Lecture 14

Last time we introduced norms on $\mathbb{R}^n$.

We also studied matrix norms.

Recall a norm on $\mathbb{R}^n$ is a function $\| \ \|$ which associates to each vector $\bar{a}$ in $\mathbb{R}^n$, a real number $\|a\|$ such that

1) $\|\bar{a}\| \geq 0 \quad \forall \bar{a} \in \mathbb{R}^n$
$= 0$ iff $\bar{a} = 0$

2) for $\alpha \in \mathbb{R}$
$\|\alpha \bar{a}\| = |\alpha| \, \|\bar{a}\|$

3) (triangle inequality)
$\|\bar{a} + \bar{b}\| \leq \|\bar{a}\| + \|\bar{b}\|$

Examples of norms

$\bar{a} = (a_1, a_2, \ldots, a_n)$

$\|\bar{a}\|_2 = \sqrt{a_1^2 + a_2^2 + \ldots + a_n^2}$.

$$\|a\|_1 = |a_1| + |a_2| + \cdots + |a_n|$$

$$\|a\|_\infty = \max_{1 \le i \le n} |a_i|$$

Matrix Norms

A matrix norm is a real valued function defined over the set of $n \times n$ matrices such that for any $n \times n$ matrices A, B and real number α we have

(1) $\|A\| \geq 0 \quad \forall A$ and $\|A\| = 0 \Leftrightarrow A = 0$

(2) $\|\alpha A\| = |\alpha| \, \|A\|$

(3) $\|A + B\| \leq \|A\| + \|B\|$

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

Theorem Let $\|x\|$ be a norm in $\mathbb{R}^n$. Then for a $n \times n$ matrix A

$$\|A\| = \max_{\|x\|=1} \|Ax\|$$

defines a matrix norm on the set of $n \times n$-matrices

We have

$$\|A\bar{x}\| \leq \|A\| \|\bar{x}\| \quad \forall \bar{x} \in \mathbb{R}^n.$$

Example

1) $\|A\|_2 = \max_{\|x\|_2 = 1} \|Ax\|_2$

2-norm on a matrix

This is difficult to compute

(2) $\|A\|_\infty = \max_{\|x\|_\infty = 1} \|Ax\|_\infty$

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

(3) $\|A\|_1 = \max_{\|x\|_1 = 1} \|Ax\|_1$

Exercise ~~Example~~ $\|A\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^{n} |a_{ij}|$

If the matrix A is invertible then

$$\bar{x} = A^{-1}(Ax)$$

So $\|\bar{x}\| = \|A^{-1}(Ax)\| \leq \|A^{-1}\| \|Ax\|$

$$\frac{\|\bar{x}\|}{\|A^{-1}\|} \leq \|Ax\| \leq \|A\| \|x\|$$

Reasons for studying norms

We want to study errors in computing solutions to linear equations.

$$Ax = b$$ exact soln $\bar{x}_0$

Solving we get approximate solution $\hat{x}$

error : $\bar{e} = \bar{x}_0 - \hat{x}$

This error is unknown to us

"residual error"

$$\bar{r} = A\bar{e} = A\bar{x}_0 - A\hat{x} = b - A\hat{x}$$

$$\bar{r} = \bar{b} - A\hat{x}$$

This we can compute

Last time I gave examples which showed that the size of the residual $\bar{r}$ is not always a reliable indicator of the size of error $\bar{e}$.

It depends on the size of A and A^{-1}.

error $\quad \bar{e} = \bar{x}_0 - \hat{x}$

residual $\quad \bar{r} = A\bar{e}$

$\therefore \quad \bar{e} = \bar{A}^{-1}\bar{r}$

Recall we have proved that for an invertible matrix B

$$\frac{\|\bar{u}\|}{\|B^{-1}\|} \leq \|Bu\| \leq \|B\|\,\|u\|$$

apply this to $B = A^{-1}$ and $u = \bar{r}$

$$\frac{\|\bar{r}\|}{\|A\|} \leq \|e\| = \|A^{-1}r\| \leq \|A^{-1}\|\,\|\bar{r}\|$$

We further proved

$$\frac{1}{\|A\|\,\|A^{-1}\|}\,\frac{\|\bar{r}\|}{\|\bar{b}\|} \leq \frac{\|\bar{e}\|}{\|\bar{x}_0\|} \leq \|A\|\,\|A^{-1}\|\,\frac{\|\bar{r}\|}{\|\bar{b}\|}$$

$\text{Cond}(A) = \|A\|\,\|A^{-1}\|$ condition number of A

We also proved

Theorem A $n \times n$ invertible matrix

$$\frac{1}{\text{cond}(A)} = \min \left\{ \frac{\|A-B\|}{\|A\|} \;\middle|\; B \text{ is NOT invertible} \right\}$$

Today we study

Perturbations of Linear systems of equations

If the linear system $A\bar{x} = \bar{b}$ derives from a practical problem, we must expect the coefficients of this system to be subject to error either because they result from other calculation or from physical measurement.

