Chapter 3. Direct Methods for Linear Systems.

Section 1. Gauss Elimination Method.

$$\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \end{pmatrix} \to (2) - (1) \times \tfrac{1}{2}$$

$$\begin{pmatrix} 2 & 1 \\ 0 & 3/2 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 3 \\ 3/2 \end{pmatrix} \to \quad \begin{aligned} &\tfrac{3}{2}y = \tfrac{3}{2} \Rightarrow y = 1 \\ &2x + y = 3 \\ &\Rightarrow 2x = 3 - y = 2 \Rightarrow x = 1 \end{aligned}$$

$$A = (a_{ij})_{n\times n} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & & \cdots & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \in \mathbb{R}^n$$

Linear System: $A\ddot{x} = b$ (1). $\quad x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n \quad b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} \in \mathbb{R}^n$

Assume A is invertible

Idea: eliminate the entries in the lower part of the matrix.

For $k = 1, 2, \cdots n-1$ do (as follows)

For $i = k+1, k+2, \cdots n$ do $\quad$ let $l_{ik} = a_{ik}/a_{kk}$

For $j = k, k+1, \cdots n$ do $\quad b_i^{(k+1)} \leftarrow b_i^{(k)} - l_{ik} b_k^{(k)}$

$$a_{ij}^{(k+1)} \leftarrow a_{ij}^{(k)} - l_{ik} a_{kj}^{(k)}$$

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ & a_{22}^{(2)} & \cdots & a_{2n}^{(2)} \\ & & a_{33}^{(3)} \cdots & a_{3n}^{(3)} \\ & & \ddots & \\ & & & a_{nn}^{(n)} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2^{(2)} \\ b_3^{(3)} \\ \vdots \\ b_n^{(n)} \end{pmatrix} \qquad (2)$$

Assumption: The diagonal entries are always non-zero in the elimination process.

$$a_{kk}^{(k)} \neq 0.$$

$$\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \quad (\times)$$

The upper/right triangular system can be solved by the backward substitution method.

For $k = n, n-1, \cdots 1$. do

$$a_{kk}^{(k)} x_k = b_k^{(k)} - \sum_{j=k+1}^{n} a_{kj}^{(k)} x_j$$

$$x_k = \left(b_k^{(k)} - \sum_{j=k+1}^{n} a_{kj}^{(k)} x_j \right) / a_{kk}^{(k)} \qquad (*).$$

algorithm complexity.

$$O(n^3) + O(n^2).$$

Section 2. LU-decomposition Method.

Decompose A into the product of a lower triangular and an upper triangular matrix.

Denote the lower and upper triangular matrix by L and U, respectively.

$$A = LU \quad \text{with} \quad L = \begin{pmatrix} 1 & & & & \\ l_{21} & 1 & & & \\ l_{31} & l_{32} & 1 & & \\ \cdots & \cdots & \cdots & \ddots & \\ l_{n1} & l_{n2} & \cdots & \cdots & 1 \end{pmatrix}$$

$$U = \begin{pmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ & u_{22} & \cdots & u_{2n} \\ & & \ddots & \vdots \\ & & & u_{nn} \end{pmatrix}$$

l_{ik}

$$Ax = b \Rightarrow LUx = b.$$

Let $Ux = y$. $\Rightarrow Ly = b$.

Step 1. solve $Ly = b$. by forward substitution.

$$\begin{pmatrix} 1 & & & \\ l_{21} & 1 & & \\ l_{31} & l_{32} & 1 & \\ \cdots & \cdots & \cdots & \cdots \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} \quad \checkmark$$

$O(n^2)$

step 2, solve $Ux = y$ by backward substitution.

$$\begin{pmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ & u_{22} & \cdots & u_{2n} \\ & & \ddots & \vdots \\ & & & u_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}.$$

$O(n^2)$

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & & \cdots & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} = \begin{pmatrix} 1 & & & \\ l_{21} & 1 & & \\ l_{31} & l_{32} & 1 & \\ l_{n1} & l_{n2} & \cdots & 1 \end{pmatrix} \begin{pmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ & u_{22} & \cdots & u_{2n} \\ & & \ddots & \\ & & & u_{nn} \end{pmatrix}$$

$a_{11} = u_{11}$ $\quad a_{12} = u_{12}.$ $\quad a_{1j} = u_{1j}$ $\quad j = 1, 2 \cdots n$

$l_{i1} u_{11} = a_{i1}$ $\quad i = 2, 3 \cdots n \Rightarrow u_{i1} = a_{i1}/u_{11}$

For $K = 1, 2, \cdots n$.

