3.4 矩阵的条件数

定理：$A \in \mathbb{R}^{n\times n}$，$\det A \neq 0$，$x$，$x+\delta x$ 分别满足

$$Ax=b, \quad (A+\delta A)(x+\delta x) = b+\delta b$$

其中 $b\neq 0$. 设 $\|\delta A\|$ 充分小，使得

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

则有

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

其中的范数为任一种向量范数及其从属矩阵范数.

证明：首先由 $\|A^{-1}\|\,\|\delta A\|<1$ 可得

$A+\delta A$ 可逆 并且

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

那么

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

$$\Rightarrow \quad \delta x = (A+\delta A)^{-1}(b+\delta b-(A+\delta A)x)$$

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

$$\Rightarrow \quad \|\delta x\| \le \frac{\|A^{-1}\|}{1-\|A^{-1}\|\,\|\delta A\|}\left(\frac{\|\delta A\|}{\|A\|}\|A\|\,\|x\| + \frac{\|\delta b\|}{\|b\|}\|A\|\,\|x\|\right)$$

$$\Rightarrow \quad \frac{\|\delta x\|}{\|x\|} \le \frac{\|A\|\,\|A^{-1}\|}{1-\|A^{-1}\|\,\|\delta A\|}\left(\frac{\|\delta A\|}{\|A\|}+\frac{\|\delta b\|}{\|b\|}\right)$$

定义: $A\in\mathbb{R}^{n\times n}$, $\det A\neq 0$, 对任一种从属矩阵范数

$$\mathrm{cond}(A) = \|A\|\,\|A^{-1}\|$$

称为矩阵 A 的条件数

矩阵的条件数具有以下性质:

(1) $\mathrm{cond}(A)\geq 1$, $\mathrm{cond}(A)=\mathrm{cond}(A^{-1})$

(2) $\mathrm{cond}(\alpha A)=\mathrm{cond}(A)$, $\forall\,\alpha\in\mathbb{R}$, $\alpha\neq 0$

(3) 若 A 正交, 则 $\mathrm{cond}(A)_2 = 1$

(4) $\|A\|_2 = \max\{y^TAx : \|x\|_2=\|y\|_2=1\}$

$$\mathrm{cond}(A)_2 = \mathrm{cond}(A^T)_2$$

(5) 若 U 为正交阵 则

$$\mathrm{cond}(A)_2 = \mathrm{cond}(UA)_2 = \mathrm{cond}(AU)_2$$

(6) 设 λ_1, λ_n 为 A 按模最大和最小的特征值, 则

$$\mathrm{cond}(A)\geq\frac{|\lambda_1|}{|\lambda_n|}$$

若A对称, 则 $\mathrm{cond}(A)_2=\dfrac{|\lambda_1|}{|\lambda_n|}$

(7)

$$\frac{1}{n}\mathrm{cond}(A)_2\leq\mathrm{cond}(A)_1\leq n\,\mathrm{cond}(A)_2$$

$$\frac{1}{n}\mathrm{cond}(A)_\infty\leq\mathrm{cond}(A)_2\leq n\,\mathrm{cond}(A)_\infty$$

$$\frac{1}{n^2}\mathrm{cond}(A)_1\leq\mathrm{cond}(A)_\infty\leq n^2\,\mathrm{cond}(A)_1$$

定理：$A\in\mathbb{R}^{n\times n}$，$\det A\neq 0$，则

$$\min_{A+\delta A\text{奇异}} \frac{\|\delta A\|_2}{\|A\|_2} = \frac{1}{\mathrm{Cond}(A)_2}$$

证明：只需证 $\min\limits_{A+\delta A\text{奇异}} \|\delta A\|_2 = \frac{1}{\|A^{-1}\|_2}$

若 $\|A^{-1}\|_2\|\delta A\|_2<1$ 则有 $A+\delta A$ 非奇异

故有

$$\min_{A+\delta A\text{奇异}} \|\delta A\|_2 \geqslant \frac{1}{\|A^{-1}\|_2}$$

另一方面 由从属范数的定义知 存在 $x\in\mathbb{R}^n$，$\|x\|_2=1$ 使得

$$\|A^{-1}x\|_2 = \|A^{-1}\|_2$$

令 $y=\frac{A^{-1}x}{\|A^{-1}x\|_2}=\frac{A^{-1}x}{\|A^{-1}\|_2}$，$\delta A = -\frac{xy^T}{\|A^{-1}\|_2}$

则有 $\|y\|_2=1$

$$(A+\delta A)y = Ay+(\delta A)y = \frac{x}{\|A^{-1}\|_2}-\frac{x}{\|A^{-1}\|_2}=0$$

