Session 6

Xuanxi Zhang

1 root finding

1.1 fixed point iteration

1.1.1 existence of fixed point

A function $g(x)$ has a fixed point if there exists a x^* such that $g(x^*) = x^*$.
ex: $g_1(x) = \frac{1}{3}\ln(x+1) + \frac{2}{3}$ has fixed point.

1.1.2 stability of fixed point

$g : [a, b] \to \mathbb{R}$ is continuous and has a fixed point ξ. Then we call ξ is stable if for any x_0 sufficiently close to ξ, the fixed point iteration $x_{k+1} = g(x_k)$ converges to ξ.

Theorem Suppose that g is continuously differentiable with $|g'(\xi)| < 1$, then the fixed point iteration converges to ξ as $k \to \infty$ provided x_0 is sufficiently close to ξ.

ex: $\xi = \frac{\pi}{2}$ is a stable fixed point for $g_2(x) = x + \cos x$.

Theorm If $g(x)$ has a fixed point ξ in $[c, d]$ and has a lipshitz constant $L < 1$, then the fixed point iteration converges to ξ as $k \to \infty$ starting from any $x_0 \in [c, d]$.

ex: starting form $[\frac{\pi}{4}, \frac{3\pi}{4}]$, the fixed point iteration $x_{k+1} = g_2(x_k)$ will always converge.
ex: starting form $[0, e-1]$, the fixed point iteration $x_{k+1} = g_1(x_k)$ will always converge.

1.1.3 speed of convergence

The sequence $\{x_k\}_{k\geq 1}$ converges (at least) linearly if

$$\lim_{k\to\infty} \frac{|x_{k+1} - \xi|}{|x_k - \xi|} = \mu > 0$$

- $\mu = 0 \to$ super-linear convergence
- $\mu \in (0, 1) \to$ linear convergence
- $\mu = 1 \to$ sub-linear convergence

ex: $x_k = 1/k$ converges super-linearly to 0.
ex: $x_{k+1} = g_1(x_k)$ converge linearly since

$$\lim_{x\to\xi} \frac{|g_1(x) - \xi|}{|x - \xi|} = g_1'(\xi) \in (0, 1)$$

Assume $x_k \to \xi$ if

$$\lim_{k\to\infty} \frac{|x_{k+1} - p|}{|x_k - \xi|^p} = \mu$$

We call p the order of convergence and μ the rate of convergence.

1.2 3 methods

1.2.1 bisection

start with $f(a)f(b) < 0$ on $[a, b]$. converge linearly.
ex: f a continuous function admits a root ξ in $[a, b]$, then the bisection method will always converge to the root ξ. Wrong, for example $f(x) = x^2$.

1.3 newton

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$$

if f is twice continuously differentiable and $f'(\xi) \neq 0$, then the newton method converges quadratically. Otherwise, it is linear.
ex: For $f(x) = e^{3x-2} - x - 1$, newton method converges quadratically to the root ξ.

1.4 secant

$$x_{k+1} = x_k - \frac{f(x_k)(x_k - x_{k-1})}{f(x_k) - f(x_{k-1})}$$

converge at least linearly. The best convergence order is aproximately 1.6.

2 LU and solve linear system

LU decomposition: $A = LU$
Pivot LU decomposition: $PA = LU$

2.1 existence of LU and PLU

Theorem For $A \in \mathbb{R}^{n\times n}$, if every leading principle submatrix $A^{(k)}$ is non-singular $k = 1, \ldots, n-1$, then $A = LU$ exists with a lower unit triangular matrix L and upper triangular matrix U.

ex Following matrix do not have a LU decomposition. If change first and second row, then it has a LU decomposition.

$$A = \begin{bmatrix} 0 & 1 & 2 \\ 3 & -1 & 1 \\ 0 & 0 & 4 \end{bmatrix}$$

Theorem PLU always exists for any matrix $A \in \mathbb{R}^{n\times n}$.

2.2 solve linear system

To solve a linear system $Ax = b$, with the decomposition $PA = LU$, the steps are

1. $\tilde{b} = Pb$, it is a permutation of b
2. $Ly = \tilde{b}$, forward substitution
3. $Ux = y$, back substitution

To solve the inverse of A, we can solve n linear systems $Ax_i = e_i$ where e_i is the ith column of the identity matrix. Then x_i is the ith column of the inverse of A.