$u_{kj} = ?$ $\quad a_{kj} = \sum_{m=1}^{k} l_{km} u_{mj}$ $\quad j = k, k+1 \cdots n$

$$= u_{kj} + \sum_{m=1}^{k-1} l_{km} u_{mj}$$

$$u_{kj} = a_{kj} - \sum_{m=1}^{k-1} l_{km} u_{mj} \leftarrow j = k, k+1 \cdots n$$

$l_{ik} = ?$ $\quad a_{ik} = \sum_{m=1}^{k} l_{im} u_{mk}$, $\quad i = k+1, \cdots n$

$$l_{ik} \cdot u_{kk} = a_{ik} - \sum_{m=1}^{k-1} l_{im} u_{mk}$$

$i = k+1, k+2, \cdots n$

$$\Rightarrow \quad l_{ik} = \left(a_{ik} - \sum_{m=1}^{k-1} l_{im} u_{mk}\right) / u_{kk}$$

Doolittle decomposition.

section 3. QR decomposition

$$A = QR.$$

First, decompose a matrix A into the product of an orthogonal matrix Q and a right triangular matrix R.

$$Ax = b \Rightarrow QRx = b.$$

Let $Rx = y$, $\quad Qy = b.$

$$\Downarrow$$

$$y = Q^T b.$$

$$\Rightarrow Rx = Q^T b.$$

$$R = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & \cdots & r_{2n} \\ & & \ddots & \vdots \\ & & & r_{nn} \end{pmatrix}. \qquad Q = (q_1, q_2 \cdots q_n)$$

q_j : j^{th} column vector of Q.

$q_j \in \mathbb{R}^n$.

$$
(q_1,\ q_2 \cdots q_n)\begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & \cdots & r_{2n} \\ & & \ddots & \vdots \\ & & & r_{nn} \end{pmatrix} = (a_1, a_2, \cdots a_n)
$$

$$
q_i^T q_j = \begin{cases} 1. & j = i \\ 0 & j \neq i \end{cases}
$$

$\|q_j\|_2 = 1.$

a_j : j^{th} column vector of A.

$a_j \in \mathbb{R}^n$

$$
\begin{cases} q_1 r_{11} = a_1 & (1) \\ q_1 r_{12} + q_2 r_{22} = a_2 & (2). \\ q_1 r_{13} + q_2 r_{23} + q_3 r_{33} = a_3 \\ q_1 r_{14} + q_2 r_{24} + q_3 r_{34} + q_4 r_{44} = a_4 \end{cases}
$$

$|r_{11}| = \|a_1\|_2 \Rightarrow r_{11} = \|a_1\|_2$

$q_1 = \dfrac{a_1}{r_{11}}$

$\cdots\ \cdots$

$$
\sum_{i=1}^{j} q_i r_{ij} = a_j \qquad (3)
$$

Take inner product of (2) with q_1

$$
(q_1, q_1) r_{12} + \overset{0}{\cancel{(q_1, q_2)}} r_{22} = (q_1, a_2) \Rightarrow r_{12} = (q_1, a_2).
$$

inner product of (2) with q_2.

$$
(q_2, q_1) r_{12} + (q_2, q_2) r_{22} = (q_2, a_2) \Rightarrow r_{22} = (\underline{q_2}, a_2)\ ?
$$

$$
q_2 r_{22} = \underline{a_2 - q_1 r_{12}} \qquad r_{22} = \|a_2 - q_1 r_{12}\|_2 \ \checkmark
$$

$$
q_2 = (a_2 - q_1 r_{12}) / r_{22}
$$

Take inner product of (3) with q_m, $m=1,2,\cdots j-1$.