Hence assuming for example the RHS is accurate we are in effect solving the system

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

instead of $Ax = b$

where $A = \hat{A} + E$.

$E \rightarrow$ matrix that contains the errors in the coefficients.

Even if all calculations are carried out exactly we only have a solution of $\hat{A}\hat{x} = b$ rather than $Ax = b$

$A\bar{x}_0 = \bar{b}$ exact solution

$\bar{x}_0 = A^{-1}\bar{b}$

$= A^{-1}\hat{A}\hat{x}_0$

$= A^{-1}(A + E)\hat{x}_0$

$\hat{x}_0$ exact sol? of $\hat{A}\hat{x}_0 = b$

$A = \hat{A} + E$

So $\bar{x}_0 = \hat{x}_0 - A^{-1}E\hat{x}_0$

$\|\bar{x}_0 - \hat{x}_0\| \leq \|A^{-1}\| \|E\| \|\hat{x}_0\|$

$= \|A^{-1}\| \|A\| \frac{\|E\|}{\|A\|} \|\hat{x}_0\|$

So $\frac{\|\bar{x}_0 - \hat{x}_0\|}{\|\hat{x}_0\|} \leq \text{cond}(A) \frac{\|E\|}{\|A\|}$

Thus if the coefficients of the linear system $A\bar{x} = \bar{b}$ are known to be correct only to about 10^{-s} (relative to size of A) and $\text{cond}(A) \approx 10^{t}$

Then there is no point in calculating the solution to a relative accuracy of 10^{t-s}

— ∝ — ∝ — ∝ —

Quite loosely we say the linear system $A\bar{x} = \bar{b}$ is "ill-conditioned" if $\text{Cond}(A)$ is large.

Iterative improvement of solution

Let $e = x_0 - \hat{x}^{(1)}$ be the (unknown) error in the approximate solution $\hat{x}^{(1)}$ for $Ax = b$

$$Ae = r = b - A\hat{x}^{(1)} \qquad (*)$$

Let $\hat{e}^{(1)}$ be approximate solution of $(*)$.

Then $\hat{e}^{(1)}$ need not equal e.

But at the very least $\hat{e}^{(1)}$ gives an indicator of the size of e

If $\dfrac{\|\hat{e}^{(1)}\|}{\|\hat{x}^{(1)}\|} \approx 10^{-s}$ then we can conclude that the first "s" decimals of $\hat{x}^{(1)}$ to agree with that of the exact answer x_0

We would then also expect $\hat{e}^{(1)}$ to be a more accurate approximation to e

So $\hat{x}^{(2)} = \hat{x}^{(1)} + \hat{e}^{(1)}$ is a better approximation to $\bar{x}_0$ than $\hat{x}^{(1)}$.

We can now, if necessary compute the new residual

$$\bar{r} = \bar{b} - A\hat{x}^{(2)}$$

and solve $Ae = \bar{r}$ to obtain a new correction $\hat{e}^{(2)}$ and a new approximation $\hat{x}^{(3)} = \hat{x}^{(2)} + \hat{e}^{(2)}$ to $\bar{x}_0$.

One keeps on doing this.

The number of places in agreement to the successive approximations $\hat{x}^{(1)}, \hat{x}^{(2)}, \ldots$ as well as an examination of the successive residuals should give an indication of the accuracy of these approximate solutions.

One normally carries out this iteration until $\frac{\| \hat{e}^{(k)} \|}{\| \hat{x}^{(k)} \|} \approx 10^{-t}$ if t decimal places are carried during the calculation

Remarks

1) Iteration steps increase with cond(A)

2) If cond(A) is "very large" then the corrections $\hat{e}^{(1)}, \hat{e}^{(2)}, \ldots$, may never decrease in size

Remark

3) For the success of iterative improvement it is absolutely mandatory that residuals be computed as accurately as possible.

If as is usual, floating pt arithmetic is used, the residual should always be calculated in "double-precision" arithmetic.

Determinants

If $A = (a_{ij})$ is a $n \times n$ matrix

$$\det(A) = \sum \sigma_P \, a_{1,P_1}, a_{2,P_2} \cdots a_{n,P_n}$$

where the sum is taken over all $n!$ permutations of degree n and $\sigma_P = 1$ or -1 depending on whether P is even or odd

This definition is NOT GOOD for PRACTICAL purposes as $n!$ grows very large with n. However one can derive some rules from this definition and then get a simpler rule to calculate determinant.