$\Rightarrow$ $A+\delta A$ 奇异

又 $\|\delta A\|_2 = \max\limits_{\|z\|_2=1}\left\|\frac{xy^T}{\|A^{-1}\|_2}z\right\|_2 = \frac{\|x\|_2}{\|A^{-1}\|_2}\max\limits_{\|z\|_2=1}|y^Tz|$

$= \frac{1}{\|A^{-1}\|_2}$

$$\Rightarrow \min_{A+\delta A\text{奇异}} \|\delta A\|_2 \leq \left\|\frac{xy^T}{\|A^{-1}\|_2}\right\|_2 = \frac{1}{\|A^{-1}\|_2}$$

定理：$A\in\mathbb{R}^{n\times n}$，$b\in\mathbb{R}^n$，$b\neq 0$，A 非奇异，x 是 $Ax=b$ 的解，$\tilde{x}$ 是方程组的一个近似解，$r=b-A\tilde{x}$，则有

$$\frac{1}{\operatorname{cond}(A)}\frac{\|r\|}{\|b\|}\le\frac{\|\tilde{x}-x\|}{\|x\|}\le\operatorname{cond}(A)\frac{\|r\|}{\|b\|}$$

3.5 列主元素消去法的舍入误差分析（扩展内容）

定理：若 $|\delta_i|\le\varepsilon$ 且 $n\varepsilon\le 0.01$ 那么

$$1-n\varepsilon\le\prod_{i=1}^{n}(1+\delta_i)\le 1+1.01n\varepsilon$$

或写成

$$\prod_{i=1}^{n}(1+\delta_i)=1+\delta,\quad |\delta|\le 1.01n\varepsilon$$

证明：由 $|\delta_i|\le\varepsilon$ 得

$$(1-\varepsilon)^n\le\prod_{i=1}^{n}(1+\delta_i)\le(1+\varepsilon)^n$$

利用函数 $(1-x)^n$ 的 Taylor 展开

$$(1-x)^n=1-nx+\frac{n(n-1)}{2}(1-\theta x)^{n-2}x^2$$

得

$$1-nx\le(1-x)^n$$

$$\Rightarrow\quad 1-1.01n\varepsilon\le 1-n\varepsilon\le(1-\varepsilon)^n$$

由 e^x 的 Taylor 展开

$$e^x=1+x+\frac{x^2}{2}+\frac{x^3}{3!}+\cdots$$
$$=1+x+\frac{x}{2}\cdot x\cdot\left(1+\frac{x}{3}+\frac{2x^2}{4!}+\cdots\right)$$

若 $0 \le x \le 0.01$ 有

$$1+x \le e^x \le 1+x+\frac{0.01}{2} x e^x \le 1+1.01x$$

这里用到了 $e^{0.01}<2$. 由此可得

$$(1+\varepsilon)^n \le e^{n\varepsilon} \le 1+1.01n\varepsilon$$

例: $x, y \in F^n$, F 为浮点数集合, 估计 $|fl(x^Ty)-x^Ty|$ 的上界

令 $S_1 = fl(x_1y_1) = x_1y_1(1+\delta_1)$, $|\delta_1| \le \varepsilon$

$$S_k = fl(S_{k-1}+fl(x_ky_k))$$
$$= (S_{k-1}+x_ky_k(1+\delta_k))(1+\gamma_k), \quad |\delta_k|, |\gamma_k| \le \varepsilon$$

于是有

$$fl(x^Ty) = S_n = \sum_{i=1}^n x_iy_i(1+\delta_i)\prod_{j=i}^n(1+\gamma_j)$$

$$= \sum_{i=1}^n (1+\alpha_i)x_iy_i$$

其中 $\alpha_i = (1+\delta_i)\prod_{j=i}^n(1+\gamma_j)-1$

$$\Rightarrow |fl(x^Ty)-x^Ty| \le \sum_{i=1}^n |\alpha_i||x_iy_i| \le 1.01n\varepsilon\sum_{i=1}^n|x_iy_i|$$

类似可得 $A, B \in F^{n\times n}$, $\alpha \in F$

$$fl(\alpha A) = \alpha A+E, \quad |E| \le \varepsilon|\alpha A|$$
$$fl(A+B) = (A+B)+E, \quad |E| \le \varepsilon|A+B|$$
$$fl(AB) = AB+E, \quad |E| \le 1.01n\varepsilon|A||B|$$

