c & d\end{vmatrix}-l\begin{vmatrix}a & b\\ a & b\end{vmatrix}$

推论：$|\det A|$ isn't changed by usual elimination from A to U ($PA=LDU$)

If U is triangular then $\det U=a_{11}a_{22}\cdots a_{nn}$ = product of diagonal entries

(5) the transpose A^T has the same determinant as A

proof: $PA=LDU$. P represents permutation matrix. D is diagonal matrix. L, U are triangular matrices

$(PA)^T=(LDU)^T$, $A^TP^T=U^TD^TL^T$

$\because \det P=\det P^T=1$, $\det U=\det U^T$, $\det L=\det L^T$, $D=D^T$

[对角线未变]

$\therefore \det A=\det A^T$

推论：every role for rows can be applied into columns since $\det A=\det A^T$

- $\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \begin{vmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{vmatrix} + \begin{vmatrix} 0 & a_{12} & 0 \\ 0 & 0 & a_{23} \\ a_{31} & 0 & 0 \end{vmatrix} + \begin{vmatrix} 0 & 0 & a_{13} \\ a_{21} & 0 & 0 \\ 0 & a_{32} & 0 \end{vmatrix} + \begin{vmatrix} a_{11} & 0 & 0 \\ 0 & 0 & a_{23} \\ 0 & a_{32} & 0 \end{vmatrix} +$

$\begin{vmatrix} 0 & a_{12} & 0 \\ a_{21} & 0 & 0 \\ 0 & 0 & a_{33} \end{vmatrix} + \begin{vmatrix} 0 & 0 & a_{13} \\ 0 & a_{22} & 0 \\ a_{31} & 0 & 0 \end{vmatrix}$ [rule (1)]

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

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

[rule (1)]

$= a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{3}$

[rule (2)]

在每一column和row取一个非零的数(为了得到kP)

由此推导出一般结论: for square matrix with order n

$\det A = a_{11}C_{11} + a_{12}C_{12} + \cdots + a_{1n}C_{1n}$ [保持column不变也可]

each cofactor C_{ij} (order $n-1$, without row i and column j) include its correct sign: $C_{ij} = (-1)^{i+j}\det M_{ij}$

- Cramer's Rule 求解 $A\vec{x} = \vec{b}$

$A\begin{pmatrix} x_1 & 0 & 0 \\ x_2 & 1 & 0 \\ x_3 & 0 & 1 \end{pmatrix} = \begin{pmatrix} b_1 & a_{12} & a_{13} \\ b_2 & a_{22} & a_{23} \\ b_3 & a_{32} & a_{33} \end{pmatrix} = B_1$

$\det A \cdot x_1 = \det B_1$, $x_1 = \dfrac{\det B_1}{\det A}$ [左右求determinant]

if $\det A \neq 0$, $A\vec{x} = \vec{b}$ is solved by determinants:

$x_1 = \dfrac{\det B_1}{\det A}$, $x_2 = \dfrac{\det B_2}{\det A}$, ..., $x_n = \dfrac{\det B_n}{\det A}$

the matrix B_j has the jth column of A replaced by $\vec{b}$

this can be used to derive inverse

the i,j entry of A^{-1} is the cofactor C_{ji} divided by $\det A$

$$(A^{-1})_{ij} = \frac{C_{ji}}{\det A}, \quad A^{-1} = \frac{C^T}{\det A}$$

where C is a cofactor matrix $\begin{pmatrix} C_{11} & C_{12} & \cdots & C_{1n} \\ C_{21} & C_{22} & \cdots & C_{2n} \\ \vdots & & \ddots & \\ C_{n1} & C_{n2} & \cdots & C_{nn} \end{pmatrix}$

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}\begin{pmatrix} C_{11} & C_{21} & \cdots & C_{n1} \\ C_{12} & C_{22} & \cdots & C_{n2} \\ \vdots & \vdots & \ddots & \\ C_{1n} & C_{2n} & \cdots & C_{nn} \end{pmatrix} = \begin{pmatrix} \det A & 0 & \cdots & 0 \\ 0 & \det A & \cdots & 0 \\ \vdots & & \ddots & \\ 0 & 0 & \cdots & \det A \end{pmatrix}$$

proof: $a_{11}C_{11} + a_{12}C_{12} + \cdots + a_{1n}C_{1n} = \det A$ by cofactor rules

除此之外，别的 column from A · 别的 row from C = 0
因为总有一 row/column 有2个元素，因此有一 row/column 全为 0
determinant = 0 [毕竟归根结底这玩意是用来算 determinant 的]
所以 $AC^T = [\det A] I$　$A^{-1} = \frac{C^T}{\det A}$

- the triangle with corners $(x_1, y_1), (x_2, y_2), (x_3, y_3)$ has area

$$A = \frac{1}{2}\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix}$$

同理可证 parallelogram 的面积

其证明为：通过对比 area/volumn 与 determination properties 证明 area/volumn 和 $|\det M|$ 是一样的

① when $M = I$, $|\det M|$ = area/volumn = 1
② when row exchange, $|\det M|$ = area volumn = same
③ when row i is multiplied by t / added up by t, change in $|\det M|$ = change in area/volumn

- $|\vec{u} \times \vec{v}| = |\vec{u}||\vec{v}||\sin\theta|$
$|\vec{u} \times \vec{v}| \cdot \vec{w} = 0$ exactly when $\vec{u}, \vec{v}, \vec{w}$ lie in same plane

五、eigenvalue & eigenvector

- if $A\vec{x} = \lambda\vec{x}$，$(A - \lambda I)\vec{x} = \vec{0}$. 说明该式有非零解
所以 $\det(A - \lambda I) = 0$
- the product of n eigenvalues equals determinant [detA = detA]
the sum of n eigenvalues equals the sum of n diagonal entries [计算公式 characteristic equation 可得]
- a symmetric matrix ($A^T = A$) can be compared to a real number
a skew-symmetic matix ($A^T = -A$) can be compared to a complex number
an orthogonal matrix ($A^T A = I$) can be compared to a complex number with $|\lambda| = 1$
- eigenvalue matrix Λ, eigenvector matrix S
$AS = S\Lambda$, so $S^{-1}AS = \Lambda$, $A = S\Lambda S^{-1}$
- different $\vec{x}$ from different λ: eigenvectors $\vec{x}_1, \vec{x}_2, \cdots, \vec{x}_j$ that correspond to distinct eigenvalues are linearly independent.

an $n\times n$ matrix that has n different eigenvalues (no repeated λ's) must be diagonalizable

proof: suppose $c_1\vec{x}_1 + c_2\vec{x}_2 = 0$

$\therefore \lambda_1(c_1\vec{x}_1 + c_2\vec{x}_2) = c_1\lambda_1\vec{x}_1 + c_2\lambda_1\vec{x}_2 = 0$

$A(c_1\vec{x}_1 + c_2\vec{x}_2) = c_1\lambda_1\vec{x}_1 + c_2\lambda_2\vec{x}_2 = 0$

$\therefore (\lambda_1 - \lambda_2)c_2\vec{x}_2 = 0 \quad \because \lambda_1 \neq \lambda_2 \quad \therefore c_2 = 0$; impossible

- Fibonacci Numbers.

assume $\vec{u}_k = \begin{pmatrix} F_{k+1} \\ F_k \end{pmatrix}$, $\vec{u}_{k+1} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}\vec{u}_k = M\vec{u}_k$

$\because M = \begin{pmatrix} \frac{1+\sqrt{5}}{2} & \frac{1-\sqrt{5}}{2} \\ 1 & 1 \end{pmatrix}\begin{pmatrix} \frac{1+\sqrt{5}}{2} & 0 \\ 0 & \frac{1-\sqrt{5}}{2} \end{pmatrix}\begin{pmatrix} \frac{1+\sqrt{5}}{2} & \frac{1-\sqrt{5}}{2} \\ 1 & 1 \end{pmatrix}^{-1}$

$\therefore u_k = \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^{k>0} - \left(\frac{1-\sqrt{5}}{2}\right)^{k>0}\right]$ [可以设 $u_k = c_1\lambda_1^k + c_2\lambda_2^k$, 求 c_1, c_2]

- how eigenvalues work: $S\Lambda^k S^{-1}\vec{u}_0 = \vec{u}_k$

① write $\vec{u}_0$ as a combination $c_1\vec{x}_1 + c_2\vec{x}_2 + \cdots + c_n\vec{x}_n$ of the eigenvectors; then $\vec{x} = S^{-1}\vec{u}_0$

② multiple each eigenvector $\vec{x}_i$ by $(\lambda_i)^k$; so we have $\Lambda^k S^{-1}\vec{u}_0$

③ add up the pieces $c_i(\lambda_i)^k\vec{x}_i$ to find solution $\vec{u}_k = A^k\vec{u}_0$. this is $S\Lambda^k S^{-1}\vec{u}_0$

- suppose A and B can be diagonalized and they share the same eigenvector matrix S if and only if $AB = BA$

only when share same eigenvector matrix S.

$AB\vec{x} = BA\vec{x} = \lambda\beta\vec{x} = \beta\lambda\vec{x}$

$(A+B)\vec{x} = (B+A)\vec{x} = (\lambda+\beta)\vec{x} = (\beta+\lambda)\vec{x}$

- solution of $\frac{d\vec{u}}{dt} = A\vec{u}$:

assume $\vec{u}$ is a vector:

if we let $\vec{u}(t) = e^{\lambda t}\vec{x}$, where $A\vec{x} = \lambda\vec{x}$

then $LHS = \lambda e^{\lambda t}\vec{x} = RHS$

$\because$ linear combination of roots is also root

$\therefore \vec{u}(t) = c_1 e^{\lambda_1 t}\vec{x}_1 + c_2 e^{\lambda_2 t}\vec{x}_2 + \cdots + c_n e^{\lambda_n t}\vec{x}_n$

其中 c_i 用初始条件计算即可

注: if two λ's are equal, with only one eigenvector, another solution is still needed: it will be $te^{\lambda t}\vec{x}$

if $y''+ay'+by=0$, then $\begin{cases}\frac{dy}{dt}=y' \\ \frac{dy'}{dt}=-ay'-by\end{cases}$

$\therefore \frac{d}{dt}\begin{pmatrix}y\\y'\end{pmatrix}=\begin{pmatrix}0&1\\-b&-a\end{pmatrix}\begin{pmatrix}y\\y'\end{pmatrix}$.

$\therefore \vec{x}_1=\begin{pmatrix}1\\\lambda_1\end{pmatrix}, \vec{x}_2=\begin{pmatrix}1\\\lambda_2\end{pmatrix}$, $\lambda^2+a\lambda+b=0$

$\therefore \vec{u}(t)=c_1e^{\lambda_1 t}\begin{pmatrix}1\\\lambda_1\end{pmatrix}+c_2e^{\lambda_2 t}\begin{pmatrix}1\\\lambda_2\end{pmatrix}$, $y=c_1e^{\lambda_1 t}+c_2e^{\lambda_2 t}$

- finite difference equation/method:
画图的时候将 $y(t)$ 用 $y(t-\Delta t)$, $z(t-\Delta t)$, Δt 等表示, 将 $z(t)$ 用 $y(t-\Delta t)$, $y(t)$, $z(t-\Delta t)$, Δt 等表达, 一步步分开算出 $y(t)$, $z(t)$ 并画图, 好处是可以将 differential equation 化成 $U_{n+1}=AU_n$

- whether stable matrix? whether $\vec{u}=\vec{0}$ when $t\to\infty$?
we want $\lambda_i<0$ or $Re(\lambda_i)<0$
so for 2x2 matrix $A\begin{pmatrix}a&b\\c&d\end{pmatrix}$, A is stable if

$\begin{cases}\lambda_1+\lambda_2<0 \to \text{trace } T=a+d<0\\ \lambda_1\lambda_2>0 \to \det A>0\end{cases}$

- 事实上, $\frac{d\vec{u}}{dt}=A\vec{u}$, $\vec{u}=e^{At}\vec{u}(0)$:

$e^{At}=I+At+\frac{1}{2}(At)^2+\frac{1}{6}(At)^3+\cdots$

$\frac{d}{dt}(e^{At})=A+A^2t+\frac{1}{2}A^3t^2+\cdots=Ae^{At}$

if $\vec{x}$ is the eigenvector of A, λ is corresponding eigenvalue:

$e^{At}\vec{x}=(I+At+\frac{1}{2}(At)^2+\frac{1}{6}(At)^3+\cdots)\vec{x}=(1+\lambda t+\frac{1}{2}(\lambda t)^2+\frac{1}{6}(\lambda t)^3+\cdots)\vec{x}$

if $A=S\Lambda S^{-1}$:

$e^{At}=I+(S\Lambda S^{-1})t+\frac{1}{2}(S\Lambda S^{-1})(S\Lambda S^{-1})t^2+\cdots$

$=I+S\Lambda S^{-1}t+\frac{1}{2}S\Lambda^2S^{-1}t^2+\cdots=S(I+\Lambda t+\frac{1}{2}(\Lambda t)^2+\cdots)S^{-1}$

$=Se^{\Lambda t}S^{-1}=S\begin{pmatrix}e^{\lambda_1 t}&&\\&\ddots&\\&&e^{\lambda_n t}\end{pmatrix}S^{-1}$

- e^{At} always has the inverse e^{-At} $[Se^{\Lambda t}S^{-1}\cdot Se^{-\Lambda t}S^{-1}]$
the eigenvalues of e^{At} are always $e^{\lambda t}$
when A is skew-symmetric ($A^T=-A$), e^{At} is orthogonal, $(e^{At})^{-1}=(e^{At})^T=e^{-At}$ $\left[\left(\sum_{k=0}^{\infty}\frac{(At)^k}{k!}\right)^T=\sum_{k=0}^{\infty}\left(\frac{(At)^k}{k!}\right)^T=\sum_{k=0}^{\infty}\frac{((A^T)t)^k}{k!}=\sum_{k=0}^{\infty}\frac{(-At)^k}{k!}\right]$

- a symmetric matrix ($A^T=A$) has only real eigenvalues, its eigenvector can be chosen orthonormal

[proof:

... and B are symmetric positive definite so is A+B

every symmetric matrix has factorization $A = Q\Lambda Q^T$ with real eigenvalues in Λ and orthonormal eigenvectors in $S = Q$; with $Q^{-1} = Q^T$:

(1) eigenvalues are real

if λ is complex, $\bar{\lambda}$ is its conjugate,

$\because A\vec{x} = \lambda\vec{x}$ $\quad \therefore A\bar{\vec{x}} = \bar{\lambda}\bar{\vec{x}}$ (probably $\vec{x}$ consists of complex elements)

$\therefore \bar{\vec{x}}^T A = \bar{\vec{x}}^T\bar{\lambda}$ (transpose), $\bar{\vec{x}}^T A\vec{x} = \bar{\vec{x}}^T\bar{\lambda}\vec{x}$

$\because \bar{\vec{x}}^T A\vec{x} = \bar{\vec{x}}^T\lambda\vec{x}$ (from $A\vec{x} = \lambda\vec{x}$) $\quad \therefore \lambda = \bar{\lambda}$, real ($\bar{\vec{x}}^T\vec{x} \neq 0$)

(2) non-repeated eigenvalues lead to orthogonal eigenvectors

assume $A\vec{x} = \lambda_1\vec{x}$. $A\vec{y} = \lambda_2\vec{y}$. $\lambda_1 \neq \lambda_2$

$$(\lambda_1\vec{x})^T\vec{y} = (A\vec{x})^T\vec{y} = \vec{x}^T A^T\vec{y} = \vec{x}^T A\vec{y} = \vec{x}^T(\lambda_2\vec{y})$$

$\because$ LHS $= \lambda_1\vec{x}^T\vec{y}$. RHS $= \lambda_2\vec{x}^T\vec{y}$ $\quad \therefore \vec{x}^T\vec{y} = 0$

$\therefore$ eigenvectors can be chosen orthonormal (unit length)

(3) schur's theorem: all symmetric matrices are diagonalizable.

schur's theorem: every square matrix factors into $A = QTQ^{-1}$, where T is upper triangular and $\bar{Q}^T = Q^{-1}$. if A has real eigenvalues, then Q and T can be real: $Q^TQ = QQ^T = I$