$$\left(q_m, \sum_{i=1}^{j} q_i r_{ij}\right) = (q_m, a_j).$$

$$(q_m, q_m)\, r_{mj} = (q_m, a_j) \Rightarrow r_{mj} = (q_m, a_j). \quad m=1,2\cdots j-1.$$

$$q_j r_{jj} = a_j - \sum_{i=1}^{j-1} q_i r_{ij}$$

$$r_{jj} = \|a_j - \sum_{i=1}^{j-1} q_i r_{ij}\|_2. \Rightarrow q_j = \left(a_j - \sum_{i=1}^{j-1} q_i r_{ij}\right) / r_{jj}$$

Gram-Schmidt orthogonalization

$(a_1, a_2, \cdots a_n)$

normalize a_1: $\quad q_1 \Leftarrow a_1, \quad \|q_1\|_2 = 1.$

$a_1 = r_{11} q_1 \quad r_{11} = \|a_1\|_2$

$q_1 = a_1 / r_{11}$

$a_2 - (a_2, q_1)\, q_1 = p_2.$

$r_{22} = \|p_2\|_2. \quad r_{22} q_2 = p_2. \Rightarrow q_2 = p_2 / r_{22}$

$$a_j - \sum_{i=1}^{j-1} (a_j, q_i)\, q_i = p_j \qquad j=1,2\cdots n.$$

$$r_{jj} = \|p_j\|_2 \qquad r_{jj} q_j = p_j \Rightarrow q_j = p_j / r_{jj}.$$

QR method to solve linear system:

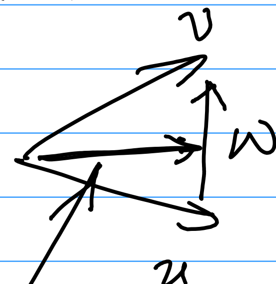
$$Ax = b$$

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n}\\ a_{21} & \boxed{a_{22}} & \cdots & a_{2n}\\ a_{31} & a_{32} & \cdots & a_{3n}\\ \vdots & \vdots & \ddots & \\ a_{n1} & a_{n2} & & a_{nn}\end{pmatrix}\begin{pmatrix}x_1\\ x_2\\ \vdots\\ x_n\end{pmatrix}=\begin{pmatrix}b_1\\ b_2\\ \vdots\\ b_n\end{pmatrix}$$

Householder matrix. (reflection matrix.)

$H = I - 2ww^T$ $\qquad \|w\|_2 = 1.$

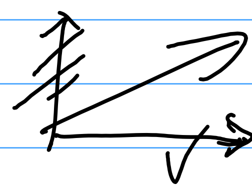
$u, v \in \mathbb{R}^n \qquad \|u\|_2 = \|v\|_2$



$$w = \frac{u-v}{\|u-v\|_2}$$

$Hu = v \qquad Hv = u.$

$(I - q_1 q_1^T) a_1$
$= a_1 - (q_1, a_1) q_1$



$(I - ww^T)v \Rightarrow$

$$Hv = (I - 2ww^T)v = v - 2\frac{(u-v)(u-v)^T}{\|u-v\|_2^2} v = u$$

H: orthogonal matrix

↓

symmetric

$H^T H$
$= (I - 2ww^T)(I - 2ww^T)$
$= I - 2ww^T - 2ww^T + 4w\underline{w^T w}w^T = I$

$$\begin{array}{ccc} u & & v \\ \| & & \| \\ \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{pmatrix} & \to & \begin{pmatrix} r_{11} \\ 0 \\ \vdots \\ 0 \end{pmatrix} \end{array}$$

$$\|v\|_2 = |r_{11}| = \|u\|_2 = \|a_1\|_2.$$

$$r_{11} = \pm\|a_1\|_2$$

$$\tilde{w}_1 = u - v. = \begin{pmatrix} a_{11} - r_{11} \\ a_{21} \\ a_{31} \\ \vdots \\ a_{n1} \end{pmatrix}$$