引理：设 $A=[a_{ij}]\in F^{n\times n}$ 有三角分解且 $1.01n\varepsilon\le 0.01$，则用 Gauss 消去法计算得到的单位下三角矩阵 $\tilde{L}$ 与上三角矩阵 $\tilde{U}$ 满足

$$\tilde{L}\tilde{U}=A+E$$

其中 $$|E|\le 2.05n\varepsilon|\tilde{L}||\tilde{U}|$$

证明：设 $\tilde{U}=[\tilde{u}_{ij}]$ $\tilde{L}=[\tilde{l}_{ij}]$ 由 Gauss 消去法的实现过程知

$$a_{ij}^{(0)}=a_{ij}$$

$$a_{ij}^{(k)}=fl(a_{ij}^{(k-1)}-fl(\tilde{l}_{ik}\tilde{u}_{kj})),\ k=1,\cdots,i-2$$

$$\tilde{u}_{ij}=a_{ij}^{(i-1)}=fl(a_{ij}^{(i-2)}-fl(\tilde{l}_{i,i-1}\tilde{u}_{i-1,j})),\quad i\le j$$

$$\Rightarrow\ a_{ij}^{(k)}=(a_{ij}^{(k-1)}-(\tilde{l}_{ik}\tilde{u}_{kj})(1+\delta_k))(1+\gamma_k)$$

其中 $|\delta_k|\le\varepsilon$，$|\gamma_k|\le\varepsilon$，$k=1,\cdots,i-1$

由此可得

$$\tilde{u}_{ij}=a_{ij}(1+\alpha_i)-\sum_{k=1}^{i-1}(\tilde{l}_{ik}\tilde{u}_{kj})(1+\alpha_k)$$

其中 $|\alpha_k|\le 1.01n\varepsilon$

$$\Rightarrow\ a_{ij}=\frac{\tilde{u}_{ij}}{1+\alpha_i}+\sum_{k=1}^{i-1}(\tilde{l}_{ik}\tilde{u}_{kj})\frac{1+\alpha_k}{1+\alpha_i}$$

$$=\sum_{k=1}^{i}\tilde{l}_{ik}\tilde{u}_{kj}-e_{ij}$$

其中 $\tilde{l}_{ii}=1$，$e_{ij}=(\tilde{l}_{ii}\tilde{u}_{ij})\frac{\alpha_i}{1+\alpha_i}+\sum_{k=1}^{i-1}(\tilde{l}_{ik}\tilde{u}_{kj})\frac{\alpha_i-\alpha_k}{1+\alpha_i}$

注意到 $|\alpha_k| \le 1.01n\varepsilon < 0.01$，则有

$$|e_{ij}| \le \frac{2.02n\varepsilon}{1-0.01}\sum_{k=1}^{i}|\tilde{l}_{ik}||\tilde{u}_{kj}| \le 2.05n\varepsilon\sum_{k=1}^{i}|\tilde{l}_{ik}||\tilde{u}_{kj}|$$

对于 $i>j$ 利用 $\tilde{l}_{ij}$ 的计算过程

$$a_{ij}^{(0)} = a_{ij}$$

$$a_{ij}^{(k)} = fl(a_{ij}^{(k-1)} - fl(\tilde{l}_{ik}\tilde{u}_{kj})), \quad k=1,2,\cdots,j-1$$

$$\tilde{l}_{ij} = fl(a_{ij}^{(j-1)}/\tilde{u}_{jj})$$

类似分析可证

$$a_{ij} = \sum_{k=1}^{j}\tilde{l}_{ik}\tilde{u}_{kj} - e_{ij}$$

$$|e_{ij}| \le 2.05n\varepsilon\sum_{k=1}^{j}|\tilde{l}_{ik}||\tilde{u}_{kj}|$$

注意到交换矩阵的行不引入舍入误差. 则有

推论：$A\in F^{n\times n}$ 非奇异，$1.01n\varepsilon \le 0.01$. 则用选列主元的 Gauss 消去法计算得到的单位下三角阵 $\tilde{L}$，上三角阵 $\tilde{U}$ 以及排列方阵 $\tilde{P}$ 满足.

$$\tilde{L}\tilde{U} = \tilde{P}A + E$$

其中 $|E| \le 2.05n\varepsilon|\tilde{L}||\tilde{U}|$

引理：三角阵 $S \in F^{n\times n}$ 非奇异，$1.01n\varepsilon \le 0.01$

则用解三角方程组的方法计算 $Sx=b$ 得到的解 $\hat{x}$ 满足

$$(S+H)\hat{x}=b$$

其中 $|H| \le 1.01n\varepsilon|S|$

证明：对 n 用数学归纳法. 不妨设 S 为下三角

$n=1$ 时显然成立.