A是n×n matrix, $\vec{x}_1$是它的一个eigenvector, $A\vec{x}_1 = \lambda_1\vec{x}_1$

对于$\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n$, $\|\vec{x}_i\| = 1$, $\vec{x}_i \cdot \vec{x}_j = 0$ if $i \neq j$

we first consider $P = [\vec{x}_1\ \vec{x}_2\ \vec{x}_3 \cdots \vec{x}_n]$;

$$\bar{P}^T A P = \begin{bmatrix} \bar{\vec{x}}_1^T \\ \bar{\vec{x}}_2^T \\ \vdots \\ \bar{\vec{x}}_n^T \end{bmatrix} A \begin{bmatrix} \vec{x}_1 & \vec{x}_2 & \cdots & \vec{x}_n \end{bmatrix} = \begin{bmatrix} \bar{\vec{x}}_1^T \\ \bar{\vec{x}}_2^T \\ \vdots \\ \bar{\vec{x}}_n^T \end{bmatrix} \begin{bmatrix} \lambda_1\vec{x}_1 & A\vec{x}_2 & A\vec{x}_3 & \cdots & A\vec{x}_n \end{bmatrix}$$

$$= \left[\begin{array}{c|c} \lambda_1 & (*) \\ \hline 0 & \\ \vdots & \bar{Q}^T U_0 Q \\ 0 & \end{array}\right]$$

[注:
我们用induction假设$(n-1)\times(n-1)$ matrix也可表示为类似于$A = QTQ^{-1}$的形式. 这里$\bar{Q}^T = Q^{-1}$, U_0是upper triangular matrix]

$$= \underbrace{\left[\begin{array}{c|c} 1 & 0 \cdots 0 \\ \hline 0 & \\ \vdots & \bar{Q}^T \\ 0 & \end{array}\right]}_{\bar{B}^T} \underbrace{\left[\begin{array}{c|c} \lambda_0 & (*)\bar{Q}^T \\ \hline 0 & \\ \vdots & U_0 \\ 0 & \end{array}\right]}_{U} \underbrace{\left[\begin{array}{c|c} 1 & 0 \cdots 0 \\ \hline 0 & \\ \vdots & Q \\ 0 & \end{array}\right]}_{B}$$

$\therefore (B\bar{P}^T)A(P\bar{B}^T) = U$ $\quad \because (B\bar{P}^T)(P\bar{B}^T) = I$

$\therefore$ if $P\bar{B}^T = X$, $\bar{X}^TX = X\bar{X}^T = I$ $\quad \therefore \bar{X}^TAX = U$ (证毕)

$\therefore A = QTQ^T$. when A is real symmetric matrix

$\therefore A = Q\Lambda Q^T$, the matrix $A = A^T$ has n orthogonal eigenvectors

- for real matrices, complex λ's and $\vec{x}$'s come in "conjugate pairs"
  if $A\vec{x} = \lambda\vec{x}$, then $A\bar{\vec{x}} = \bar{\lambda}\bar{\vec{x}}$

- generally, for all matrices:

  products of pivots = determinant = products of eigenvalues

specially: for all symmetric matrices:

the number of positive eigenvalues = the number of positive pivot

proof:

we firstly consider $A = LDL^T$, we can find the pivots $d_1, d_2, \dots d_n$.
we gradually change L into I, until A being turned into D in the end, D has eigenvalues $d_1, d_2, \dots d_n$
assume the sign of eigenvalues of A, $\lambda_1, \lambda_2, \dots \lambda_n$, are not the same as those of D, $d_1, d_2, \dots, d_n$
then there exists a time such that $A' = L' D (L')^T$ is singular so in such way $\lambda'_i = 0$ for some i
impossible since pivots $d_1, d_2, \dots, d_n$, $\det D = \det A' \neq 0$.
since $\lambda'_1, \lambda'_2, \dots \lambda'_n$ never cross 0, then $\lambda_1, \lambda_2, \dots, \lambda_n$ have the same signs as $d_1, d_2, \dots, d_n$

- positive definite matrix.

for a symmetric matrix $A = A^T$, it is a positive definite matrix if all eigenvalues are positive

或者,以下五个条件,满足一个则满足全部

(1) all n pivots are positive
(2) all n upper left determinants are positive (本质是为了查pivots)
(3) all n eigenvalues are positive
(4) $\vec{x}^T A \vec{x} > 0$ when $\|\vec{x}\| \neq 0$ (energy-based definition)
(5) $A = R^T R$ for matrix R with independent columns (R行大于不一定)

proof:

(4) assume we have orthogonal eigenvectors $\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n$ and corresponding positive eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$
if $\vec{x} = \sum c_i \vec{x}_i$, c_i is constant

$\therefore \vec{x}^T A \vec{x} = (\sum c_i \vec{x}_i)^T A (\sum c_i \vec{x}_i) = \sum c_i^2 \vec{x}_i^T A \vec{x}_i$ since $\vec{x}_i^T \vec{x}_j = 0, i \neq j$

$= \sum c_i^2 \lambda_i \|x_i\|^2 > 0$

(5) $\vec{x}^T A \vec{x} = \vec{x}^T R^T R \vec{x} = (\vec{x} R^T)(R\vec{x}) = (R\vec{x})^T (R\vec{x}) = \|R\vec{x}\|^2 > 0$

If A and B are symmetric positive definite, so is A+B

proof $\vec{x}^T(A+B)\vec{x} = \vec{x}^T A\vec{x} + \vec{x}^T B\vec{x} > 0$

- symmetric matrices with non-negative eigenvalues are positive semidefinite matrices

同理.以下条件相互满足:

(1) non-negative eigenvalues

(2) non-negative pivots

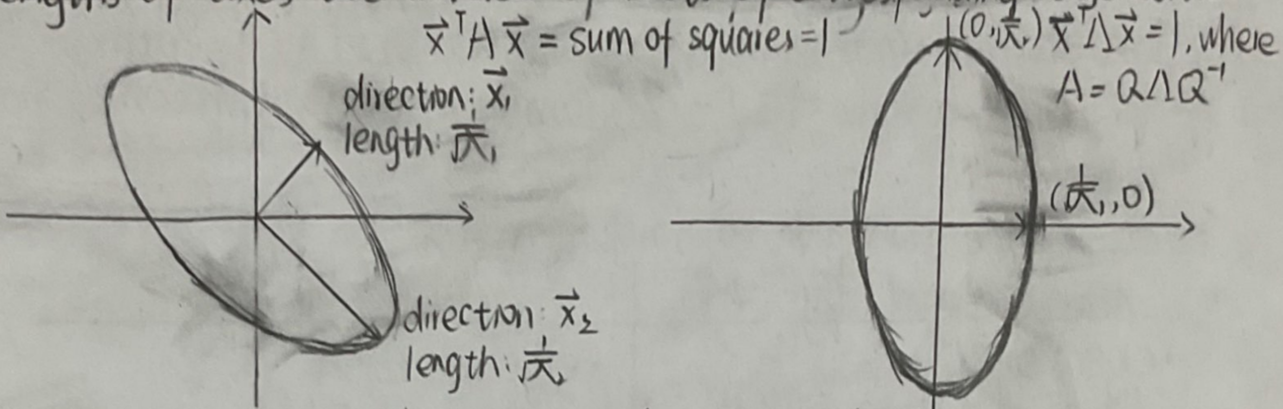
(3) $A = R^T R$, with dependent or independent columns in R.

- application of positive definite matrix in ellipse

consider $\vec{x}^T A\vec{x} = 1$, where A is a positive definite matrix

axes of ellipse point along eigenvectors

lengths of axes are inverse of roots of corresponding eigenvalues



- let M be any invertible matrix, then $B = M^{-1}AM$ is similar to A, because if A has eigenvalues λ_i and corresponding eigenvectors $\vec{x}_i$, then B has the same eigenvalues λ_i and related corresponding eigenvectors $M^{-1}\vec{x}_i$

proof: $\because B = M^{-1}AM \quad \therefore A = MBM^{-1}$

$\therefore BM^{-1}\vec{x}_i = M^{-1}MBM^{-1}\vec{x}_i = M^{-1}A\vec{x}_i = \lambda_i M^{-1}\vec{x}_i$

- not every square matrix can be diagonalised, but all of them can be expressed in Jordan matrix

algebraic multiplicity represents the number of eigenvalues (including repeated ones)

geometric multiplicity represents the number of independent eigenvectors

if A has s independent eigenvectors, it is similar to a matrix J that has a Jordan blocks on its diagonal: some matrix M puts A into Jordan form

Jordan form:

$$J = M^{-1}AM = \begin{bmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_s \end{bmatrix}$$

each block in J has one (unique) eigenvalue λ_i, one eigenvector, and 1's above the diagonal.

Jordan block:

$$J_i = \begin{bmatrix} \lambda_i & 1 & & \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{bmatrix}$$

A is similar to B if they share the same Jordan form (matrix) J.

[注意: (1) J_i 中 λ_i 出现几次取决于 λ_i 为几次 roots]

(2) Jordan matrix 是最接近 diagonalised matrix 的 matrix.

(3) 一个 Jordan matrix 对应了一整个 family of matrix with corresponding eigenvalues

(4) during the transition to Jordan matrix:

not changed by M	changed by M
• eigenvalues	• eigenvectors
• trace and determinant	• nullspace
• rank	• row space
• number of independent eigenvectors	• left nullspace
• Jordan form	• column space
	• singular matrix

(5) $A^k = (M^{-1}JM)(M^{-1}JM)\cdots(M^{-1}JM) = M^{-1}J^kM$]

- application of Jordan matrix in differential equation.

if $\frac{d\vec{u}}{dt} = A\vec{u}$, $J = M^{-1}AM$,

assume $\vec{u} = M\vec{v}$: $M\frac{d\vec{v}}{dt} = AM\vec{v}$, $\frac{d\vec{v}}{dt} = M^{-1}AM\vec{v}$

$$\frac{d\vec{v}}{dt} = J\vec{v}$$

设 $\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$, 在计算中会有几种可能:

(1) $\frac{dv_i}{dt} = \lambda_i v_i$, so $v_i = v_i(0)e^{\lambda_i t}$ [对应的eigenvalue首次出现/不重复

(2) $\frac{dv_i}{dt} = \lambda_i v_i + v_{i-1}$, so $v_i = \begin{cases} [v_i(0) + t v_{i-1}(0)]e^{\lambda_i t} \\ [v_i(0) + t v_{i-1}(0) + \frac{1}{2}t^2 v_{i-2}(0)]e^{\lambda_i t} \\ \cdots \end{cases}$

[对应的eigenvalue repeat]

- singular value decomposition (SVD)

对于所有的matrix，它们的eigenvectors大多有三个问题

(1) not orthogonal (2) not enough eigenvectors (3) $A\vec{x}=\lambda\vec{x}$, A is square

所以我们让 $A=U\Sigma V^T$; U & V are orthogonal, Σ is diagonal

$$\underbrace{A}_{m\times n}=\underbrace{\begin{bmatrix}\vec{u}_1 & \vec{u}_2 & \cdots & \vec{u}_r\end{bmatrix}}_{m\times r}\underbrace{\begin{bmatrix}\sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r\end{bmatrix}}_{r\times r}\underbrace{\begin{bmatrix}\vec{v}_1^T \\ \vec{v}_2^T \\ \vdots \\ \vec{v}_r^T\end{bmatrix}}_{r\times n}$$

其中 $\vec{u}_i$'s are eigenvectors of AA^T, $\vec{v}_i$'s are eigenvectors of A^TA

since AA^T and A^TA are symmetric, u_i's and v_i's can be chosen to be orthonormal

$AA^T=(U\Sigma V^T)(V\Sigma^T U^T)=U(\Sigma\Sigma^T)U^T$ [注: $V^TV=I$]

$A^TA=(V\Sigma^T U^T)(U\Sigma V^T)=V(\Sigma^T\Sigma)V^T$ [注: $U^TU=I$]

σ_i's are positive, they are the square root of eigenvalues of AA^T and A^TA

$A(A^TA)=(AA^T)A$ gives:

$$A\begin{bmatrix}\vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_r\end{bmatrix}=\begin{bmatrix}\vec{u}_1 & \vec{u}_2 & \cdots & \vec{u}_r\end{bmatrix}\begin{bmatrix}\sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r\end{bmatrix}$$

$A\vec{v}_1=\sigma_1\vec{u}_1$, $A\vec{v}_2=\sigma_2\vec{u}_2$ $\cdots$ $A\vec{v}_r=\sigma_r\vec{u}_r$

SVD: $A=\sum_{i=1}^{r}\vec{u}_i\sigma_i\vec{v}_i^T$

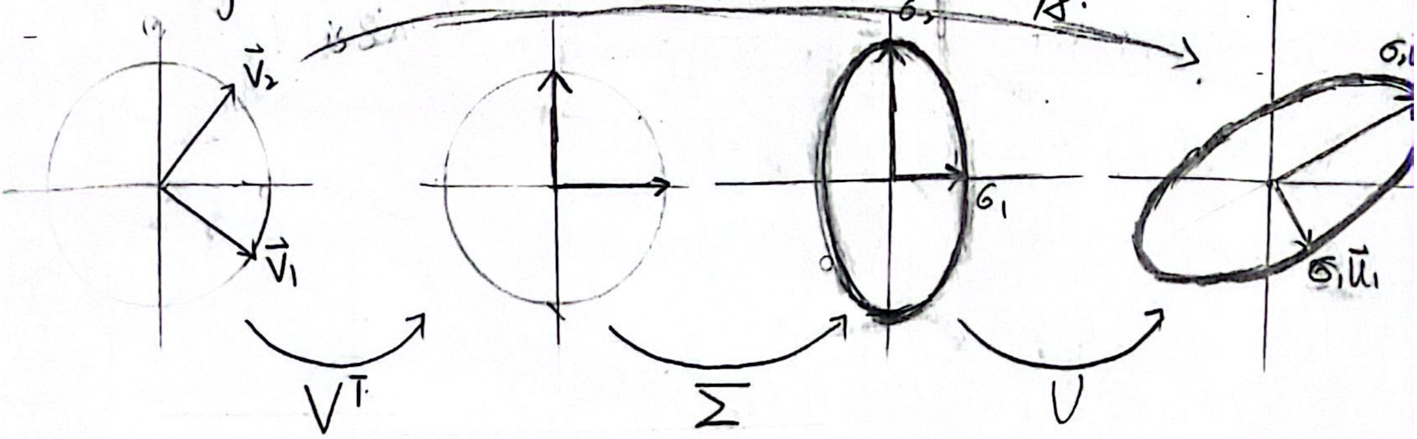
[注意 (1) $\vec{v}_i$'s from row space of A; $\vec{u}_i$'s from column space of A

if we add up $(n-r)$ more $\vec{v}$'s and $(m-r)$ more $\vec{u}$'s from nullspace $N(A)$ and left nullspace $N(A^T)$ and make them orthonormal:

$$A\begin{bmatrix}\vec{v}_1 & \cdots & \vec{v}_r & \cdots & \vec{v}_n\end{bmatrix}=\begin{bmatrix}\vec{u}_1 & \cdots & \vec{u}_r & \cdots & \vec{u}_m\end{bmatrix}\begin{bmatrix}\sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & 0 \\ & & & & \ddots\end{bmatrix}$$

still holds. [用square matrix找inverse更加合理(?)]

(2) $U\Sigma V^T$ is $S\Lambda S^{-1}$ if A is positive semidefinite (definite) symmetric matrix $Q\Lambda Q^T$.