$$w_1 = \frac{\hat{w}_1}{\|\hat{w}\|_2} = \frac{u-v}{\|u-v\|_2}$$

$$a_{11} - r_{11} = ?$$

$$H_1 = I - 2w_1 w_1^T.$$

Next,

$$H_1 A x = H_1 b$$

$$\hat{a}_2 = \begin{pmatrix} a_{22} \\ a_{23} \\ \vdots \\ a_{2n} \end{pmatrix} \in \mathbb{R}^{n-1} \to \begin{pmatrix} r_2 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$$\| u_2 \qquad \| v_2$$

$$|r_2| = \pm\|\hat{a}_2\|_2$$

$$w_2 = \frac{u_2 - v_2}{\|u_2 - v_2\|_2} \in \mathbb{R}^{n-1}. \qquad H_2 = I - 2w_2 w_2^T \in \mathbb{R}^{(n-1)\times(n-1)}$$

$$\underline{H_2 H_1 A} x = H_2 H_1 b.$$

$$\downarrow$$

$$\begin{pmatrix} a_{33} \\ a_{43} \\ \vdots \\ a_{n3} \end{pmatrix} \in \mathbb{R}^{n-2} \to \begin{pmatrix} r_{33} \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in \mathbb{R}^{n-2}.$$

$$\underline{H_{n-1} \cdots H_2 H_1 A} x = \underline{H_{n-1} \cdots H_2 H_1} b$$

$$\|\|$$

$$R x = H_{n-1} \cdots H_2 H_1 b.$$

algorithm complexity $O(n^3)$.

Section 4. Stability Analysis

$$Ax = b. \qquad \text{(original)}$$

$$\tilde{A}\tilde{x} = \hat{b} \qquad \text{(practical)}$$

Step 1. Only b is perturbed.

$$A\hat{x} = \hat{b}$$

perturbation

Let $\hat{b} = b + \delta b$ $\qquad \delta b \in \mathbb{R}^n$.

$\tilde{x} = x + \delta x$. $\qquad \delta x$: error.

$$A(x+\delta x) = b + \delta b \qquad Ax + A\delta x = b + \delta b$$

$$\Rightarrow \quad A\delta x = \delta b \quad \Rightarrow \quad \delta x = \underline{A^{-1}\delta b}.$$

$\|\delta x\|$: absolute error.

relative error : $\dfrac{\|\delta x\|}{\|x\|}$

$$\|\delta x\| = \|A^{-1}\delta b\| \leq \|A^{-1}\| \cdot \|\delta b\|.$$

$$\frac{\|\delta x\|}{\|x\|} \leq \frac{\|A^{-1}\| \cdot \|\delta b\|}{\|x\|} \qquad Ax = b.$$

$$\|b\| = \|Ax\| \leq \|A\| \cdot \|x\| \quad \Rightarrow \quad \|x\| \geq \|b\| / \|A\|.$$

$$\Rightarrow \quad \frac{\|\delta x\|}{\|x\|} \leq \frac{\|A^{-1}\| \cdot \|\delta b\|}{\|b\| / \|A\|} = \|A\| \cdot \|A^{-1}\| \cdot \frac{\|\delta b\|}{\|b\|}$$

relative error $\qquad$ relative perturbation

condition number of A

norm of matrix A: related to vector norm

Definition of Vector Norm $\|\cdot\|$

1). $v \in \mathbb{R}^n$. $\|v\| \geq 0$ $\quad \|v\| = 0$ iff $v = 0$

non-negativity.

2). $\forall \lambda \in \mathbb{R}$. $\|\lambda v\| = |\lambda| \cdot \|v\|$

homogeneity

3). $\forall v, w \in \mathbb{R}^n$ $\quad \|v + w\| \leq \|v\| + \|w\|$

Example: $\|v\|_2 = \sqrt{\sum_{i=1}^{n} v_i^2}$

$\|v\|_\infty = \max_{1 \leq i \leq n} |v_i|$. $\quad \|v\|_1 = \sum_{i=1}^{n} |v_i|$.

The norms are equivalent

(Norm Equivalence)

$\|\cdot\|_a$. $\|\cdot\|_b$

There exists two positive constants $\mu, M > 0$

s.t. $\quad \mu \|v\|_b \leq \|v\|_a \leq M \|v\|_b \quad \forall v \in \mathbb{R}^n$

(μ. M are independent of v).