2.3 computing complexity

- Matrix times vector A*b: $2n^2 \sim \mathcal{O}(n^2)$
- LU decomposition: $\frac{2}{3}n^3 - \frac{1}{2}n^2 \sim \mathcal{O}(n^3)$
- back and forward substitution: $2n^2 \sim \mathcal{O}(n^2)$
- solve linear system given PLU (or LU): $\mathcal{O}(n^2)$
- solve the inverse of A: $\mathcal{O}(n^3)$
- by the way using cramer's rule to solve the inverse of A is $\mathcal{O}(n!)$, which is rouphly $\mathcal{O}(e^n)$

3 norm and condition number

3.1 definition

V is a vector space. If a function $\|\cdot\| : V \to \mathbb{R}$ satisfies

- Positive definiteness: $\|v\| = 0 \iff v = 0, \forall v \in V$.
- Absolute homogeneity: $\|\lambda v\| = |\lambda|\|v\|, \forall v \in V, \lambda \in \mathbb{R}$.
- Triangular inequality: $\|v + w\| \leq \|v\| + \|w\|, \forall v, w \in V$,

then $\|\cdot\|$ is a norm.

ex Is $\|\cdot\|_1\|\cdot\|_2$ a norm? No, absolute homogeneity is not satisfied.

vector norms:

- 1-norm: $\|x\|_1 = \sum_{i=1}^n |x_i|$
- 2-norm: $\|x\|_2 = \sqrt{\sum_{i=1}^n x_i^2}$
- ∞-norm: $\|x\|_\infty = \max_{1\leq i\leq n} |x_i|$

matrix norms:

- induced norm: $\|A\| = \max_{x\neq 0} \frac{\|Ax\|}{\|x\|} = \max_{\|x\|=1} \|Ax\|$
- 1-norm: $\|A\|_1 = \max_{1\leq j\leq n} \sum_{i=1}^n |a_{ij}|$
- ∞-norm: $\|A\|_\infty = \max_{1\leq i\leq n} \sum_{j=1}^n |a_{ij}|$
- 2-norm: $\|A\|_2 = \sqrt{\rho(A^T A)}$
- Frobenius norm: $\|A\|_F = \sqrt{\sum_{i=1}^n \sum_{j=1}^n a_{ij}^2}$

In a space with finite dimention, norms are all equavalent. That is, there exists $c_>0$ such that $\|x\|_a \leq c\|x\|_b$ and $\|x\|_b \leq c\|x\|_a$. Therefore, in practical problems, the choice of norm does not significantly affect the result, as they all lead to comparable outcomes.

matrix 2-norm has propety $\|A\|_2 = \|AQ_1\|_2 = \|Q_2 A\|_2$, where $A \in \mathbb{R}^{m\times n}$, $Q_1 \in \mathbb{R}^{n\times n}$, $Q_2 \in \mathbb{R}^{m\times m}$ are orthogonal matrices. Proof by definition.

For a digonal matrix D, $\|D\|_2 = \max_{1\leq i\leq n} |d_{ii}|$.

3.2 condition number

For a matrix A, the condition number is defined as

$$\kappa(A) = \|A\|\|A^{-1}\|$$

ex so $\kappa(A^{-1}) = \|A^{-1}\|\|A^{-1^{-1}}\| = \kappa(A)$

ex For a permutation matrix, $\kappa_2(A) = \kappa_2(PA)$ because P is also orthogonal.

4 least square

For $A \in \mathbb{R}^{m\times n}$, when there is no exat solution for $Ax = b$, in order to obtain a solution x such that $Ax \sim b$, we solve the least quare problem $\min_x \|Ax - b\|_2^2$, which is equal to solve normal quation $A^T Ax = A^T b$. The normal equation will always have a solution no matter what A is, because range of $A^T A$ is the same as range of A. The solution is unique if A has full column rank.

The QR decomposition for A is $A = QR$, where Q is orthogonal and R is upper triangular. If $m > n$ and A has full column rank, then $R = \begin{bmatrix} \hat{R} \\ 0 \end{bmatrix}$, also seperate $Q = \begin{bmatrix} \hat{Q} & \hat{\hat{Q}} \end{bmatrix}$ in the same way. Then we have $A = QR = \hat{Q}\hat{R}$, and

$$\min_x \|Ax - b\|_2 = \min_x \|QRx - b\|_2 = \min_x \|Rx - Q^T b\|_2 = \min_x \|\hat{R}x - \hat{Q}^T b\|_2 + \|\hat{\hat{Q}}^T b\|_2.$$

The least square problem is equivalent to solve $\hat{R}x = \hat{Q}^T b$.

note: Given QR, the flop to sovle least square is $\mathcal{O}(n^2)$

4.1 hoseholder

hoseholder transform $H = I - 2\frac{vv^T}{v^v}$ is a orthogonal transfrom.

note: $\|H\|_2 = \sqrt{\rho(H^T H)} = \sqrt{\rho(I)} = 1$ **note**: if $x \perp v$, Then $Hx = (I - 2\frac{vv^T}{v^T v})x = x - 0 = x$.