(3) 在求SVD的时候，用 AA^T 与 A^TA 求 eigenvectors & eigenvalues
或者 $A\vec{v}_i = \sigma_i \vec{u}_i$ 均可（注意 $\vec{u}_i$ 与 $\vec{v}_i$ 为unit vectors）
可以只要 row space & column space，也可加上 nullspace & left nullspace
(4) why $A\vec{v}_i$ falls in the direction of $\vec{u}_i$?
$\because A^TA\vec{v}_i = \sigma_i^2 \vec{v}_i$
$\therefore \vec{v}_i^T A^T A \vec{v}_i = (A\vec{v}_i)^T(A\vec{v}_i) = \|A\vec{v}_i\|^2 = \sigma_i^2 \vec{v}_i^T \vec{v}_i = \sigma_i^2$
$|A\vec{v}_i| = \sigma_i$

$$\therefore AA^TA\vec{v}_i = \begin{cases} A(A^TA\vec{v}_i) = \sigma_i^2 A\vec{v}_i & \text{[通过上述推导]} \\ (AA^T)(A\vec{v}_i) = \sigma_i^2 A\vec{v}_i & \text{[是eigenvalues的形式]} \end{cases}$$

$\therefore A\vec{v}_i$ falls in the direction of $\vec{u}_i$，$\vec{u}_i = \frac{A\vec{v}_i}{\sigma_i}$ [unit vector]
and $A\vec{v}_i$'s are orthogonal:
$(A\vec{v}_i)^T(A\vec{v}_j) = \vec{v}_i^T A^T A \vec{v}_j = \sigma_j^2 \vec{v}_i^T \vec{v}_j = 0$

- we can use SVD for compression

when we put the singular values in descending order, $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$
the splitting in SVD gives r rank-one pieces of A in order of importance.
比如我们在压缩图片，用矩阵表示像素点阵，可以用SVD将矩阵压缩
if we want best rank k approximation, $A \approx \sigma_1 \vec{u}_1 \vec{v}_1^T + \cdots + \sigma_k \vec{u}_k \vec{v}_k^T$
corresponding titles of books with key words 同理
再或者我们考虑 search engine，所有网站分成两种：authority, hub

$$\begin{cases} \text{authority: links come in from many hubs} \\ \text{hub: links go out to many authority} \end{cases}$$

我们想用 $x_1, x_2, \ldots, x_n$ 与 $y_1, y_2, \ldots, y_m$ 去评价authority, hub
1st stage: good authority (hub) is linked to many important (hub (authority)
[右上角数据代表几次数据加工]

$$\begin{cases} \text{authority:} & x_i^1 = \sum_{j \text{ links to } i} y_j^0 \longrightarrow \vec{x}^1 = A^T \vec{y}^0 \\ \text{hub:} & y_i^1 = \sum_{i \text{ links to } j} x_j^0 \longrightarrow \vec{y}^1 = A\vec{x}^0 \end{cases}$$

A contains 1's and 0's, with $a_{ij} = 1$ if i links to j
2nd stage:

$$\begin{cases} \text{authority:} & \vec{x}^2 = A^T \vec{y}^1 = A^TA\vec{x}^0 \\ \text{hub:} & \vec{y}^2 = A\vec{x}^1 = AA^T\vec{y}^0 \end{cases}$$

重复多次，σ_i 的作用开始发力，与SVD类似

六 linear transformations

- a transformation T assigns an output $T(\vec{v})$ to each input vector $\vec{v}$ in V. the transformation is linear if it meets these requirements for all $\vec{v}$ and $\vec{w}$ in V:

(1) $T(\vec{v}+\vec{w}) = T(\vec{v}) + T(\vec{w})$ (2) $T(c\vec{v}) = cT(\vec{v})$ / for all real c

(3) if $\vec{v}=\vec{0}$, $T(\vec{v})=\vec{0}$

注意: input space V and output space W.

- shift transformation alone is not linear (apart from identity transformation) the linear-plus-shift transformation is called affine
- linearity: $\vec{u} = c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_n\vec{v}_n$ transforms to

$$T(\vec{u}) = c_1T(\vec{v}_1) + c_2T(\vec{v}_2) + \cdots + c_nT(\vec{v}_n)$$

equally-spaced points go to equally-spaced points

- range of T = set of all output $T(\vec{v})$ = colume space

kernel of T = set of all input for which $T(\vec{v}) = \vec{0}$ = nullspace

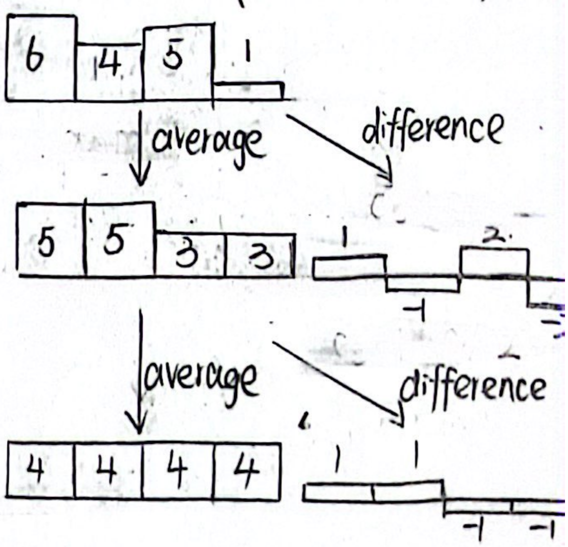
- 所有的linear transformation都能用matrix表达. 只要考虑好input space V中的basis vectors经过transformation后怎样用output space W中的basis vectors表达即可
- Wavelet Transform

这是一种压缩数据的方法:

haar basis: $\vec{w}_1 = \begin{pmatrix}1\\1\\1\\1\end{pmatrix}$ $\vec{w}_2 = \begin{pmatrix}1\\1\\-1\\-1\end{pmatrix}$ $\vec{w}_3 = \begin{pmatrix}1\\-1\\0\\0\end{pmatrix}$ $\vec{w}_4 = \begin{pmatrix}0\\0\\1\\-1\end{pmatrix}$

for any given vectors $\vec{v}$, we can find $\vec{c}$ such that $\vec{v} = \begin{pmatrix}\vec{w}_1 & \vec{w}_2 & \vec{w}_3 & \vec{w}_4\end{pmatrix}\vec{c}$

通过右例方法, 可以快速求出$\vec{c}$. 其中, c_1, c_2描述了general shape, c_3, c_4描述了具体细节



$c_3=1$

$c_4=2$

假设我们要进行压缩, 如果直接压缩$\vec{v}$便会失去成片的数据; 但如果加工$\vec{c}$, 删减细节, 仍可继续使用整体数据——因为我们本来也分辨不清细节

$c_1=4$

$c_2=1$

input $\vec{v}$ $\xrightarrow{\text{lossless}}$ coefficient $\vec{c}$ $\xrightarrow{\text{lossy}}$ compressed $\hat{c}$ $\xrightarrow{\text{reconstruct}}$ compressed $\hat{v}$

similar use can be seen in Fourier Transformation

- Change the input and output bases: transform相同 bases不同

一般来说，如果对于一个matrix A，我们想让input bases变为W，则AW

如果我们想让output bases变成S，则$S^{-1}A$

S^{-1}	A	W
把standard form变成$\vec{s}$'s form	$\vec{w}$'s form计算可方便，但结果standard form	确保输入是$\vec{w}$'s form

A就是一个单纯的matrix，有些时候我们change bases单纯因为这样简便计算，如：$S^{-1}AS$ (eigenvectors)；U^TAV (SVD)（因为我们想让结果diagonal）

* A与$B = S^{-1}AW$描述了同一个变化，但A是standard form，B是input $\vec{w}$'s form，output $\vec{s}$'s form

- polar decomposition

every real square matrix can be factored into $A = QH = KQ$, where Q is orthogonal and H and K are symmetric positive semidefinite

if A is invertible, H is positive definite

proof: $A = U\Sigma V^T = (UV^T)(V\Sigma V^T) = QH$

$= (U\Sigma U^T)(UV^T) = KQ$

[注意：(1) the product of orthogonal matrix is also orthogonal matrix

(2) H is positive semidefinite because its eigenvalues are in Σ, which is non-negative (eigenvalue may be 0)

(3) $H^2 = V\Sigma^2V^T = V\Sigma^T\Sigma V^T = A^TA$

$K^2 = U\Sigma^2U^T = U\Sigma\Sigma^TU^T = AA^T$

(4) in mechanics, polar decomposition seperates the rotation (in Q) from sketching (in H)

- pseudoinverse A^+

其满足几个条件：$A^+AA^+ = A^+$, $AA^+A = A$，但

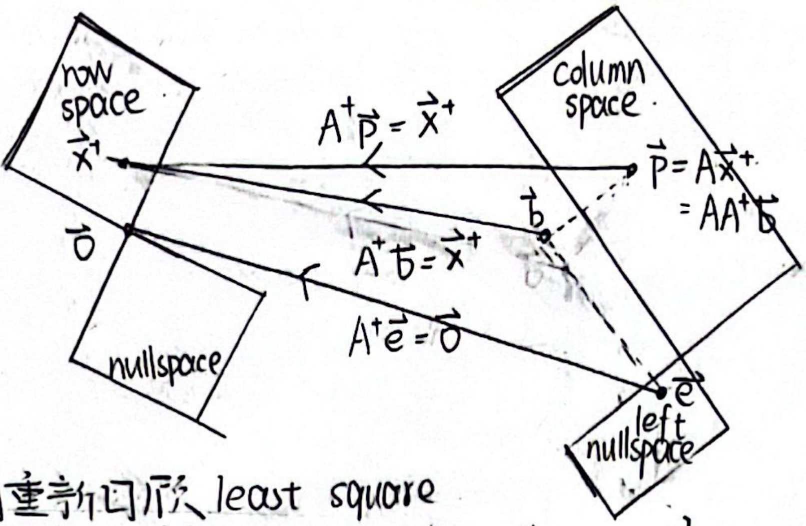
但是A^+ is not neccessary A^{-1}; if A^{-1} exists, $A^+ = A^{-1}$

$$\underbrace{A^+}_{n\times m} = V\Sigma^+U^T = \underbrace{\begin{bmatrix} \vec{v}_1 \cdots \vec{v}_r \cdots \vec{v}_n \end{bmatrix}}_{n\times n} \underbrace{\begin{bmatrix} \sigma_1^{-1} & & & & \\ & \ddots & & & \\ & & \sigma_r^{-1} & & \\ & & & 0 & \\ & & & & \ddots \ 0 \end{bmatrix}}_{n\times m} \underbrace{\begin{bmatrix} \vec{u}_1^T \\ \vdots \\ \vec{u}_r^T \\ \vdots \\ \vec{u}_m^T \end{bmatrix}}_{m\times m}$$

$A^+\vec{u}_i = \frac{1}{\sigma_i}\vec{v}_i$ for $i \le r$ and $A^+\vec{u} = \vec{0}$ for $i > r$ since $A^+U = V\Sigma^+$

$\vec{u}_1 \cdots \vec{u}_r$ come from the column space of A, A^+ sends them back to $\vec{v}_1 \cdots \vec{v}_r$ in row space of A. $\vec{u}_{r+1} \cdots \vec{u}_n$ come from left nullspace of A, A^+ sends them to $\vec{0}$

$\Sigma^+\Sigma = \begin{bmatrix} I_{r\times r} & \\ & \end{bmatrix}$ [整体$n\times n$] $\Sigma\Sigma^+ = \begin{bmatrix} I_{r\times r} & \\ & \end{bmatrix}$ [整体$m\times m$]



A^+A = projection matrix onto row space of A

AA^+ = projection matrix on column space of A

- 我们重新回顾 least square

for unsolvable matrix $A\vec{x}=\vec{b}$, we choose $A^TA\hat{x}=A^T\vec{b}$ to approximate it if A^TA is invertible

but what if A^TA is not invertible? then there will be many solution for $A^TA\hat{x}=A^T\vec{b}$ but $\vec{x}^+=A^+\vec{b}$ will be the best since $A^TAA^+\vec{b}=A^T\vec{p}$ $=A^T\vec{b}$ since left nullspace bases · column space bases $=0$

the shortest least squares solution to $A\vec{x}=\vec{b}$ is $\vec{x}^+=A^+\vec{b}$

X. application

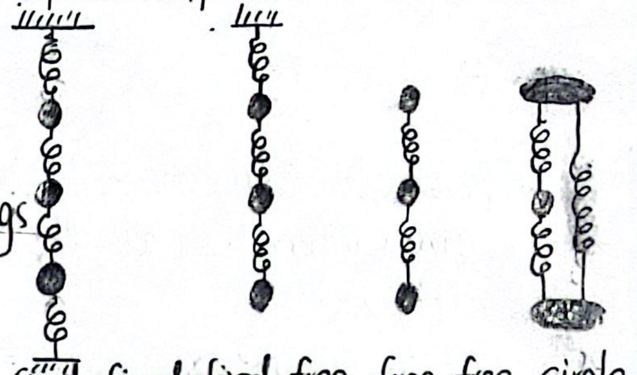
- 在engineering中linear algebra的运用：

direct way: the problem has only a finite number of piece, any many laws indicate linear relationship, so we may use matrix to connect them. (example: stiffness matrix for spring system)

indirect way: continuous system. we can find accurate solutions by approximation. (example: finite difference method)

[注意: stiffness matrix

$\vec{u}$: movements of n masses
$\vec{e}$: elongations of m springs
C: elasticity coefficients of m springs
$\vec{y}$: internal force in m springs
f: external forces on n masses



fixed-fixed fixed-free free-free circle

我们分三步计算

$$\begin{tikzcd} \boxed{u} \arrow[d, "A"'] & \boxed{f} \\ \boxed{e} \arrow[r, "C"'] & \boxed{y} \arrow[u, "A^T"'] \end{tikzcd}$$

A: m×n
C: m×m
A^T: n×m

先看 fixed-fixed system, if there are 3 masses and 4 springs

$$\begin{aligned} e_1 &= u_1 \\ e_2 &= u_2-u_1 \\ e_3 &= u_3-u_2 \\ e_4 &= \quad -u_3 \end{aligned} \Rightarrow \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{pmatrix}\begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} = \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{pmatrix}, \quad A\vec{u}=\vec{e}$$

$y_1 = c_1 e_1$
$y_2 = c_2 e_2$
$y_3 = c_3 e_3$
$y_4 = c_4 e_4$

$$\Rightarrow \begin{pmatrix} c_1 & 0 & 0 & 0 \\ 0 & c_2 & 0 & 0 \\ 0 & 0 & c_3 & 0 \\ 0 & 0 & 0 & c_4 \end{pmatrix} \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix}, \quad C\vec{e} = \vec{y}$$

$f_1 = y_1 - y_2$
$f_2 = y_2 - y_3$
$f_3 = y_3 - y_4$

$$\Rightarrow \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} f_1 \\ f_2 \\ f_3 \end{pmatrix}, \quad A^T\vec{y} = \vec{f}$$

so $A^TCA\vec{u} = \vec{f}$, $K\vec{u} = \vec{f}$

some properties of K: (1) tridiagonal because mass 1 & 3 aren't connec
(2) symmetric, because $C^T = C$
(3) positive definite, since $c_i > 0$ and A full rank
(4) K^{-1} is full matrix with all positive entries

notice that $\vec{u} = K^{-1}\vec{f}$ not necessarily means $\vec{u} = A^{-1}C^{-1}(A^T)^{-1}\vec{f}$

for fixed-free system.

$$K = A^TCA = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} c_1 & 0 & 0 \\ 0 & c_2 & 0 \\ 0 & 0 & c_3 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix}$$

for free-free system:

$$K = A^TCA = \begin{pmatrix} -1 & 0 \\ 1 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} c_2 & 0 \\ 0 & c_3 \end{pmatrix} \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix}$$

since K is singular here, $f_1 + f_2 + f_3 = 0$ for real $\vec{u}$

for circle system.

$$K = A^TCA = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} c_1 & 0 & 0 \\ 0 & c_2 & 0 \\ 0 & 0 & c_3 \end{pmatrix} \begin{pmatrix} 1 & 0 & -1 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix}.$$

- 在 differential equation 中 linear algebra 的运用

for a given interval, we divide it into $(n+1)$ equal piece, and each piece has width Δx

let $u_0 = a$, $u_1 = a + \Delta x$, ..., $u_n = a + n\Delta x$, $u_{n+1} = b$,

if $-\frac{d^2u}{dx^2} = 1$, $u(0) = u(1) = 0$, to find $u(x)$. let $n = 3$, $\Delta x = \frac{1}{4}$

$$\text{so } -\frac{d^2u}{dx^2} = 1 \Rightarrow \frac{1}{(\Delta x)^2} \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{bmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

$$A^TA\vec{u} = (\Delta x)^2\vec{f}$$

A = forward difference, A^T = − backward difference, A^TA = − centere

- 在 graph 与电路中 linear algebra 的运用

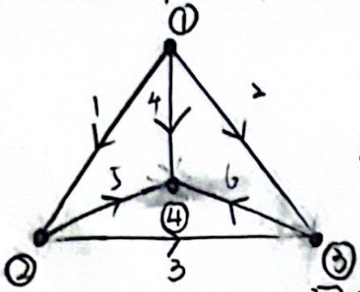
对于每一个graph，具本质上有限是node与edge的组合

利用incidence matrix，我们可以找到matrix与(directed) graph的biject

incidence matrix的每一行代表一条edge，只包含一个"-1"(起点)和一个"1"(终点

incidence matrix的每一列代表一个node

every entry of an incidence matrix is 0 or 1 or -1. this continue to hold during elimination. all pivots and multipliers are ±1. therefore both factors in A=LU contain 0, -1, 1, so does the nullspace matrix.