($\|v\|_1$ $\|v\|_\infty$) $\quad$ ($\|v\|_2$. $\|v\|_1$) $\quad$ ($\|v\|_2$, $\|v\|_\infty$)

$\|v\|_2 \sim \|v\|_\infty$

$$\|v\|_\infty = \max_{1 \leq i \leq n} |v_i| \leq \|v\|_2 = \sqrt{\sum_{i=1}^{n} v_i^2} \leq \sqrt{n \max_{1 \leq i \leq n} v_i^2} = \sqrt{n}\, \|v\|_\infty$$

$\mu = 1$ $\quad M = \sqrt{n}$

$$\|v\|_2 = \sqrt{\sum_{i=1}^{n} v_i^2} \qquad \left\{ \begin{array}{l} \|v\|_{2,n} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} v_i^2} \\ \|v\|_{1,n} = \frac{1}{n}\sum_{i=1}^{n} |v_i| \end{array} \right.$$

$$\|v\|_1 = \sum_{i=1}^{n} |v_i|$$

Matrix Norm : $A \in \mathbb{R}^{n\times n}$

1). $\|A\| \geq 0$. $\|A\| = 0$ iff $A = 0$ non-negativity.
2). $\|\lambda A\| = |\lambda| \cdot \|A\|$. $\lambda \in \mathbb{R}$. homogeneity.
3). $\|A+B\| \leq \|A\| + \|B\|$. triangle inequality.
4). $\|AB\| \leq \|A\| \|B\|$.

Frobenius norm : $A = (a_{ij})_{n\times n} \in \mathbb{R}^{n\times n}$

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

Vector-norm <u>induced</u> norm of a matrix. A

$$\|A\| = \sup_{\substack{v \in \mathbb{R}^n \\ v \neq 0}} \frac{\|Av\|}{\|v\|} \geq \frac{\|Aw\|}{\|w\|} \quad \forall w \in \mathbb{R}^n$$

Example : $\|A\|_2 = \sup \frac{\|Av\|_2}{\|v\|_2}$

$$\|A\|_\infty = \sup \frac{\|Av\|_\infty}{\|v\|_\infty} \qquad \|A\|_1 = \sup \frac{\|Av\|_1}{\|v\|_1}$$

Property : $\|A\| \geq \frac{\|Aw\|}{\|w\|} \Rightarrow \|Aw\| \leq \|A\| \cdot \|w\|$

Remark: We can also use other vector norm to define an induced matrix norm.

$$\|v\|_p = \left(\sum_{i=1}^{n} |v_i|^p\right)^{\frac{1}{p}} \qquad p \geq 1.$$

$$\|A\|_p = \sup \frac{\|Av\|_p}{\|v\|_p}.$$

Propositn: ① $\|A\|_\infty = \max\limits_{1\leq i\leq n} \sum\limits_{j=1}^{n} |a_{ij}|$.

$$\begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{pmatrix}$$

$$\| \quad \max_{1\leq i\leq n} \|r_i\|_1$$

$r_1, r_2 \cdots r_n \in \mathbb{R}^{1\times n}$

② $\|A\|_1 = \max\limits_{1\leq j\leq n} \sum\limits_{i=1}^{n} |a_{ij}| = \|A^T\|_\infty$

③ $\|A\|_2 = [\rho(A^TA)]^{\frac{1}{2}}$

proof: ①. $\|A\|_\infty = \sup \frac{\|Av\|_\infty}{\|v\|_\infty} = \sup\limits_{\|v\|_\infty=1} \|Av\|_\infty$

$$= \max_{1\leq i\leq n} \sum_{j=1}^{n} |a_{ij}|.$$

Let $v = (v_1, v_2 \cdots v_n)^T$ be a normalized vector with $\|v\|_\infty = \max\limits_{1\leq i\leq n} |v_i| = 1$.

$$\|Av\|_\infty = \max_{1\leq i\leq n} \left|\sum_{j=1}^{n} a_{ij} v_j\right| \leq \max_{1\leq i\leq n} \sum_{j=1}^{n} |a_{ij}| \cdot |v_j|$$

$$\leq \max_{1\leq i\leq n} \sum_{j=1}^{n} |a_{ij}|. \Rightarrow \|A\|_\infty \leq \max_{1\leq i\leq n} \sum_{j=1}^{n} |a_{ij}|. \quad (1).$$