假设对于 $n-1$ 阶下三角矩阵引理成立.

设 $b=\begin{bmatrix} b_1 \\ c \end{bmatrix}$，$\hat{x}=\begin{bmatrix} \hat{x}_1 \\ \tilde{y} \end{bmatrix}$

$$S=\begin{bmatrix} s_{11} & 0 \\ s_1 & S_1 \end{bmatrix},$$

$S_1 \in F^{(n-1)\times(n-1)}$，$c \in F^{n-1}$，$\tilde{y} \in F^{n-1}$

则有

$$\hat{x}_1 = fl\left(\frac{b_1}{s_{11}}\right) = \frac{b_1}{s_{11}(1+\delta_1)}, \quad |\delta_1| \le \varepsilon$$

$\tilde{y}$ 是 $S_1 y = fl(c-\hat{x}_1 s_1)$ 的数值解

由归纳假设得 $\tilde{y}$ 满足

$$(S_1+H_1)\tilde{y} = fl(c-\hat{x}_1 s_1)$$

其中

$$|H_1| \le 1.01(n-1)\varepsilon|S_1|$$

$$fl(c-\hat{x}_1 s_1) = fl(c - fl(\hat{x}_1 s_1))$$
$$= (I+D_\gamma)^{-1}(c - \hat{x}_1 s_1 - \hat{x}_1 D_\delta s_1)$$

其中

$$D_\gamma = diag(\gamma_2 \cdots \gamma_n), \quad D_\delta = diag(\delta_2 \cdots \delta_n)$$
$$|\gamma_i| \le \varepsilon, \quad |\delta_i| \le \varepsilon$$

于是

$$\hat{x}_1 s_1 + \hat{x}_1 D_\delta s_1 + (I+D_\gamma)(S_1+H_1)\hat{y} = c$$

从而有

$$(S+H)\hat{x} = b$$

其中

$$H = \begin{bmatrix} \delta_1 s_{11} & 0 \\ D_\delta s_1 & H_1 + D_\gamma(S_1+H_1) \end{bmatrix}$$

$$\Rightarrow \quad |H| \le \begin{bmatrix} \varepsilon|s_{11}| & 0 \\ \varepsilon|s_1| & |H_1| + \varepsilon(|S_1|+|H_1|) \end{bmatrix}$$

$$\le \varepsilon \begin{bmatrix} |s_{11}| & 0 \\ |s_1| & (1.01(n-1)+1+1.01(n-1)\varepsilon)|S_1| \end{bmatrix}$$

$$\le 1.01 n \varepsilon |S|$$

最终选列主元的 Gauss 消去法得到的解通过求解三角方程组

$$\tilde{L}y = \tilde{P}b, \quad \tilde{U}x = y$$

得到

由上述定理知 $\hat{x}$ 满足

$$(\tilde{L}+F)(\tilde{U}+G)\hat{x} = \tilde{P}b$$

即 $$(\tilde{L}\tilde{U} + F\tilde{U} + \tilde{L}G + FG)\hat{x} = \tilde{P}b$$

其中

$$|F| \le 1.01n\varepsilon|\tilde{L}|, \quad |G| \le 1.01n\varepsilon|\tilde{U}|$$

将 $\tilde{L}\tilde{U} = \tilde{P}A + E$ 代入得

$$(A+\delta A)\hat{x} = b$$

其中

$$\delta A = \tilde{P}^T(E + F\tilde{U} + \tilde{L}G + FG)$$

$$\Rightarrow |\delta A| \le \tilde{P}^T(|E| + |F||\tilde{U}| + |\tilde{L}||G| + |F||G|)$$
$$\le 4.09n\varepsilon\tilde{P}^T|\tilde{L}||\tilde{U}|$$

$$\Rightarrow \|\delta A\|_\infty \le 4.09n\varepsilon\|\tilde{L}\|_\infty\|\tilde{U}\|_\infty$$

注意到 $|\tilde{L}| \le 1 \Rightarrow \|\tilde{L}\|_\infty \le n$

$$\|\tilde{U}\|_\infty \le n\max_{i,j}|\tilde{u}_{ij}| = n\rho\max_{i,j}|a_{ij}| \le n\rho\|A\|_\infty$$

其中

$$\rho = \max_{i,j}|\tilde{u}_{ij}| \Big/ \max_{i,j}|a_{ij}|$$

称为选列主元 Gauss 消去法的增长因子

最终可得 $\|\delta A\|_\infty \le 4.09\, n^3 \rho \varepsilon \|A\|_\infty$