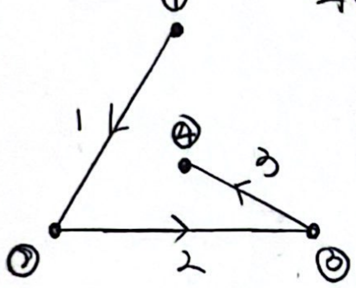


图(a)

$$
A=\begin{array}{c}
\text{node} \\
\begin{array}{cccc} ① & ② & ③ & ④ \end{array} \\
\begin{pmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 \\ -1 & 0 & 0 & 1 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 1 \end{pmatrix}
\end{array}
\begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array}
\text{ edge}
$$

the graph (a) is completed because every pair of node is connected by edges

the graph (b) is a tree since there aren't closed loops.



$$
B=\begin{array}{c}
\text{node} \\
\begin{array}{cccc} ① & ② & ③ & ④ \end{array} \\
\begin{pmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{pmatrix}
\end{array}
\begin{array}{c} 1 \\ 2 \\ 3 \end{array}
\text{ edge}
$$

for n nodes, there are maximum $[\frac{1}{2}n(n-1)]$ edges (completed) or $(n-1)$ edges (tree)

rows of B matches nonzero rows U, where A=LU: elimination reduces every graph to a tree: when two edges share a node. elimination produces the "shortcut edge" without that node. if graph already have this "shortcut edge", elimation gives a row of zeros: rows are dependent once form a loop: because we can go to the same destination (from same starting point) (get same result) through different ways (through different combinations)

事实上，four subspace在incidence matrix中很有用：想象这是一个电路 the unknown $\vec{x}=\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}$ represents potentials / voltage at node

(1) nullspace: $A\vec{x}=\vec{0} \Rightarrow U\vec{x}=\vec{0} \Rightarrow B\vec{x}=\vec{0}$

$$
\begin{pmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}=\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \Rightarrow \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}=\begin{pmatrix} C \\ C \\ C \\ C \end{pmatrix}
$$

shows combination of voltage such that no

(2) row space: if it is perpendicular to (1,1,1,1) (nullspace)

(3) column space: if we add difference around a closed loop in the graph we will have 0.
so the component of $A\vec{x}$ add to zero around every loop.
this is Kirchhoff's voltage law

[注意: on each edge, flow with arrow is positive, flow in opposi direction is negative]

(4) left nullspace: flow in equals to flow out for each node. flow around a loop can be added up without breaking the law
this is Kirchhoff's current law.

summary: for a (m×n) incidence matrix describing network.
- n nodes, m edge
- row/column space is rank (n-1) (a tree)
- nullspace: $\begin{pmatrix} c \\ c \\ \vdots \\ c \end{pmatrix}$ as the only basis.
- voltage law: the components of $A\vec{x}$ add to 0 around every loop
- current law: $A^T\vec{y} = \vec{0}$ is solved by loop currents. there are $(m-n+1)$ independent loops

Euler formula: $V - E + F$
= (number of node) − (number of edges) + (number of independent loops)
$= n - m + (m-n+1) = 1$. for all planar graph

in reality. current along edge $(\vec{y}_i)$ = conductance (c_i) × potential difference $(A\vec{x}_i)$

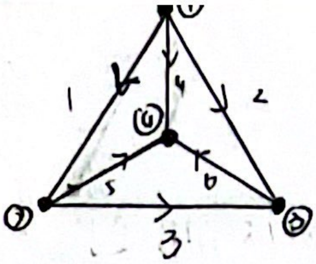
so $\vec{y} = CA\vec{x}$, where $C = \begin{pmatrix} c_1 & & \\ & \ddots & \\ & & c_m \end{pmatrix}$. $A^T\vec{y} = \vec{f}$. where $\vec{f}$ indicates the source from outside

$A^TCA\vec{x} = \vec{f}$. $K\vec{x} = \vec{f}$ where $K = A^TCA$ [balance 内外 source

注意 K 不是 full rank 不能 invertible because $\begin{pmatrix} c \\ c \\ \vdots \\ c \end{pmatrix}$ exists in null space

hence we assume a suitable node has voltage 0. for instance

$c_i = 1$ for all i, with source S in node ①.



$$\begin{pmatrix} -1 & -1 & 0 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 & -1 & 0 \\ 0 & 1 & 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 \\ -1 & 0 & 0 & 1 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 1 \end{pmatrix} = \begin{pmatrix} 3 & -1 & -1 & -1 \\ -1 & 3 & -1 & -1 \\ -1 & -1 & 3 & -1 \\ -1 & -1 & -1 & 3 \end{pmatrix}$$

$$\underbrace{\begin{pmatrix} 3 & -1 & -1 & -1 \\ -1 & 3 & -1 & -1 \\ -1 & -1 & 3 & -1 \\ -1 & -1 & -1 & 3 \end{pmatrix}}_{\text{invertible}} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} S \\ 0 \\ 0 \\ 0 \end{pmatrix} \Longrightarrow \underbrace{\begin{pmatrix} 3 & -1 & -1 \\ -1 & 3 & -1 \\ -1 & -1 & 3 \end{pmatrix}}_{\text{vertible,}} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} S \\ 0 \\ 0 \end{pmatrix}$$

answer is $\vec{x}_n + \vec{x}_p$.

by letting $x_4 = 0$

answer is $\vec{x}_n$ now

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} \frac{S}{2} \\ \frac{S}{4} \\ \frac{S}{4} \\ 0 \end{pmatrix}, \quad \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{pmatrix} = \begin{pmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 \\ -1 & 0 & 0 & 1 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} \frac{S}{2} \\ \frac{S}{4} \\ \frac{S}{4} \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{S}{4} \\ \frac{S}{4} \\ 0 \\ \frac{S}{2} \\ \frac{S}{4} \\ \frac{S}{4} \end{pmatrix}$$

lastly, we may recall A^TA from least squares

nature distributes the current to minimise heat loss.

• 在概率、人口统计、经济上linear algebra的应用

首先介绍 Perron - Frobenius Theorem:

for $A > 0$ (meaning all elements are positive), all number in $A\vec{x} = \lambda_{max}\vec{x}$ are strictly positive.

proof:

for all positive numbers t such that $A\vec{x} \geq t\vec{x}$ for some nonnegative vectors $\vec{x}$ (except for $\vec{0}$, $\vec{x} \neq \vec{0}$), $A\vec{x} = t_{max}\vec{x}$

if not, $\begin{cases} A\vec{x} \geq t_{max}\vec{x} \text{ (这里指有的row对应相同, 有的row LHS更大)} \\ A\vec{x} \neq t_{max}\vec{x} \end{cases}$

$\therefore \begin{cases} A\vec{x} - t_{max}\vec{x} \geq \vec{0} \\ A\vec{x} - t_{max}\vec{x} \neq 0 \end{cases}$

$\therefore A(A\vec{x} - t_{max}\vec{x}) > \vec{0}$ (至少一项为 positive × positive)

$\therefore A^2\vec{x} > t_{max}A\vec{x}$, $A(A\vec{x}) > t_{max}(A\vec{x})$,

$A\vec{y} > t_{max}\vec{y}$ where $\vec{y} = A\vec{x}$, impossible since no t larger than t_{max}

by contradiction, $A\vec{x} = t_{max}\vec{x}$. since LHS is positive, $\vec{x}$ is positive

we now want to prove that no eigenvalues can be larger than t_{max}.

suppose $A\vec{z} = \lambda\vec{z}$, since λ and $\vec{z}$ may involve negative or complex element

we take absolute values

$A|\underline{z}| \geq |Az| = \overline{|\lambda|}|z|$

non-negative　　positive

因此 $|\lambda|$ is one of possible t, $t_{max} = \lambda_{max}$.

现在我们可以根据 λ_{max} 的不同来分类不同 positive matrix 的用途

- Markov matrix: $\lambda_{max} = 1$ (see below)
- population matrix: $\lambda_{max} > 1$ (we are not going to extinct, right? M^k, 当 k 很大时, 结果与 λ_{max} 有很大关系)
- consumption matrix: $\lambda_{max} < 1$. (see below)

(1) Markov matrix 马尔可夫矩阵, random walk 概率的一个展示方式

it has several properties:

- every entry of A is positive
- every column of A adds up to 1.

so:

- multiplying a non-negative $\vec{u}_0$ by A produces a non-negative $\vec{u}_1 = A\vec{u}_0$
- if the components of $\vec{u}_0$ add up to 1, so do the components of $\vec{u}_1 = A\vec{u}_0$

proof: $\because [1\,1 \cdots 1]\vec{u}_0 = 1$, $[1\,1 \cdots 1]A = [1\,1 \cdots 1]$

$\therefore [1\,1 \cdots 1]A\vec{u}_0 = [1\,1 \cdots 1]\vec{u}_1 = [1\,1 \cdots 1]\vec{u}_0 = 1$.

- $A^{\infty}u_0$ approximate to the unique vector u_{∞}, u_{∞} is an eigenvector with eigenvalue of 1

proof: $\lambda = 1$ will definitely be one of its eigenvalue

since $A - I$ makes the sum of each column be 0

add up all the rows now, and the final sum is 0

meaning $A - I$ is singular, so $\lambda_1 = 1$

no eigenvalues are $|\lambda_i| > 1$, with such properties, the power of A^k will grow, but A^k is still a Markov matrix with non-negative entries and columns adding up to 1 - soon there's no room for enlargement.

(2) consumption matrix: tell how much of each input goes into unit outp

$$\begin{pmatrix} \text{M output} \\ \text{N output} \end{pmatrix} = \begin{pmatrix} \text{M needed for unit M output} & \text{N needed for unit M output} \\ \text{M needed for unit N output} & \text{N needed for unit N output} \end{pmatrix} \begin{pmatrix} \text{M input} \\ \text{N input} \end{pmatrix}$$

如果 demand 为 $\vec{y}$, 我们要保证: input 在生产完别的产品后, 依旧能满足市场

$\vec{p} - A\vec{p} = \vec{y}$　　$\vec{p} = (I - A)^{-1}\vec{y}$

→ → → → for a productive economy $(I-A)^{-1}$ should be non-negative satisfy all possible combination of demand

$$\begin{cases} \text{if } \lambda_{max} > 1, & (I-A)^{-1} \text{ has negative entries} \\ \text{if } \lambda_{max} = 1, & (I-A)^{-1} \text{ fails to exist} \\ \text{if } \lambda_{max} < 1, & (I-A)^{-1} \geq 0 \end{cases}$$

besides, for $|\lambda_i| < 1$ for all i:

$(I-A)^{-1} = I + A + A^2 + A^3 + \cdots$

proof: let $S = I + A + A^2 + A^3 + \cdots$, for A, $|\lambda_i| < 1$ for all i

$S - AS = I$ ($A^k \longrightarrow \begin{pmatrix} 0 & \\ & \ddots \end{pmatrix}$ when $k \longrightarrow \infty$)

$(I-A)^{-1} = S$

- 在linear programming中linear algebra的运用

一般的format为: $A\vec{x} = \vec{b}$, $\vec{x} \geq \vec{0}$, minimise $\vec{c}^T\vec{x}$, 有几种方法:

(1) 我们将$A\vec{x} = \vec{b}$在图中表示出来, 有可能bounded or unbounded

我们主要讨论bounded的情况(毕竟unbounded时$\vec{c}^T\vec{x}^*$最小值可能为负无限)

consider $\vec{c}^T\vec{x} = K$, where K is a constant, so $\vec{c}^T\vec{x} = K$ is also a plane

when $A\vec{x} = \vec{b}$ and $\vec{c}^T\vec{x} = K$ only intersect at one point, at the first time when K arises gradually, $\vec{c}^T\vec{x}$ is minimised, and $\vec{x}$ must be at the corner of $A\vec{x} = \vec{b}$ (similar things happen if we want to maximise $\vec{c}^T\vec{x}$)

因此如果A为rank r, 则最多有r个nonzero elements for $\vec{x}$. 我们可以先选一corner, 并检查这个选择对不对.

我们通过simplex method检验: 我们对于任意一个zero element, change it into a suitable positive value, so in such way, we find $A\vec{x}_i = \vec{b}$. then calculate $r_i = \vec{c}^T\vec{x}_i - \vec{c}^T\vec{x}$, which is called reduced cost. for the smallest reduced cost, the corresponding changed initially-zero element (entering element) should be positive, and the first initially-positive element going to zero (leaving element) should be zero forever

可以理解为被更高效的选择代替了, 如果reduced cost都为正, 则得到了

(2) in linear programming, problems come in pairs: maximise $\vec{b}^T\vec{y}$ given $A^T\vec{y} \leq$

if either problem has a best solution then so does the other,

minimum $\vec{c}^T\vec{x}$ = maximum $\vec{b}^T\vec{y}$

proof: $\vec{b}^T\vec{y} = (A\vec{x})^T\vec{y} = \vec{x}^T A^T \vec{y} \leq \vec{x}^T\vec{c}$

because $\vec{x} \geq 0$, $\vec{s} = \vec{c} - A^T\vec{y} \geq 0$, $\vec{x}^T\vec{s} \geq 0$, so $x^T A^T \vec{y} \leq \vec{x}^T\vec{c}$

optimal solution has $x_j = 0$ or $s_j = 0$ for all j

(3) simplex method尝试所有corner, primal & dual problem可以相互转化

interior point method试图通过approximation直接得到答案

将原问题转化为barrier problem

minimise $\vec{c}^T\vec{x} - \theta(\log x_1 + \log x_2 + \cdots + \log x_n)$ with $A\vec{x}=\vec{b}$
in this way, $x_i > 0$ for sure. as $\theta \to 0$, we gradually get $\vec{x}$ we want (which is on the edge, so it is reasonable to let $x>0$ when $\theta \neq 0$
在$A\vec{x}=\vec{b}$中有m个constraints, which have m Lagrange multipliers $y_1, y_2, \cdots y_m$
因此 $L(\vec{x},\vec{y},\theta) = \vec{c}^T\vec{x} - \theta(\sum \log x_i) - \vec{y}^T(A\vec{x}-\vec{b})$
求极值,则 $\frac{\partial L}{\partial \theta}, \frac{\partial L}{\partial x_i}, \frac{\partial L}{\partial y_i}$ 应均为0 (Lagrange Theorem)

$$\frac{\partial L}{\partial y_i} = A\vec{x} - \vec{b} = 0 \quad \text{(for sure)}$$

$$\frac{\partial L}{\partial x_i} = c_i - \frac{\theta}{x_i} - (A^T\vec{y})_i = 0 \implies x_i s_i = \theta \quad (\text{when } \theta \to 0,\ x_i s_i = 0)$$

通过x_i求y_i并看合不合适,若不合适则迭代,试图精确(当然,θ已假设
by Newton's method, we take a step $\Delta\vec{x}, \Delta\vec{y}, \Delta\vec{s}$
in such way, by ignoring $\Delta\vec{x}\Delta\vec{s}$, we convert nonlinear equations into linear one

$$\begin{cases} A\Delta\vec{x} = \vec{0} \\ A^T\Delta\vec{y} + \Delta\vec{s} = \vec{0} \\ s_i\Delta x_i + x_i\Delta s_i = \theta - x_i s_i \end{cases}$$

随着θ假设的越来越小, approximation越精确

- 在function中linear algebra的运用

what if a vector is in an infinite-dimensional space? we can express it
(1) $\vec{v} = (v_1, v_2, v_3, \cdots)^T$ (2) vector is a function $f(x)$
for each of them, dot product and vector length both work, and they are connected by Fourier Series

form	dot product	vector length
(1)	$\vec{v}\cdot\vec{w} = v_1w_1 + v_2w_2 + \cdots$ if $\lvert\vec{v}\rvert, \lvert\vec{w}\rvert$ both in Hilbert space, $v\cdot w$ converge, since $\lvert\vec{v}\cdot\vec{w}\rvert \le \lvert\vec{v}\rvert\lvert\vec{w}\rvert$ $\vec{v}$ and $\vec{w}$ are orthogonal if $\vec{v}\cdot\vec{w}=0$	it is in our infinite-dimensional "Hilbert space" if and only if $\vec{v}\cdot\vec{v} / \lVert\vec{v}\rVert$ is finite $\lvert\vec{v}\rvert^2 = v_1^2 + v_2^2 + \cdots$
(2) [over (a,b)]	$\int_a^b f(x)\cdot g(x)\,dx$ $f(x)$ and $g(x)$ are orthogonal if $\int_a^b f(x)g(x)dx = 0$.	$\lvert f\rvert = \int_a^b (f(x))^2 dx$

注意到,根据定义,在$[0, 2\pi]$上, $1, \sin x, \cos x, \sin 2x, \cos 2x, \cdots$ 等function均orthogonal, $\frac{1}{\sqrt{2\pi}}, \frac{\sin x}{\sqrt{\pi}}, \frac{\cos x}{\sqrt{\pi}}, \frac{\sin 2x}{\sqrt{\pi}}, \frac{\cos 2x}{\sqrt{\pi}}, \cdots$ 均为orthonormal unit vector,
因此可以将它们视作bases