Let k be the row index so that

$$\sum_{j=1}^{n} |a_{kj}| = \max_{1\leq i\leq n} \sum_{j=1}^{n} |a_{ij}|.$$

Choose a vector $w = (w_1, w_2 \cdots w_n)^T \in \mathbb{R}^n$ s.t.

$$w_j = \begin{cases} 1 & \text{if } a_{kj} \geq 0 \\ -1 & \text{otherwise} \end{cases} \qquad \|w\|_\infty = 1.$$

$$\|Aw\|_\infty \geq |(Aw)_k| = \left|\sum_{j=1}^n a_{kj} w_j\right| = \sum_{j=1}^n |a_{kj}|$$

$$\|A\|_\infty \geq \|Aw\|_\infty$$

$$= \max_{1\leq i\leq n} \sum_{j=1}^n |a_{ij}|$$

$$\Rightarrow \|A\|_\infty \geq \max_{1\leq i\leq n} \sum_{j=1}^n |a_{ij}| \qquad (2)$$

(1). (2) $\Rightarrow$ $$\|A\|_\infty = \max_{1\leq i\leq n} \sum_{j=1}^n |a_{ij}|.$$

(3). $\rho(B)$: spectral radius

$$\rho(B) = \max_{1\leq i\leq n} |\lambda_i(B)|$$

$$\|A\|_2 = [\rho(A^TA)]^{\frac{1}{2}}$$

$$\|A\|_2 = \sup \frac{\|Av\|_2}{\|v\|_2} \Rightarrow \|A\|_2^2 = \sup \frac{\|Av\|_2^2}{\|v\|_2^2}$$

$$= \sup \frac{(Av, Av)}{(v, v)} = \sup \frac{v^TA^TAv}{v^Tv}$$

A^TA : symmetric. non-negative definite

There exist a normalized mutually orthogonal basis, consisting of eigenvectors of A^TA.

associated with eigenvalues $\quad \Downarrow \quad \{r_1, r_2 \cdots r_n \in \mathbb{R}^n\}$

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$$

$$r_i^T r_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

For any vector v, we have

$$v = c_1 r_1 + \cdots + c_n r_n \qquad c_i \in \mathbb{R}.$$

$$v^T v = \sum_{i=1}^{n} c_i^2$$

$$A^T A v = \sum_{i=1}^{n} \lambda_i c_i r_i \ . \Rightarrow v^T A^T A v = \left(\sum_{i=1}^{n} c_i r_i\right)^T \sum_{i=1}^{n} \lambda_i c_i r_i$$

$$= \sum_{i=1}^{n} \lambda_i c_i^2$$

$$\frac{v^T A^T A v}{v^T v} = \frac{\sum_{i=1}^{n} \lambda_i c_i^2}{\sum_{i=1}^{n} c_i^2} \le \frac{\lambda_1 \sum_{i=1}^{n} c_i^2}{\sum_{i=1}^{n} c_i^2} = \lambda_1 = \rho(A^T A).$$

The equality holds when $c_2 = c_3 = \cdots = c_n = 0$

$$\|A\|_2^2 = \sup_{\|v\| \neq 0} \frac{v^T A^T A v}{v^T v} = \lambda_1 = \rho(A^T A)$$

$$\Rightarrow \|A\|_2 = \rho(A^T A)^{\frac{1}{2}}$$

#

property: Let $\|\cdot\|$ be an induced matrix norm. if $\|A\| < 1$ for some matrix A, then $I + A$ is invertible and

$$\|(I+A)^{-1}\| \le \frac{1}{1 - \|A\|}.$$

Proof: Assume $I+A$ is not invertible. there exists a non-zero vector $v \in \mathbb{R}$ s.t.

$$(I+A) v = 0$$

$$\Rightarrow v = -Av. \quad \Rightarrow \|v\| \le \|A\| \cdot \|v\| \Rightarrow 1 \le \|A\|$$

This is a contradiction

Let $C = (I+A)^{-1}$. $\quad (I+A) C = I$.