[illegible] at covariance matrix

$$f(x) = \frac{A_0}{\sqrt{2\pi}} + \frac{A_1}{\sqrt{\pi}}\cos x + \frac{B_1}{\sqrt{\pi}}\sin x + \frac{A_2}{\sqrt{\pi}}\cos 2x + \frac{B_2}{\sqrt{\pi}}\sin 2x + \cdots$$

in this way, function length = vector length $= \sqrt{A_0^2 + A_1^2 + B_1^2 + A_2^2 + B_2^2 + \cdots}$

the function space contains $f(x)$ exactly when the Hilbert space contai the vector $\vec{v} = (A_0, A_1, B_1, A_2, B_2, \ldots)$ of Fourier coefficient. both $f(x)$ and $\vec{v}$ have finite length

当然为了简便，我还是喜欢写

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

$$|f(x)|^2 = \int_0^{2\pi} (a_0 + a_1\cos x + b_1\sin x + a_2\cos 2x + b_2\sin 2x + \cdots)^2 dx$$

$$= \int_0^{2\pi} a_0^2 + a_1^2\cos^2 x + b_1^2\sin^2 x + a_2^2\cos^2 2x + b_2^2\sin^2 2x + \cdots dx \text{ [orthogonal}$$

$$= 2\pi a_0^2 + \pi(a_1^2 + b_1^2 + a_2^2 + b_2^2 + \cdots)$$

想求 a_i, b_i, a_0 (for $i \geq 1$ 可巧用 orthogonal

$$a_i = \frac{1}{\pi}\int_0^{2\pi} f(x)\cos ix\, dx \qquad b_i = \frac{1}{\pi}\int_0^{2\pi} f(x)\sin ix\, dx$$

$$a_0 = \frac{1}{2\pi}\int_0^{2\pi} f(x)\cdot 1\, dx$$

- 在 statistics 与 probability 中 linear algebra 的运用

(1) 以前我们求 least square 有一些 independent experiments, if $A\vec{x} = \vec{b}$ has no solutions, then find $A^TA\hat{x} = A^T\vec{b}$

但问题是 并不是所有的 experiment 都一样可靠，我们可以给 variance 小的数据 (from 多次测量 / 工具误差 / ...) 更大的 weight.

so $A^TA\hat{x} = A^T\vec{b}$ can be better turned into $(WA)^TWA\hat{x} = (WA)^TW\vec{b}$

so $A^TCA\hat{x} = A^TC\vec{b}$, $C = W^TW = \begin{pmatrix} \sigma_1^2 & & \\ & \sigma_2^2 & \\ & & \ddots \end{pmatrix}$ would be a better approximation for $A\vec{x} = \vec{b}$. because for $WA\vec{x} = W\vec{b}$, where $W = \begin{pmatrix} \sigma_1 & & \\ & \sigma_2 & \\ & & \ddots \end{pmatrix}$, has variances of 1 for all samples

(2) 协方差 $Cov(X,Y) = E[(X-\mu_x)(Y-\mu_y)] = E(e_x e_y)$

$$= E[XY - X\mu_y - Y\mu_x + \mu_x\mu_y]$$

$$= E(XY) - E(X)\mu_y - E(Y)\mu_x + \mu_x\mu_y$$

$$= E(XY) - \mu_x\mu_y$$

如果 X, Y independent, $Cov(X,Y) = 0$

如果 X, Y 正相关, $Cov(X,Y) > 0$

如果 X, Y 负相关, $Cov(X,Y) < 0$

现在我们引入 covariance matrix $\Sigma = \begin{pmatrix} E(e_1e_1) & E(e_1e_2) & \cdots \\ E(e_2e_1) & E(e_2e_2) & \cdots \\ & & \ddots \end{pmatrix}$, 其中 $E(e_ie_i) = \sigma_i^2$ 我们可以把 Σ 本身看成一个 data matrix 的 variance.

注意到 $E(e_ie_j) = E(e_je_i)$, 这是一个 symmetric matrix

我们可以改写出 multivariate normal:

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$: single variable; mean μ, variance σ^2

$$p(\vec{b}) = \frac{1}{(2\pi)^{\frac{m}{2}}\sqrt{|\Sigma|}} e^{-\frac{(\vec{b}-\vec{\mu})^T\Sigma^{-1}(\vec{b}-\vec{\mu})}{2}}, \text{ where } \vec{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}.$$

m variables, mean $\vec{\mu}$, covariance matrix Σ

let $\Sigma = Q\Lambda Q^T$, $\vec{b}-\vec{\mu} = Q\vec{c}$, $\vec{c}$ are statistically independent

$$\exp\left(-\frac{(\vec{b}-\vec{\mu})^T\Sigma(\vec{b}-\vec{\mu})}{2}\right) = \exp\left(-\frac{\vec{c}^T\Lambda^{-1}\vec{c}}{2}\right)$$

$$= \left(e^{-\frac{c_1^2}{2\lambda_1}}\right)\left(e^{-\frac{c_2^2}{2\lambda_2}}\right)\cdots\left(e^{-\frac{c_m^2}{2\lambda_m}}\right).$$

let $\vec{b}-\vec{\mu} = \sqrt{\Lambda}Q\vec{z}$ we could even find out the standard form.

$$p(\vec{z}) = \frac{1}{(2\pi)^{\frac{m}{2}}} e^{-\frac{\vec{z}^T\vec{z}}{2}} d\vec{z}$$

如果我们对 $p(\vec{b})$ 全域做积分:

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} p(\vec{b})\,d\vec{b} = \prod_{i=1}^{m}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\lambda_i}} e^{-\frac{c_i^2}{2\lambda_i}} dc_i = 1$$

另一个有趣的结论，根据协方差矩阵的定义

$$\int \vec{b}\vec{b}^T p(\vec{b})\,d\vec{b} = \Sigma$$

3) 我们看完(2)之后来回顾(1)：本质上 $WA\vec{x} = W\vec{b}$ 也是在找standard form

因此，本质上 $C = W^TW = \Sigma^{-1}$，不过如果用 Σ^{-1}，我们便可以让experiment变dependent

so $A^T\Sigma^{-1}A\hat{x} = A^T\Sigma^{-1}\vec{b}$, where Σ is covariance matrix, would be the best approximation for $A\vec{x} = \vec{b}$, allowing dependent experiment

高斯认为这是 best linear unbaised estimate $\hat{x}$，因为:

① unbiased: since $E(\hat{x}) = \vec{x}$

$\because \hat{x} = (A^T\Sigma^{-1}A)^{-1}A^T\Sigma^{-1}\vec{b}$, $\vec{b} = A\vec{x} + \vec{e}$

$$\therefore \hat{x} = (A^T\Sigma^{-1}A)^{-1}A^T\Sigma^{-1}A\vec{x} + (A^T\Sigma^{-1}A)^{-1}A^T\Sigma^{-1}\vec{e}$$

$$= \vec{x} + (A^T\Sigma^{-1}A)^{-1}A^T\Sigma^{-1}\vec{e}$$

$\because E(\hat{x}) = \vec{x} + (A^T\Sigma^{-1}A)^{-1}A^T\Sigma^{-1}E(\vec{e})$, $E(\vec{e}) = \vec{0}$

$\therefore E(\hat{x}) = \vec{x}$

② best: covariance of $\hat{x}$ doesn't depend on $\vec{b}$'s in real experiment

$$P = E[(\vec{x}-\hat{x})(\vec{x}-\hat{x})^T]$$

$$= E\left[(-(A^T\Sigma^{-1}A)^{-1}A^T\Sigma^{-1}\vec{e})(-\underbrace{(A^T\Sigma^{-1}A)}_{\text{symmetric}}{}^{-1}A^T\Sigma^{-1}e)^T\right]$$

$$= (A^T\Sigma^{-1}A)^{-1}A^T\Sigma^{-1}E(\vec{e}\vec{e}^T)\Sigma^{-1}A(A^T\Sigma^{-1}A)^{-1}$$

$$= (A^T\Sigma^{-1}A)^{-1}A^T\Sigma^{-1}\Sigma\Sigma^{-1}A(A^T\Sigma^{-1}A)^{-1}$$

$$= (A^T\Sigma^{-1}A)^{-1}(A^T\Sigma^{-1}A)(A^T\Sigma^{-1}A)^{-1}$$

$$= (A^T\Sigma^{-1}A^{-1})^{-1}$$

(5) further knowledge of covariance matrix

★ $\Sigma = \sum\sum_{\text{all } i,j} p_{ij}\Sigma_{ij} = \sum\sum_{\text{all } i,j} p_{ij}\begin{pmatrix}(x_i-m_1)^2 & (x_i-m_1)(y_j-m_2) & \cdots \\ (x_i-m_1)(y_j-m_2) & (y_j-m_2)^2 & \cdots \\ \vdots & & \ddots\end{pmatrix}$

in fact, $\Sigma_{ij} = \vec{U}\vec{U}^T$, where $\vec{U} = \begin{pmatrix} x_i-m_1 \\ y_j-m_2 \\ \vdots \end{pmatrix}$

thus, the covariance matrix is at least semidefinite:

proof①: $\vec{a}^T\Sigma\vec{a} = \sum\sum_{\text{all } i,j} p_{ij}\vec{a}^T\vec{U}\vec{U}^T\vec{a} = \sum\sum_{\text{all } i,j} p_{ij}|\vec{U}^T\vec{a}|^2 \geq 0$

proof②: $\Sigma = E[(\vec{X}-\vec{\mu})(\vec{X}-\vec{\mu})^T]$ where $\vec{\mu} = E(\vec{X})$, and $\vec{x}$ is a vector with all dates from each experiment

consider linear combination $\vec{c}^T\vec{X} = c_1X_1 + \cdots + c_nX_n$

$Var(\vec{c}^T\vec{X}) = E[(c^T\vec{X} - c^T\vec{\mu})(c^T\vec{X} - c^T\vec{\mu})^T]$

$= \vec{c}^T E[(\vec{X}-\vec{\mu})(\vec{X}-\vec{\mu})^T]\vec{c} = \vec{c}^T\Sigma\vec{c} \geq 0$

since variance is non-negative

in fact, the covariance matrix is positive definite unless the experiments dependent (毕竟不能一直让 $\vec{U}^T\vec{a}=0$ 吧,除非 $\vec{a}$ 方向固定 即各个 experiment 之间有倍数关系)

★ diagonalising the covariance matrix means finding n independent experiments as combination of the original n experiments

★ $Var(x+y) = \sum\sum_{\text{all } i,j} p_{ij}(x_i+y_j-m_x-m_y)^2$

$= \sum\sum_{\text{all } i,j} p_{ij}(x_i-m_x)^2 + \sum\sum_{\text{all } i,j} p_{ij}(y_j-m_y)^2 + 2\sum\sum_{\text{all } i,j} p_{ij}(x_i-m_i)(y_j-m_y)$

$= \sigma_x^2 + \sigma_y^2 + 2\sigma_{xy}$

now for the main point: vector $\vec{X}$ can have n components from n experiments. those experiments will have an $n \times n$ covariance matrix $\Sigma_{\vec{x}}$. $K \times n$ matrix A makes $A\vec{X}$ a combination of M components

$\Sigma_{A\vec{X}} = A\Sigma_{\vec{X}}A^T$

★ correlation $\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y} = Cov(\frac{X}{\sigma_x}, \frac{Y}{\sigma_y})$, ρ_{XY} is Pearson coefficient

this shows the relationship between X and Y.

since $\sigma_{XY}^2 \leq \sigma_X^2\sigma_Y^2$, $\rho_{XY}^2 \leq 1$

$\begin{cases} \text{independent} & \rho_{XY}=0 \\ \text{positively related} & \rho_{XY}=1 \\ \text{negatively related} & \rho_{XY}=-1 \end{cases}$

correlation matrix $R = \begin{pmatrix} 1 & \rho_{XY} \\ \rho_{XY} & 1 \end{pmatrix}$

$R = D\Sigma D$ for $D = \begin{pmatrix} \frac{1}{\sigma_1} & \\ & \frac{1}{\sigma_2} \end{pmatrix}$. if Σ is positive definite, so is R this can be used to know the reliability of the whole experiment: since often the question is not the best $\hat{x}$ for one particular set of measurement $\vec{b}$ (this is only one sample!)

这里 $\vec{x}$ 为population的对应特征，虽然我们不知道 $\vec{x}$ 的值，但是通过计算我们发现我们不需要知道 $\vec{x}$ 也可求 $E[(\hat{x}-\vec{x})(\hat{x}-\vec{x})^T]=(A^T\Sigma^{-1}A)^{-1}$

(4) Kalman Filter.

- Kalman Filter is an algorithm in dynamic least square: new measurements $\vec{b}_k$ keep coming, so best estimate $\hat{x}_k$ keeps changing.

  for example: if $A_0\vec{x}_0=\vec{b}_0$ is estimated by $\hat{x}_0$ through

  $$A_0^T\Sigma_0^{-1}A_0\hat{x}_0=A_0^T\Sigma^{-1}\vec{b}_0$$

  now we have: $\begin{pmatrix}A_0\\A_1\end{pmatrix}\vec{x}_1=\begin{pmatrix}\vec{b}_1\\\vec{b}_2\end{pmatrix}$

  $$\therefore (A_0^T\ A_1^T)\begin{pmatrix}\Sigma_0^{-1} & \\ & \Sigma_1^{-1}\end{pmatrix}\begin{pmatrix}A_0\\A_1\end{pmatrix}\hat{x}_1=(A_0^T\ A_1^T)\begin{pmatrix}\Sigma_0^{-1} & \\ & \Sigma_1^{-1}\end{pmatrix}\begin{pmatrix}\vec{b}_0\\\vec{b}_1\end{pmatrix}$$

  $\therefore \hat{x}_1=\hat{x}_0+K_1(\vec{b}_1-A_1\hat{x}_0)$, where $K_1=P_1A_1^T\Sigma_1^{-1}$,

  $P_1^{-1}=P_0^{-1}+A_1^T\Sigma_1^{-1}A$

(6) principal component analysis.

start by m properties from n samples for each row. substract the average so the sample mean are 0

we are looking for a combination of properties and samples for which datas provides the most information

找到方差最大的一组数据进行分析：通过SVD找最大的singular value即可

- 在computer graphic中linear algebra的运用

我们可以把基础图像分为以下几种：

translation: shift the origin to an another point $P_0=(x_0,y_0,z_0)$
rescaling by c_1, c_2, c_3 in different directions. may with P_0 as center
rotate around an axis. may with P_0 as center
projection: parallel / perspective projection. may with P_0 as a center
reflection: may with P_0 as center

我们主要讨论3D space，为此我们用homogenerous coordinates $(x,y,z,1)$表示坐标，用4x4 matrix表示操作：因为并不是所有操作都是linear的，3x3 matrix办不到. $(x,y,z,1)=(cx,cy,cz,c)$ for real $c\neq 0$

注意，计算机从左往右处理数据：row times matrix now.

(1) translation: $T=\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\x_0&y_0&z_0&1\end{pmatrix}$

(2) rescaling: $S=\begin{pmatrix}c_1&0&0&0\\0&c_2&0&0\\0&0&c_3&0\\0&0&0&1\end{pmatrix}$

如果不以origin为center，可以先将整个space translate，再rescale，最后再translate back，或者 $\begin{pmatrix}T(1,0,0)&0\\T(0,1,0)&0\\T(0,0,1)&0\\T(0,0,0)&1\end{pmatrix}$

(3) rotate: if we rotate around unit vector $a=(a_1,a_2,a_3)$

我们将$\vec{v}$分成两部分：$\vec{v}_{\parallel}$平行于$\vec{a}$方向，$\vec{v}_{\perp}$垂直于$\vec{a}$方向

$$\vec{v}_{\parallel}=(\vec{a}\cdot\vec{v})\vec{a},\quad \vec{v}_{\perp}=\vec{v}-(\vec{a}\cdot\vec{v})\vec{a}$$

$$\vec{v}'=\underbrace{\vec{v}_{\parallel}}_{\text{不变}}+\underbrace{\cos\theta\,\vec{v}_{\perp}+\sin\theta\cdot(\vec{a}\times\vec{v}_{\perp})}_{\text{旋转}}$$

$$=(\vec{a}\cdot\vec{v})\vec{a}+\cos\theta\left(\vec{v}-(\vec{a}\cdot\vec{v})\vec{a}\right)+\sin\theta\,(\vec{a}\times\vec{v})$$

$$=\cos\theta\,\vec{v}+(1-\cos\theta)(\vec{a}\cdot\vec{v})\vec{a}+\sin\theta(\vec{a}\times\vec{v})$$

$\therefore Q = \cos\theta\, I + (1-\cos\theta)\, \vec{a}\vec{a}^T - \sin\theta\, [\vec{a}]_\times$

叉积矩阵

$$= \cos\theta \cdot I + (1-\cos\theta)\begin{pmatrix} a_1^2 & a_1a_2 & a_1a_3 \\ a_1a_2 & a_2^2 & a_2a_3 \\ a_1a_3 & a_2a_3 & a_3^2 \end{pmatrix} - \sin\theta \begin{pmatrix} 0 & a_3 & -a_2 \\ -a_3 & 0 & a_1 \\ a_2 & -a_1 & 0 \end{pmatrix}$$

$$R = \begin{pmatrix} Q & 0 \\ 0 & 1 \end{pmatrix}$$

如果center不为origin，做法同translation

(4) projection: for parallel projection:

if we are going to project all thing into a plane with normal unit vector $\vec{n}$, the vectors in plane satisfying $\vec{n}^T\vec{v} = 0$

so usual projection is: $I - \vec{n}\vec{n}^T$, since

$$\begin{cases} (I - \vec{n}\vec{n}^T)\vec{n} = \vec{n} - \vec{n}(\vec{n}^T\vec{n}) = \vec{0} \\ (I - \vec{n}\vec{n}^T)\vec{v} = \vec{v} - \vec{n}(\vec{n}^T\vec{v}) = \vec{v} \end{cases}$$

$$P = \begin{pmatrix} I - \vec{n}\vec{n}^T & 0 \\ 0 & 1 \end{pmatrix}$$

如果center不为origin，做法同上

for perspective projection:

设投影平面为 $ax+by+cz=d$，中心为origin，我们将 $\vec{v}=(x,y,z)$ 投到该平面上

从原点出发的投影线为：$\vec{r}(t) = t\vec{v} = (tx, ty, tz)$

$\therefore\ a(tx)+b(ty)+c(tz) = d$ [看线上哪一点在面上]

$$t = \frac{d}{ax+by+cz}$$

因此，面上交点为 $\vec{v}' = \left(\frac{dx}{ax+by+cz}, \frac{dy}{ax+by+cz}, \frac{dz}{ax+by+cz}\right)$

通过homogenerous coordinate, perspective projection matrix:

$$P = \begin{pmatrix} d & 0 & 0 & 0 \\ 0 & d & 0 & 0 \\ 0 & 0 & d & 0 \\ a & b & c & 0 \end{pmatrix}$$

如果center不为origin，做法同上

(5) reflection: $R = \begin{pmatrix} I - 2\vec{n}\vec{n}^T & 0 \\ 0 & 1 \end{pmatrix}$,

where $\vec{n}$ 为symmetric plane的normal unit vector

如果center不为origin，做法同上

在cryptography(密码学)中linear algebra的运用

①基础思想就是在matrix上mod p. 当同余与矩阵结合，我们有以下问题：

(1) when can we solve $A\vec{x}=\vec{b} \pmod p$? or when does an inverse matrix (mod p) exist whenever the determinant is nonzero mod p?

(2) do we still have four subspaces $C(A)$, $N(A)$, $C(A^T)$, $N(A^T)$? are they still orthogonal in pairs?

问题(2)的答案是肯定的，但我们重点关注问题(1):

through cofactor formula: $A^{-1} \equiv (\det A)^{-1} C^T \pmod p$

thus, $\gcd(\det A, p)=1$, if A^{-1} exists for all $\det A \neq 0$, then p is prime.

Hill Cipher加密的基本过程：

A~Z分别代表0~25，整体mod 26. 选择n×n的一个矩阵E为加密矩阵

其中 $\gcd(\det E, 26)=1$；由此则存在解密矩阵 $D \equiv E^{-1} \pmod{26}$

对于其安全性，由E与n决定，可以多次重复此操作以多次加密。

① 我们了解下finite field有限域，如 $\mathbb{F}_p$，以及finite vector spaces. 如 $(\mathbb{F}_p)^n$, where p is prime, notice that for a subspace, $\vec{0}$ must be included.

其实通过定义新元素，a field with p^k elements can also exist, where p is a prime and k is a positive integer.

八. numerical linear algebra

- roundoff error & partial pivoting

计算机在计算时精度有限. 如果pivot上element绝对值很小，那么在后续的elimination可能会导致multiplier过大，导致结果精确度/误差大

因此：the k^{th} pivot is decided when we reach and search column k: choose the largest (absolute value) number in row k or below, then exchange its row with row k.

- 计算机计算 $A\vec{x}=\vec{b}$ by elimination instead of finding inverse

① 当A=LU简化时，如果A为n×n矩阵，则第一个pivot要消去剩下(n-1)行的第一列。那么每一个row减去第一个row都是n次操作…… repeat.

total operations (notice that there are n pivots) are:

$$\sum_{i=1}^{n} i(i-1) = \frac{n^3-n}{3}.$$

for large n, we can just say that the multiply-substract count for forwards elimination (A to U, producing L simultaneously) is $\frac{1}{3}n^3$.

once we have $LU\vec{x}=\vec{b}$, we can solve it quickly: we first look at the single pivot, which is 1 operation; then we look at the pivot whose row has two elements, which is 2 operations... repeat, total operations (notice that we have both $L\vec{c}=\vec{b}$ and $U\vec{x}=\vec{c}$) are

$$2\sum_{i=1}^{n} i = n^2+n.$$

for large n, we can just say that the count for forwards part and back substitutions is n^2

② 当计算 inverse 时, 我们先花 $\frac{1}{3}n^3$ operations 找到 $A=LU$

if $A^{-1}=(\vec{a}_1\ \vec{a}_2 \cdots \vec{a}_n)$, $I=(\vec{i}_1\ \vec{i}_2 \cdots \vec{i}_n)$, we have $LU\vec{a}_j=\vec{i}_j$

在 $L\vec{c}_j=\vec{i}_j$ 时, no work is required until we reach row j, so the forward part needs $\frac{1}{2}(n-j)^2$ operations. repeat, total operations are

$$\sum_{i=1}^{n}\frac{1}{2}(n-i)^2=\frac{1}{2}\sum_{i=1}^{n-1} i^2=\frac{1}{12}n(n-1)(2n-1)$$

for large n, we can just say that the count for forwards part is $\frac{1}{6}n^3$

从而在 $U\vec{a}_j=\vec{c}_j$, repeat, total operations are

$$n\cdot\sum_{i=1}^{n} i=\frac{n^3+n^2}{2}$$

for large n, we can just say that the count for back substitutions is $\frac{1}{2}n^3$

totally, for large n, we can just say that the count for finding inverse is $\frac{1}{3}n^3+\frac{1}{6}n^3+\frac{1}{2}n^3=n^3$

to find $\vec{x}=A^{-1}\vec{b}$, it requires n^2 operations for sure

③ 当计算 $A=QR$ 时, each multiplier is decided by a dot product, taking n operations. then there are n "multiply-substract" operations to remove from column k its projection along column $j<k$. repeat, total operations are:

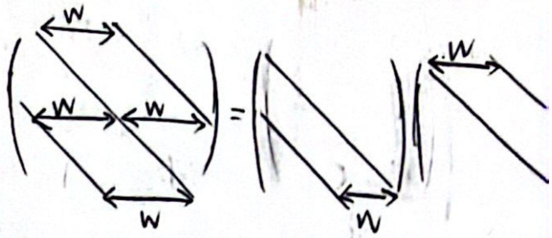
$$2\sum_{i=1}^{n} i^2=\frac{1}{3}(2n+1)(n+1)n$$

for large n, we can just say that the count for finding $A=QR$ is $\frac{2}{3}n^3$

④ 我们通过对比, 发现高斯 elimination operations 数量最少. 其实, 对于 banded matrix (非零元素集中在主对角线及其附近的几条对角线上), 如果其 half-bandwidth (非零元素所集中的那几条对角线的水平宽度的一半) 为 w, 则:

- 在计算 $A\vec{x}=\vec{b}$ 时，得到的 $\vec{x}$ 会有误差，这误差可能来自计算机/测量

$a_{ij}=0$ if $|i-j|>w$



$w=1$ for diagonal matrix, $w=2$ for tridiagonal, $w=n$ for dense

each stage of elimination is completed after $w(w-1)$ operations, and the band structure survives. so: elimination on a band matrix (A to LU) needs less that w^2n operations for large n. similarly, in this situation, solving $L\vec{c}=\vec{b}$ and $U\vec{x}=\vec{c}$ only needs $2wn$ operations

- 通过以下两种方法可以较快找到 $A=QR$.

① 因为 product of orthogonal matrix is also orthogonal，对于一个 3×3 matrix 我们可以通过一次次的 orthogonal rotation matrix to find R.

$A=\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$，$Q_{21}=\begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$ 在 xy plane 上的 rotation.

where $\cos\theta=\dfrac{a_{11}}{\sqrt{a_{11}^2+a_{21}^2}}$，$\sin\theta=\dfrac{-a_{21}}{\sqrt{a_{11}^2+a_{21}^2}}$，$Q_{21}A=\begin{pmatrix} b_{11} & b_{12} & b_{13} \\ 0 & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix}$，$a_{3i}=b_{3i}$

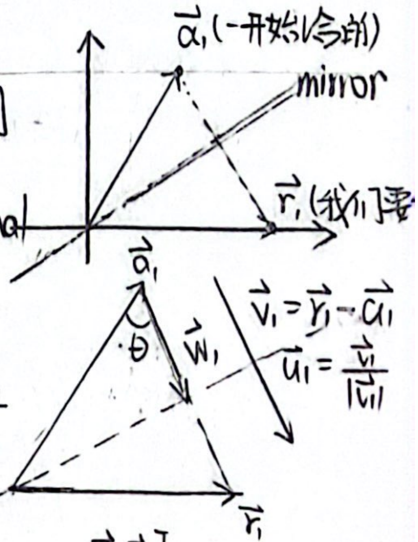
随后 $Q_{31}=\begin{pmatrix} \cos\alpha & 0 & -\sin\alpha \\ 0 & 1 & 0 \\ \sin\alpha & 0 & \cos\alpha \end{pmatrix}$ 在 xz plane 上的 rotation，where $\cos\alpha=\dfrac{b_{11}}{\sqrt{b_{11}^2+b_{31}^2}}$，$\sin\alpha=\dfrac{-b_{31}}{\sqrt{b_{11}^2+b_{31}^2}}$，so $Q_{31}Q_{21}A=\begin{pmatrix} c_{11} & c_{12} & c_{13} \\ 0 & c_{22} & c_{23} \\ 0 & c_{32} & c_{33} \end{pmatrix}$，$b_{2i}=c_{2i}$

随后 $Q_{32}=\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\varphi & -\sin\varphi \\ 0 & \sin\varphi & \cos\varphi \end{pmatrix}$，在 yz plane 上的 rotation，where $\cos\varphi=\dfrac{c_{22}}{\sqrt{c_{22}^2+c_{32}^2}}$，$\sin\varphi=\dfrac{-c_{32}}{\sqrt{c_{22}^2+c_{32}^2}}$，so $Q_{32}Q_{31}Q_{21}A=\begin{pmatrix} d_{11} & d_{12} & d_{13} \\ 0 & d_{22} & d_{23} \\ 0 & 0 & d_{33} \end{pmatrix}$，$c_{1i}=d_{1i}$

so $A=(Q_{32}Q_{31}Q_{21})^{-1}R=QR$.

② Household Reflection 是想人为制造 R

对于一个 $n\times n$ 矩阵，我们先关注 $\vec{a}_1=\begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{pmatrix}$，我们想让其变成 $\vec{r}_1=\begin{pmatrix} |\vec{a}_1| \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ or $\begin{pmatrix} -|\vec{a}_1| \\ 0 \\ \vdots \\ 0 \end{pmatrix}$，通过一个 orthogonal H. 注意 $|\vec{a}_1|=|H\vec{a}_1|=|\vec{r}_1|$ by definition of orthogonal matrix



我们先考虑 $\vec{w}$：其数值 $|\vec{w}_1|=|\vec{a}_1|\cos\theta=|\vec{a}_1|\cdot\dfrac{(-\vec{a}_1)\cdot\vec{v}_1}{|\vec{a}_1||\vec{v}_1|}=-\dfrac{\vec{a}^T\vec{v}_1}{|\vec{v}_1|}$，其方向 $\dfrac{\vec{w}_1}{|\vec{w}_1|}=\dfrac{\vec{v}_1}{|\vec{v}_1|}$，因此 $\vec{w}_1=-\vec{v}_1\dfrac{\vec{a}_1^T\vec{v}_1}{\vec{v}_1^T\vec{v}_1}$

则 $H_1A = \begin{pmatrix} a_{11} & a_{12} & \cdots \\ 0 & a_{22} & \cdots \\ \vdots & \vdots & \ddots \\ 0 & a_{2n} & \end{pmatrix}$，随后我们考虑 $\begin{pmatrix} a_{22} & \cdots & a_{2n} \\ \vdots & \ddots & \vdots \\ a_{2n} & \cdots & a_{nn} \end{pmatrix}$ 这个 $(n-1)\times(n-1)$ matrix，

做对应的 H_2 ... repeat，则 $H_n \cdots H_2H_1A = R$，$A = (H_n \cdots H_2H_1)^{-1}R$

- 对于一个 sparse matrix，在 elimination 的过程中可能会出现 zero elements be "filled-in" and turned into nonzero elements 的情况，因此需要提前 reordering the columns and rows during this process. "approximate minimum degree" algorithm is useful
- the norm of matrix A is the largest ratio of $\frac{|A\vec{x}|}{|\vec{x}|}$：$|A| = \max\limits_{\vec{x}\neq\vec{0}} \frac{|A\vec{x}|}{|\vec{x}|}$

因此：(1) $|A+B| \leq |A| + |B|$ where A, B are matrices, $\vec{x}$ is a vector

proof: $|(A+B)\vec{x}| = |A\vec{x} + B\vec{x}| \leq |A\vec{x}| + |B\vec{x}|$

$$|A+B| = \max_{\vec{x}\neq\vec{0}} \frac{|(A+B)\vec{x}|}{|\vec{x}|} \leq \max_{\vec{x}\neq\vec{0}} \frac{|A\vec{x}|}{|\vec{x}|} + \frac{|B\vec{x}|}{|\vec{x}|} \leq |A| + |B|$$

(2) $|cA| = |c||A|$, where c is a constant

(3) $|A\vec{x}| \leq |A||\vec{x}|$ by definition

(4) $|AB| \leq |A||B|$

proof: $|A\underbrace{B\vec{x}}_{\text{先看为一组}}| \leq |A||B\vec{x}| \leq |A||B||\vec{x}|$

$$|AB| = \max_{\vec{x}\neq\vec{0}} \frac{|AB\vec{x}|}{|\vec{x}|} \leq \max_{\vec{x}\neq\vec{0}} \frac{|A||B||\vec{x}|}{|\vec{x}|} = |A||B|$$

(5) for a positive definite symmetric matrix, the norm is $|A| = \lambda_{max}(A)$ (notice $\lambda_i \geq 0$)

proof $A = Q\Lambda Q^T$，Q 与 Q^T leave the length unchanged so in such way $|A| = \max\limits_{\vec{x}\neq\vec{0}} \frac{|\Lambda\vec{x}|}{|\vec{x}|}$, notice that

$$\lambda_{min}(A) \leq \frac{|\Lambda\vec{x}|}{|\vec{x}|} \leq \lambda_{max}(A) \quad \text{for } \vec{x}\neq\vec{0}$$

(6) for un-symmetric matrix, A^TA is a good way because it is always symmetric. in fact, $|A| = \sigma_{max}(A)$ (notice $\sigma_i \geq 0$)

proof: $|A|^2 = \max\limits_{\vec{x}\neq\vec{0}} \frac{|A\vec{x}|^2}{|\vec{x}|} = \max\limits_{\vec{x}\neq\vec{0}} \frac{\vec{x}^TA^TA\vec{x}}{\vec{x}^T\vec{x}} =$

$$= \max_{\vec{x}\neq\vec{0}} \frac{(c_1\vec{q}_1 + \cdots + c_n\vec{q}_n)^T(c_1\lambda_1\vec{q}_1 + \cdots + c_n\lambda_n\vec{q}_n)}{(c_1\vec{q}_1 + \cdots + c_n\vec{q}_n)^T(c_1\vec{q}_1 + \cdots + c_n\vec{q}_n)}$$

$$= \max_{\vec{x}\neq\vec{0}} \frac{c_1^2\lambda_1 + \cdots + c_n^2\lambda_n}{c_1^2 + \cdots + c_n^2} \quad (\text{notice } \vec{q}_i\vec{q}_j = 0, |\vec{q}_i| = 1)$$