$C = I - AC \qquad \|C\| \le \|I\| + \|A\|\,\|C\| = 1 + \|A\|\cdot\|C\|$

$\Rightarrow \quad (1 - \|A\|)\,\|C\| \le 1$

$\Rightarrow \quad \|C\| \le \dfrac{1}{1 - \|A\|}$ (*)

Proposit: $\|A\|\cdot\|A^{-1}\|$: condition number of A

$\|A\|\cdot\|A^{-1}\| \ge 1$. in induced norm.

$1 = \|I\| = \|AA^{-1}\| \le \|A\|\cdot\|A^{-1}\|$

Example of an ill conditioned matrix

$$H = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \\ \frac{1}{4} & \frac{1}{5} & & \end{pmatrix} \qquad \text{Hilbert matrix}$$

$x \in \mathbb{R}^n \qquad Hx = b \qquad$ solve $Hy = b$ using LU or QR

$Hx \to b$

$\|x - y\| = ?$

Case II. matrix is perturbed

$\tilde{A} = A + \delta A \qquad \delta A \in \mathbb{R}^{n\times n}$.

$\hat{A}\hat{x} = b$.

$\Rightarrow \quad (A + \delta A)(x + \delta x) = b. \qquad \delta x \in \mathbb{R}^n$

$\cancel{Ax} + A\delta x + \delta A\, x + \delta A\cdot\delta x = \cancel{b}\ 0$

$(A + \delta A)\,\delta x = -\delta A\cdot x$.

$$\delta x = -(A+\delta A)^{-1}\cdot \delta A \cdot x$$

$$\|\delta x\| \leq \|(A+\delta A)^{-1}\|\cdot \|\delta A\|\cdot \|x\|.$$

$$\Rightarrow \quad \frac{\|\delta x\|}{\|x\|} \leq \|(A+\delta A)^{-1}\|\cdot \|\delta A\|. \qquad (3)$$

$$\Uparrow$$

$$A+\delta A = A(I + A^{-1}\delta A)$$

$$(A+\delta A)^{-1} = (I + A^{-1}\delta A)^{-1}\cdot A^{-1}$$

$$(3) \Rightarrow \quad \frac{\|\delta x\|}{\|x\|} \leq \|(I+A^{-1}\delta A)^{-1}\|\cdot \|A^{-1}\|\cdot \|\delta A\|.$$

Assume $\|A^{-1}\delta A\| < 1$

Assume $\|A^{-1}\|\cdot\|\delta A\| < 1$

$$\leq \frac{1}{1-\|A^{-1}\delta A\|}\|A^{-1}\|\cdot\|\delta A\|.$$

$$\leq \frac{1}{1-\|A^{-1}\|\cdot\|\delta A\|}\cdot\|A^{-1}\|\cdot\|\delta A\|$$

$$= \frac{1}{1-\|A^{-1}\|\cdot\|A\|\cdot\frac{\|\delta A\|}{\|A\|}}\cdot\|A^{-1}\|\cdot\|A\|\cdot\frac{\|\delta A\|}{\|A\|}$$

$$= \frac{\|A^{-1}\|\cdot\|A\|}{1-\|A^{-1}\|\cdot\|A\|\cdot\frac{\|\delta A\|}{\|A\|}}\cdot\frac{\|\delta A\|}{\|A\|} = \frac{\text{cond}(A)}{1-\text{cond}(A)\cdot\frac{\|\delta A\|}{\|A\|}}\cdot\frac{\|\delta A\|}{\|A\|}$$

The relative error also depends on the condition number of the matrix A.