$= \lambda_{max}(A^TA)$

$|A| = \sigma_{max}(A)$.

in fact $\sigma_{min}(A) \leq \frac{|A\vec{x}|}{\ldots} \leq \sigma_{max}(A)$

则 $H_1A = \begin{pmatrix} a_{11} & a_{12} & \\ 0 & a_{22} & \cdots \\ \vdots & \vdots & \ddots \\ 0 & a_{2n} & \end{pmatrix}$，从而后我们考虑 $\begin{pmatrix} a_{22} & \cdots & a_{2n} \\ \vdots & \ddots & \\ a_{2n} & \cdots & a_{nn} \end{pmatrix}$ 这个 $(n-1)\times(n-1)$ matr

做对应的 $H_2 \ldots$ repeat，则 $H_n \cdots H_2H_1A = R$，$A = (H_n \cdots H_2H_1)^{-1}R$

- 对于一个 sparse matrix，在 elimination 的过程中可能会出现 zero elements be "filled-in" and turned into nonzero elements 的情况，因此需要提前 reordering the columns and rows during this process. "approximate minimum degree" algorithm is useful
- the norm of matrix A is the largest ratio of $\frac{|A\vec{x}|}{|\vec{x}|}$: $|A| = \max_{\vec{x}\neq\vec{0}} \frac{|A\vec{x}|}{|\vec{x}|}$

因此：(1) $|A+B| \leq |A| + |B|$ where A, B are matrices, $\vec{x}$ is a vector

proof: $|(A+B)\vec{x}| = |A\vec{x} + B\vec{x}| \leq |A\vec{x}| + |B\vec{x}|$

$$|A+B| = \max_{\vec{x}\neq\vec{0}} \frac{|(A+B)\vec{x}|}{|\vec{x}|} \leq \max_{\vec{x}\neq\vec{0}} \frac{|A\vec{x}|}{|\vec{x}|} + \frac{|B\vec{x}|}{|\vec{x}|} \leq |A| + |B|$$

(2) $|cA| = |c||A|$, where c is a constant

(3) $|A\vec{x}| \leq |A||\vec{x}|$ by definition

(4) $|AB| \leq |A||B|$

proof $|A\underbrace{B\vec{x}}_{\text{先看为一组}}| \leq |A||B\vec{x}| \leq |A||B||\vec{x}|$

$$|AB| = \max_{\vec{x}\neq\vec{0}} \frac{|AB\vec{x}|}{|\vec{x}|} \leq \max_{\vec{x}\neq\vec{0}} \frac{|A||B||\vec{x}|}{|\vec{x}|} = |A||B|$$

(5) for a positive definite symmetric matrix, the norm is $|A| = \lambda_{max}(A)$ (notice $\lambda_i \geq 0$)

proof $A = Q\Lambda Q^T$, Q 与 Q^T leave the length unchanged

so in such way $|A| = \max_{\vec{x}\neq\vec{0}} \frac{|\Lambda\vec{x}|}{|\vec{x}|}$, notice that

$$\lambda_{min}(A) \leq \frac{|\Lambda\vec{x}|}{|\vec{x}|} \leq \lambda_{max}(A) \text{ for } \vec{x}\neq\vec{0}$$

(6) for un-symmetric matrix. A^TA is a good way because it is always symmetric. in fact, $|A| = \sigma_{max}(A)$ (notice $\sigma_i \geq 0$)

proof: $|A|^2 = \max_{\vec{x}\neq\vec{0}} \frac{|A\vec{x}|^2}{|\vec{x}|} = \max_{\vec{x}\neq\vec{0}} \frac{\vec{x}^TA^TA\vec{x}}{\vec{x}^T\vec{x}} = \lambda_{max}(A^TA)$

$$= \max_{\vec{x}\neq\vec{0}} \frac{(c_1\vec{q}_1 + \cdots + c_n\vec{q}_n)^T(c_1\lambda_1\vec{q}_1 + \cdots + c_n\lambda_n\vec{q}_n)}{(c_1\vec{q}_1 + \cdots + c_n\vec{q}_n)^T(c_1\vec{q}_1 + \cdots + c_n\vec{q}_n)}$$

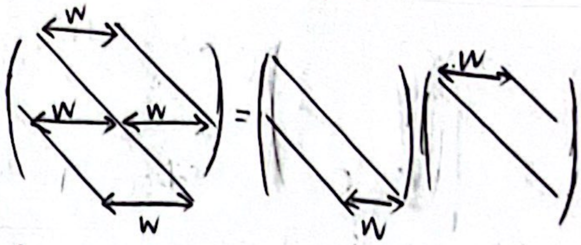
$$= \max_{\vec{x}\neq\vec{0}} \frac{c_1^2\lambda_1 + \cdots + c_n^2\lambda_n}{c_1^2 + \cdots + c_n^2} \quad (\text{notice } \vec{q}_i\vec{q}_j = 0, |\vec{q}_i| = 1)$$

$= \lambda_{max}(A^TA)$

$|A| = \sigma_{max}(A)$.

in fact, $\sigma_{min}(A) \leq \frac{|A\vec{x}|}{|\vec{x}|} \leq \sigma_{max}(A)$

$a_{ij} = 0$ if $|i-j| > w$



$w=1$ for diagonal matrix, $w=2$ for tridiagonal, $w=n$ for dense

each stage of elimination is completed after $w(w-1)$ operations, and the band structure survives. so: elimination on a band matrix (A to LU) needs less that $w^2 n$ operations. for large n. similarly, in this situation, solving $L\vec{c}=\vec{b}$ and $U\vec{x}=\vec{c}$ only needs $2wn$ operations

- 通过以下两种方法可以较快找到$A=QR$.

① 因为 product of orthogonal matrix is also orthogonal. 对于一个3×3 matrix 我们可以通过一次次的 orthogonal rotation matrix to find R.

$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$, $Q_{21} = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$ 在 xy plane 上的 rotation.

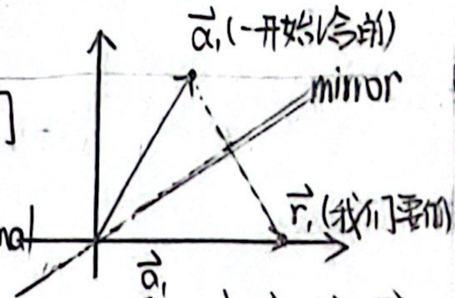
where $\cos\theta = \frac{a_{11}}{\sqrt{a_{11}^2 + a_{21}^2}}$, $\sin\theta = \frac{-a_{21}}{\sqrt{a_{11}^2+a_{21}^2}}$. $Q_{21}A = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ 0 & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix}$, $a_{3i} = b_{3i}$

随后 $Q_{31} = \begin{pmatrix} \cos\alpha & 0 & -\sin\alpha \\ 0 & 1 & 0 \\ \sin\alpha & 0 & \cos\alpha \end{pmatrix}$ 在 xz plane 上的 rotation, where $\cos\alpha = \frac{b_{11}}{\sqrt{b_{11}^2+b_{31}^2}}$, $\sin\alpha = \frac{-b_{31}}{\sqrt{b_{11}^2+b_{31}^2}}$, so $Q_{31}Q_{21}A = \begin{pmatrix} c_{11} & c_{12} & c_{13} \\ 0 & c_{22} & c_{23} \\ 0 & c_{32} & c_{33} \end{pmatrix}$, $b_{2i} = c_{2i}$

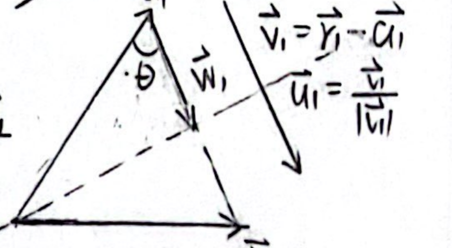
随后 $Q_{32} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\varphi & -\sin\varphi \\ 0 & \sin\varphi & \cos\varphi \end{pmatrix}$, 在 yz plane 上的 rotation, where $\cos\varphi = \frac{c_{22}}{\sqrt{c_{22}^2+c_{32}^2}}$, $\sin\varphi = \frac{-c_{32}}{\sqrt{c_{22}^2+c_{32}^2}}$; so $Q_{32}Q_{31}Q_{21}A = \begin{pmatrix} d_{11} & d_{12} & d_{13} \\ 0 & d_{22} & d_{23} \\ 0 & 0 & d_{33} \end{pmatrix}$, $c_{1i} = d_{1i}$

so $A = (Q_{32}Q_{31}Q_{21})^{-1}R = QR$

② Household Reflection 是想人为制造 R



对于一个 $n\times n$ 矩阵, 我们先关注 $\vec{a}_1 = \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{pmatrix}$, 我们想让其变成 $\vec{r}_1 = \begin{pmatrix} |\vec{a}_1| \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ or $\begin{pmatrix} -|\vec{a}_1| \\ 0 \\ \vdots \\ 0 \end{pmatrix}$, 通过一个 orthogonal H. 注意 $|\vec{a}_1| = |H\vec{a}_1| = |\vec{r}_1|$ by definition of orthogonal matrix



我们先考虑 $\vec{w}$: 其数值 $|\vec{w}_1| = |\vec{a}_1|\cos\theta = |\vec{a}_1| \frac{(-\vec{a}_1)\cdot\vec{v}_1}{|\vec{a}_1||\vec{v}_1|} = -\frac{\vec{a}^T\vec{v}_1}{|\vec{v}_1|}$, 其方向: $\frac{\vec{w}_1}{|\vec{w}_1|} = \frac{\vec{v}_1}{|\vec{v}_1|}$, 因此 $\vec{w}_1 = -\vec{v}_1 \frac{\vec{a}_1^T\vec{v}_1}{\vec{v}_1^T\vec{v}_1}$

so $H_1\vec{a}_1 = \vec{r}_1 = a_1 - 2\vec{w}_1 = a_1 - 2\vec{v}_1 \frac{\vec{a}_1^T\vec{v}_1}{\ldots}$ $H = I - 2\frac{\vec{v}_1\vec{v}_1^T}{\ldots}$ ...

在计算中，通过使用multigrid与conjugate gradient，可以简便计算

- 在计算 $A\vec{x}=\vec{b}$ 时，得到的 $\vec{x}$ 会有误差，这误差可能来自计算机/测量

(i) $\Delta\vec{b}$ leads to $\Delta\vec{x}$.

$A(\vec{x}+\Delta\vec{x})=\vec{b}+\Delta\vec{b}$，$A\Delta\vec{x}=\Delta\vec{b}$，$\Delta\vec{x}=A\Delta\vec{b}$

单纯比较数值是无意义的，我们只对 $\frac{|\Delta\vec{x}|}{|\vec{x}|}$ 感兴趣，relative error.

注意到 $|\Delta\vec{x}|\leq|A^{-1}||\Delta\vec{b}|$，(consider eigenvectors)，因此

the solution error is less than condition number $c=|A||A^{-1}|$ times the problem error:

$$\frac{|\Delta\vec{x}|}{|\vec{x}|}\leq c\frac{|\Delta\vec{b}|}{|\vec{b}|}$$

proof: $|\Delta\vec{x}|\leq|A^{-1}||\Delta\vec{b}|$，$|\vec{b}|\leq|A||\vec{x}|$

(ii) ΔA leads to $\Delta\vec{x}$

if the problem error is ΔA, c still controls $\Delta\vec{x}$: $\frac{|\Delta\vec{x}|}{|\vec{x}+\Delta\vec{x}|}\leq c\frac{|\Delta A|}{|A|}$

proof: $\because (A+\Delta A)(\vec{x}+\Delta\vec{x})=\vec{b}$，$A\vec{x}=\vec{b}$

$\therefore A\Delta\vec{x}=-(\Delta A)(\vec{x}+\Delta\vec{x})$，$\Delta\vec{x}=-A^{-1}(\Delta A)(\vec{x}+\Delta\vec{x})$

$\because |\Delta\vec{x}|\leq|A^{-1}||\Delta A||\vec{x}+\Delta\vec{x}|$

$$\therefore \frac{|\Delta\vec{x}|}{|\vec{x}+\Delta\vec{x}|}\leq c\frac{|\Delta A|}{|A|}$$

[注意：$c=|A||A^{-1}|$ 本身就意味着简单的rescale并不会改变relative error]

- iterative method in matrix calculation.

当计算机在计算 $A\vec{x}=\vec{b}$，我们或许可以考虑找一个与A相关但方便计算的矩阵S, and corresponding difference $T=S-A$. by using iterative method, we rewrite $A\vec{x}=\vec{b}$ into: $S\vec{x}=T\vec{x}+\vec{b}$.

通过猜一个 $\vec{x}_0$ 并 repeat $S\vec{x}_{k+1}=T\vec{x}_k+\vec{b}$，当 residual $\vec{r}_k=\vec{b}-\vec{x}_k$ 足够小，我们便可以停止计算并得到 $\vec{x}$ 的一个approximation. 毕竟 if converge $\vec{x}_\infty=\vec{x}$.

(1) 我们在决定 $A=S-T$ 时要考虑① speed per step (由S是否好算决定)；② fast convergence (由 $S^{-1}T$ 的 $|\lambda_{max}(S^{-1}T)|$ 决定)

proof: 我们主要关注②：

$\because \vec{e}_k=\vec{x}-\vec{x}_k$，$S\vec{e}_{k+1}=S\vec{x}-S\vec{x}_{k+1}$，$T\vec{e}_k=T\vec{x}-T\vec{x}_k$

$\therefore S\vec{x}-T\vec{x}=A\vec{x}=\vec{b}=T\vec{x}_k+\vec{b}-T\vec{x}_k=S\vec{x}_{k+1}-T\vec{x}_k$

$\therefore S\vec{e}_{k+1}=T\vec{e}_k$，$\vec{e}_{k+1}=S^{-1}T\vec{e}_k$

if $B=S^{-1}T$，$\vec{e}_k=B^k\vec{e}_0$

since $\vec{e}_\infty=\vec{0}$, the powers B^k approach 0 if and only if every eigenvalue of B has $|\lambda|<1$ the rate of convergence

在计算中，同时使用multigrid与conjugate gradients，可以简便运算

is controlled by the spectral radius of B $\rho = \max|\lambda(B)|$

notice that there is a technical difficulty when B doesn't ha[ve] enough independent eigenvectors. in such way we express B in terms of Jordan form. but even in this situation we still need keep all $|\lambda|<1$

② with enough independent eigenvectors:

$$e_0 = c_1\vec{x}_1 + \cdots + c_n\vec{x}_n,\quad e_k = c_1\lambda_1^k\vec{x}_1 + \cdots + c_n\lambda_n^k\vec{x}_n$$

without enough independent eigenvectors (2x2 matrix)

$$B = MJM^{-1},\quad B^k = MJ^kM^{-1},\quad J = \begin{pmatrix}\lambda & 1\\ 0 & \lambda\end{pmatrix},\quad J^k = \begin{pmatrix}\lambda^k & k\lambda^{k-1}\\ 0 & \lambda^k\end{pmatrix}$$

(2) here are some popular choices for iterative methods:

① Jacobi iteration: S = diagonal of A

it works well when the main diagonal of A is large compared to the off-diagonal part

② Gauss-Seidel: S = lower triangular part of A, including diago[nal]

③ successive overrelaxation method (SOR):

we introduce a parameter ω into the iteration, then choose ω to make the spectral radius of $S^{-1}T$ as small as possible

rewrite $A\vec{x} = \vec{b}$ into $\omega A\vec{x} = \omega\vec{b}$. the matrix S in SOR has the diagonal of the original A. but below the diagonal we use ωA:

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots \\ a_{21} & a_{22} & a_{23} & \cdots \\ a_{31} & a_{32} & a_{33} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix},\quad S = \begin{pmatrix} a_{11} & & & \\ \omega a_{21} & a_{22} & & \\ \omega a_{31} & \omega a_{32} & a_{33} & \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix},\quad T = S - \omega A$$

$$S\vec{x}_{k+1} = T\vec{x}_k + \omega\vec{b}$$

④ incomplete LU: set the small nonzeros in L and U back to zero.

[注意：在我们得到L_0与U_0后，不要求L_0U_0，还是用forward part & ba[ck] substitution求解]

$$L_0U_0\vec{x}_{k+1} = (L_0U_0 - A)\vec{x}_k + \vec{b}$$

if L_0U_0 is close to A, then $|\lambda|_{max}$ is small and a few iteratio[ns] will give a close answer

(3) iterations are intended for large sparse matrices - when a hig[h] percentage of the entries are zero. the not good zeros are inside th[e] band, which is wide. they become nonzeros in L and U, which is why eliminations become expensive

在计算中，通过使用multigrid与conjugate gradients，可以简便运算

① multigrid: solve smaller problems (often coming from coarser grids and doubled step sizes Δx and Δy). each iteration will be cheaper and convergence will be faster. then interpolate between the values computed on the coarse grid to get a quick and close head-start on the full-size problem

② conjugate gradients: $\vec{x}_{k+1}$ is not the best combination of $\vec{x}_0, A\vec{x}_0, \ldots A^k\vec{x}_0$. thus this method finds the best combination x_k at each step，其中bases之间存在共轭关系，以避免重复修正某一方向的误差

(4) 用iteration求eigenvalue

当matrix比较复杂时 $\det(A-\lambda I)=0$ 很难计算（因为degree $n>4$ 的polynomials没有通解），通过使用power and inverse power method 与 QR method，可以简便得到eigenvalue

① power and inverse power method

$$\vec{u}_k = A^k\vec{u}_0 = c_1\lambda_1^k\vec{x}_1 + \cdots + c_n\lambda_n^k\vec{x}_n$$

当k足够大，the largest eigenvalue begins to dominate，the vector $\vec{u}_k$ point towards that dominant eigenvector

the speed of convergence depends on the ratio of the second largest eigenvalue to the largest one

同理，利用 $\vec{u}_k = A^{-k}\vec{u}_0 = \frac{c_1\vec{x}_1}{\lambda_1^k} + \cdots + \frac{c_n\vec{x}_n}{\lambda_n^k}$，we can find the smallest eigenvalue as k is large enough

for high speed, we make change to λ by shifting matrix to $(A-\lambda^* I)$. this shift doesn't change eigenvectors. so we may let λ^* be an element on the diagonal or $\frac{\vec{x}^T A\vec{x}}{\vec{x}^T\vec{x}}$ (Rayleigh quotient) to simplify calculation

② QR method

if $A_0 = Q_0R_0$ we find $A_1 = R_0Q_0$, A与A_1 share the same eigenvalues since $A_1 = Q^{-1}AQ$, they are similar (Jordan form)

we can continue to factor $A_k = Q_kR_k$ and find $A_{k+1} = R_kQ_k$ until we can find eigenvalues easily (often on the diagonal).

we can also factor $A_k - c_kI = Q_kR_k$, and find $A_{k+1} = R_kQ_k + c_kI$.

and A_k and A_{k+1} still share the same eigenvalues.

proof:

$\because R_k = Q_k^T(A_k - c_k I)$

$\therefore R_k Q_k + c_k I = Q_k^T A_k Q_k - Q_k^T c_k I Q_k + c_k I = Q_k^T A_k Q_k = A_{k+1}$

$\therefore A_k$ and A_{k+1} are similar

a good shift choose c_k near an (unknown) eigenvalue that eigenvalue appears more accurately on the diagonal of A_{k+1} – which tells us a better c_{k+1} for the next step to A_{k+2}

the other idea is to obtain off-diagonal zero before the QR method starts ($E^{-1}AE$, do not forget E^{-1} to keep λ)

11、complex vector and matrices

- when we transpose a complex vector $\vec{z}$ or complex matrix A, take the conjugate too: $\bar{\vec{z}}^T = \vec{z}^H$, which is $\vec{z}$ Hermitian

(1) the length $|\vec{z}|$ is $\sqrt{\vec{z}^H \vec{z}}$, wh

(2) the inner product of real or complex vectors $\vec{u}$ and $\vec{v}$ is $\vec{u}^H \vec{v}$

notice that at that time $\vec{u}^H \vec{v}$ is the conjugate of $\vec{v}^H \vec{u}$

a zero inner product still means the complex vectors are orthogonal.

(3) the inner product of $A\vec{u}$ with $\vec{v}$ equals the inner product of $\vec{u}$ with $A^H \vec{u}$: $(A\vec{u})^H \vec{v} = \vec{u}^H A^H \vec{v}$.

(4) $(AB)^H = B^H A^H$

- complex symmetric matrix is Hermitian matrix. $A = A^H$

(1) if $A = A^H$ and $\vec{z}$ is any vector. $z^H A z$ is real.

proof: $\because (z^H A z)^H = z^H A^H (z^H)^H = z^H A z$

(2) every eigenvalue of a Hermitian matrix is real

proof: suppose $A\vec{z} = \lambda \vec{z}$ $\quad \therefore \vec{z}^H A \vec{z} = \lambda \vec{z}^H \vec{z}$

(3) the eigenvectors of a Hermitian matrix are orthogonal when the correspond to different eigenvalues.

proof: if $A\vec{z} = \lambda \vec{z}$, $A\vec{y} = \beta \vec{y}$, $\lambda \neq \beta$

$\therefore \vec{y}^H A \vec{z} = \lambda \vec{y}^H \vec{z} = (A\vec{y})^H \vec{z} = \beta \vec{y}^H \vec{z}$ $\quad \therefore \vec{y}^H \vec{z} = 0$

(4) the eigenvectors of a Hermitian matrix can be chosen to be orthonormal

proof. $A = QTQ^{-1}$ for all square A, where T is upper triangular, $Q^H = Q^{-1}$

A is complex symmetric, so $T = \Lambda$

- a unitary matrix U is a (complex) square matrix that has orthonormal columns

(1) $U^H U = I$ for all unitary matrices U

(2) if U is square, $U^H = U^{-1}$

(3) $|U\vec{z}| = |\vec{z}|$, so all possible $|\lambda| = 1$ (similar thing happens for Q. since $|Q\vec{x}| = |\vec{x}|$, so all possible $|\lambda| = 1$)

- Fast Fourier Transformation (FFT)

the $n\times n$ fourier matrix contains powers of $w = e^{\frac{2\pi i}{n}}$

$$F_n\vec{c} = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & w & w^2 & \cdots & w^{n-1} \\ 1 & w^2 & w^4 & \cdots & w^{2(n-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & w^{n-1} & w^{2(n-1)} & \cdots & w^{(n-1)^2} \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \\ \vdots \\ c_{n-1} \end{pmatrix} = \begin{pmatrix} y_0 \\ y_1 \\ \vdots \\ y_{n-1} \end{pmatrix} = \vec{y}$$

F_n is symmetric but not Hermitian. its columns are orthogonal.

$F_n \overline{F_n} = nI$, $F_n^{-1} = \frac{\overline{F_n}}{n}$.

if we consider the starting row/column as 0^{th} row/column. the entry in row j column k is w^{jk}.

(1) when $f(x)$ has period of 2π, and we change x to $e^{i\theta}$, then the function is defined around the unit circle (where $z = e^{i\theta}$). then the Discrete Fourier Transform from $\vec{y}$ to $\vec{c}$ is matching n values of $f(x)$ by a polynomial $p(z) = c_0 + c_1 z + \cdots + c_{n-1} z^{n-1}$

$F_n\vec{c} = \vec{y}$: finds $c_0, \dots, c_{n-1}$ such that $p(z) = f(z)$ at $z = 1, w, \dots, w^{n}$

(2) 一种计算FFT的简便方法 ($m = \frac{1}{2}n$, n is even, $w_x = e^{\frac{2\pi i}{x}}$)

$$F_n = \begin{pmatrix} I_{m\times m} & D_{m\times m} \\ I_{m\times m} & -D_{m\times m} \end{pmatrix} \begin{pmatrix} F_m & \\ & F_m \end{pmatrix} \begin{pmatrix} \text{even-odd} \\ \text{permutation} \end{pmatrix}$$

其中 $D_{m\times m} = \begin{pmatrix} 1 & & \\ & w_m & \\ & & \ddots \\ & & & w_m^{m-1} \end{pmatrix}$, even-odd permutation makes all even c_i's $(c_0, c_2, \dots)$ ahead of odd c_i's $(c_1, c_3, \dots)$

so we firstly divide $\vec{c}$ into two parts: $\vec{c}' = \begin{pmatrix} c_0 \\ c_2 \\ \vdots \end{pmatrix}$, $\vec{c}'' = \begin{pmatrix} c_1 \\ c_3 \\ \vdots \end{pmatrix}$

then do half-size transformation $\vec{y}' = F_m\vec{c}'$, $\vec{y}'' = F_m\vec{c}''$

finally combine them to reconstruct $\vec{y}$:

$$\begin{cases} y_j = y_j' + w_n^j y_j'' & \text{for } j = 0, \dots, m-1 \\ y_{j+m} = y_j' - w_n^j y_j'' & \text{for } j = 0, \dots, m-1 \end{cases}$$

proof: $y_j = \sum^{n-1} w_n^{jk} c_k = \sum^{m-1} w_n^{2jk} c_{2k} + \sum^{m-1} w_n^{j(2k+1)} c_{2k+1}$

$$= \sum_{k=0}^{m-1} W_m^{jk} c_k' + (w_n)^j \sum_{k=0}^{m-1} W_m^{jk} c_k''$$

$$= y_j' + (w_n)^j y_j'' \quad (\text{for } j \geq m, (w_n^j) \text{ has factor } (w_n)^m = -1)$$

we can keep that again and again... (if $u = \frac{1}{2}m$, m is even)

$$\begin{pmatrix} F_m & \\ & F_m \end{pmatrix} = \begin{pmatrix} I_{u\times u} & D_{u\times u} & & \\ I_{u\times u} & -D_{u\times u} & & \\ & & I_{u\times u} & D_{u\times u} \\ & & I_{u\times u} & -D_{u\times u} \end{pmatrix} \begin{pmatrix} F_u & & & \\ & F_u & & \\ & & F_u & \\ & & & F_u \end{pmatrix} \begin{pmatrix} \text{pick } 0,4,8\ldots \\ \text{pick } 2,6,10\ldots \\ \text{pick } 1,5,9,\ldots \\ \text{pick } 3,7,11,\ldots \end{pmatrix}$$

③ for a nxn fourier matrix, using this method, we need $\frac{1}{2}n\log_2 n$ operations to yield final answer

MATRIX FACTORIZATIONS

1. $A = LU = \begin{pmatrix} \text{lower triangular } L \\ \text{1's on the diagonal} \end{pmatrix} \begin{pmatrix} \text{upper triangular } U \\ \text{pivots on the diagonal} \end{pmatrix}$

   Requirements: No row exchanges as Gaussian elimination reduces A to U.

2. $A = LDU = \begin{pmatrix} \text{lower triangular } L \\ \text{1's on the diagonal} \end{pmatrix} \begin{pmatrix} \text{pivot matrix} \\ D \text{ is diagonal} \end{pmatrix} \begin{pmatrix} \text{upper triangular } U \\ \text{1's on the diagonal} \end{pmatrix}$

   Requirements: No row exchanges. The pivots in D are divided out to leave 1's on the diagonal of U. If A is symmetric then U is L^{T} and $A = LDL^{\mathrm{T}}$.

3. $PA = LU$ (permutation matrix P to avoid zeros in the pivot positions).

   Requirements: A is invertible. Then P, L, U are invertible. P does all of the row exchanges in advance, to allow normal LU. Alternative: $A = L_1 P_1 U_1$.

4. $EA = R$ (m by m invertible E) (any matrix A) = rref(A).

   Requirements: None! *The reduced row echelon form* R has r pivot rows and pivot columns. The only nonzero in a pivot column is the unit pivot. The last $m - r$ rows of E are a basis for the left nullspace of A; they multiply A to give zero rows in R. The first r columns of E^{-1} are a basis for the column space of A.

5. $A = C^{\mathrm{T}}C$ = (lower triangular) (upper triangular) with $\sqrt{D}$ on both diagonals

   Requirements: A is symmetric and positive definite (all n pivots in D are positive). This *Cholesky factorization* C = chol(A) has $C^{\mathrm{T}} = L\sqrt{D}$, so $C^{\mathrm{T}}C = LDL^{\mathrm{T}}$.

6. $A = QR$ = (orthonormal columns in Q) (upper triangular R).

   Requirements: A has independent columns. Those are *orthogonalized* in Q by the Gram-Schmidt or Householder process. If A is square then $Q^{-1} = Q^{\mathrm{T}}$.

7. $A = S\Lambda S^{-1}$ = (eigenvectors in S) (eigenvalues in Λ) (left eigenvectors in S^{-1}).

   Requirements: A must have n linearly independent eigenvectors.

8. $A = Q\Lambda Q^{\mathrm{T}}$ = (orthogonal matrix Q) (real eigenvalue matrix Λ) (Q^{T} is Q^{-1}).

   Requirements: A is *real and symmetric*. This is the Spectral Theorem.

9. $A = MJM^{-1} =$ (generalized eigenvectors in M) (Jordan blocks in J) (M^{-1}).

Requirements: A is any square matrix. This *Jordan form* J has a block for each independent eigenvector of A. Every block has only one eigenvalue.

10. $A = U\Sigma V^{\mathrm{T}} = \begin{pmatrix} \text{orthogonal} \\ U \text{ is } m \times \cancel{n}\, r \end{pmatrix} \begin{pmatrix} \overset{r}{\cancel{n}} \times \overset{r}{\cancel{n}} \text{ singular value matrix} \\ \sigma_1, \ldots, \sigma_r \text{ on its diagonal} \end{pmatrix} \begin{pmatrix} \text{orthogonal} \\ V^{T} \text{is } \underset{r}{\cancel{n}} \times n \end{pmatrix}.$

Requirements: None. This *singular value decomposition* (SVD) has the eigenvectors of AA^{T} in U and eigenvectors of $A^{\mathrm{T}}A$ in V; $\sigma_i = \sqrt{\lambda_i(A^{\mathrm{T}}A)} = \sqrt{\lambda_i(AA^{\mathrm{T}})}$.

11. $A^+ = V\Sigma^+U^{\mathrm{T}} = \begin{pmatrix} \text{orthogonal} \\ n \times n \end{pmatrix} \begin{pmatrix} n \times m \text{ pseudoinverse of } \Sigma \\ 1/\sigma_1, \ldots, 1/\sigma_r \text{ on diagonal} \end{pmatrix} \begin{pmatrix} \text{orthogonal} \\ m \times m \end{pmatrix}.$

Requirements: None. The *pseudoinverse* A^+ has $A^+A =$ projection onto row space of A and $AA^+ =$ projection onto column space. The shortest least-squares solution to $Ax = b$ is $\widehat{x} = A^+b$. This solves $A^{\mathrm{T}}A\widehat{x} = A^{\mathrm{T}}b$.

12. $A = QH =$ (orthogonal matrix Q) (symmetric positive definite matrix H).

Requirements: ~~A is invertible.~~ This *polar decomposition* has $H^2 = A^{\mathrm{T}}A$. The factor H is semidefinite if A is singular. The reverse polar decomposition $A = KQ$ has $K^2 = AA^{\mathrm{T}}$. Both have $Q = UV^{\mathrm{T}}$ from the SVD.

13. $A = U\Lambda U^{-1} =$ (unitary U) (eigenvalue matrix Λ) (U^{-1} which is $U^{\mathrm{H}} = \overline{U}^{\mathrm{T}}$).

Requirements: A is *normal*: $A^{\mathrm{H}}A = AA^{\mathrm{H}}$. Its orthonormal (and possibly complex) eigenvectors are the columns of U. Complex λ's unless $A = A^{\mathrm{H}}$: Hermitian case.

14. $A = UTU^{-1} =$ (unitary U) (triangular T with λ's on diagonal) ($U^{-1} = U^{\mathrm{H}}$).

Requirements: *Schur triangularization* of any square A. There is a matrix U with orthonormal columns that makes $U^{-1}AU$ triangular: Section 6.4.

15. $F_n = \begin{bmatrix} I & D \\ I & -D \end{bmatrix} \begin{bmatrix} F_{n/2} & \\ & F_{n/2} \end{bmatrix} \begin{bmatrix} \text{even-odd} \\ \text{permutation} \end{bmatrix} =$ one step of the (recursive) **FFT**.

Requirements: $F_n =$ Fourier matrix with entries w^{jk} where $w^n = 1$: $F_n\overline{F}_n = nI$. D has $1, w, \ldots, w^{n/2-1}$ on its diagonal. For $n = 2^\ell$ the *Fast Fourier Transform* will compute F_nx with only $\frac{1}{2}n\ell = \frac{1}{2}n\log_2 n$ multiplications from ℓ stages of D's.