Case III. Both b and A are perturbed.

$$(A+\delta A)(x+\delta x) = b+\delta b.$$

$$\Rightarrow \cancel{Ax} + \delta A\cdot x + (A+\delta A)\delta x = \cancel{b} + \delta b$$

$$(A+\delta A)\,\delta x = -\delta A\cdot x + \delta b.$$

$$\delta x = (A+\delta A)^{-1}\left[-\delta A\cdot x + \delta b\right]$$

$$\|\delta x\| \lesssim \|(A+\delta A)^{-1}\|\cdot\left(\|\delta A\|\cdot\|x\| + \|\delta b\|\right)$$

$$\frac{\|\delta x\|}{\|x\|} \lesssim \|(A+\delta A)^{-1}\|\cdot\left(\|\delta A\| + \frac{\|\delta b\|}{\|x\|}\right) \qquad (5)$$

$$b = Ax \Rightarrow \|b\| \le \|A\|\cdot\|x\| \Rightarrow \|x\| \ge \frac{\|b\|}{\|A\|}$$

$$(5) \le \|(A+\delta A)^{-1}\|\cdot\left(\|\delta A\| + \|A\|\cdot\frac{\|\delta b\|}{\|b\|}\right)$$

$$= \|(A+\delta A)^{-1}\|\cdot\|A\|\left(\frac{\|\delta A\|}{\|A\|} + \frac{\|\delta b\|}{\|b\|}\right)$$

$$\le \frac{\|A^{-1}\|\cdot\|A\|}{1-\|A^{-1}\|\|A\|\cdot\frac{\|\delta A\|}{\|A\|}}\cdot\left(\frac{\|\delta A\|}{\|A\|} + \frac{\|\delta b\|}{\|b\|}\right).$$

$\#$

Remark: 1. When the matrix A is symmetric and positive definite, the LU decomposit may take the form $LL^T = A$.

with
$$L = \begin{pmatrix} l_{11} & & & \\ l_{21} & l_{22} & & \\ \vdots & & \ddots & \\ l_{n1} & l_{n2} & \cdots & l_{nn} \end{pmatrix}$$

$$\begin{pmatrix} l_{11} & & & \\ l_{21} & l_{22} & & \\ \vdots & & \ddots & \\ l_{n1} & l_{n2} & & l_{nn} \end{pmatrix} \begin{pmatrix} l_{11} & l_{21} & l_{31} & \\ & l_{22} & l_{32} & \\ & & l_{33} & \\ & & & \ddots \; l_{nn} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots \\ a_{21} & & \\ a_{31} & & \\ \vdots & & \\ a_{n1} & & \end{pmatrix}$$

$$l_{11}^2 = a_{11} \quad \Rightarrow \quad l_{11} = \sqrt{a_{11}}$$

$$l_{i1} l_{11} = a_{i1} \qquad i = 1, 2 \cdots n$$

$$\Rightarrow \quad l_{i1} = a_{i1}/l_{11}$$

$$l_{21}^2 + l_{22}^2 = a_{22} \Rightarrow \quad l_{22} = \sqrt{a_{22} - l_{21}^2}$$

$$\cdots \cdots$$

The method is called the Cholosky decomposition

2. When the matrix A is tridiagonal

$$A = \begin{pmatrix} a_1 & c_1 & & & \\ b_1 & a_2 & c_2 & & \\ & b_2 & a_3 & \ddots & \\ & & \ddots & \ddots & c_{n-1} \\ & & & b_{n-1} & a_n \end{pmatrix} = LU.$$

$$L = \begin{pmatrix} 1 & & & \\ \beta_1 & 1 & & \\ & \beta_2 & 1 & \\ & & \ddots & \ddots \\ & & \beta_{n-1} & 1 \end{pmatrix} \qquad U = \begin{pmatrix} \alpha_1 & \gamma_1 & & & \\ & \alpha_2 & \gamma_2 & & \\ & & \alpha_3 & \ddots & \\ & & & \ddots & \gamma_{n-1} \\ & & & & \alpha_n \end{pmatrix}$$

$$\alpha_1 = a_1 \qquad \beta_1 \alpha_1 = b_1 \quad \Rightarrow \quad \beta_1 = b_1/a_1$$

$\gamma_1 = c_1 \qquad \beta_1 \gamma_1 + \alpha_2 = a_2 \Rightarrow \alpha_2 = a_2 - \beta_1 c_1.$

$\beta_2 \alpha_2 = b_2 \Rightarrow \beta_2 = b_2 / \alpha_2.$

$\cdots \quad \cdots$

$$\boxed{LUx = f} \Rightarrow \begin{cases} Ly = f \\ Ux = y \end{cases}$$

The algorithm complexity is $O(n)$.