The determinant of a matrix is of importance because of the following theorem

Theorem Let A be a $n \times n$ matrix.

Then A is invertible if and only if

$$\det(A) \neq 0.$$

Rule 1 If A is an upper (or lower) triangular matrix then

$$\det(A) = a_{11} a_{22} \cdots a_{nn}$$

Proof Assume A is upper triangular

If p is any permutation other than identity then for some i, $p_i < i$

$$\Rightarrow a_{i,p_i} = 0$$

$$\text{So } a_{1,p_1} \cdots a_{n,p_n} = 0$$

$$\therefore \det(A) = \sum \sigma_p a_{1,p_1} \cdots a_{n,p_n}$$

$$= a_{11} a_{22} \cdots a_{nn}.$$

Rule 2

If P is a $n \times n$ - permutation matrix given by $P_{ij} = {}_{i}P_{j}$ $\quad j = 1, 2, \ldots, n$

for some permutation P

Then $\det(P) = \begin{cases} 1 & \text{if } P \text{ is even} \\ -1 & \text{if } P \text{ is odd} \end{cases}$

Rule 3

If the matrix B results from the matrix A by the interchange of two columns (rows) of A then

$$\det(B) = -\det(A)$$

Rule 4 If the matrix B is formed by multiplying all entries of one column (row) of A by the same number α then

$$\det B = \alpha \det A.$$

Rule 5

Suppose 3 matrices A_1, A_2, A_3 differ in one column (row), say the j^{th} and the j^{th} column (row) of A_3 is the vector sum of the j^{th} column of A_1 and A_2

Then $\det(A_1) + \det(A_2) = \det(A_3)$

Theorem If A and B are $n \times n$ matrices then $\det(AB) = \det A \cdot \det B$.

Practical method to find determinant

Give factorization

$$PA = LU$$

P permutation matrix

$(\det P.) = (-1)^i$ $\quad i$ = # number of row interchanges $\in E$

$\det A = (-1)^i 1 \cdot u_{11} \cdots u_{nn}$

Exercise $A = LU$ A invertible

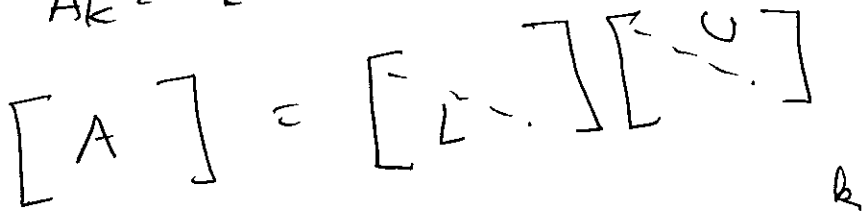
factorization

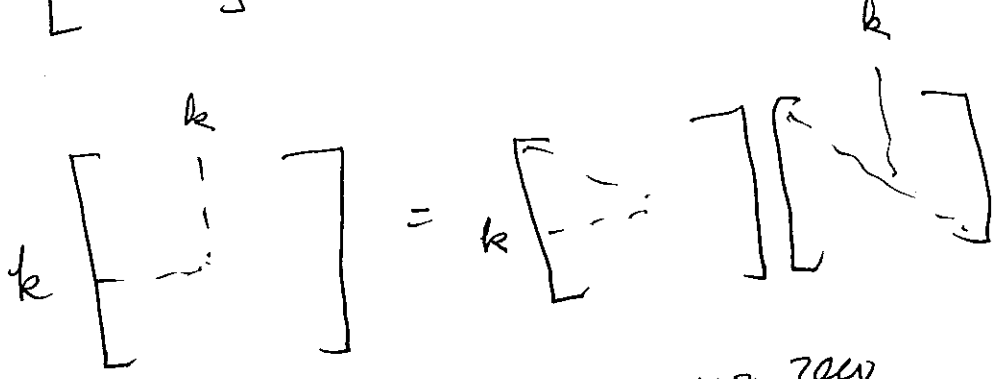
$\det A_k$ = principal submatrix of A formed by the first k-rows and the first k-columns

Then $\det(A_k) \neq 0$ for $k = 1, 2, \ldots, n$

Proof $A = LU$

$A_k = L_k U_k$





diagonal entries of L_k, U_k are non-zero

So $\det A_k = \det L_k \det U_k \neq 0$

Remark A true definite matrix can be factored into LU

One can prove a symmetric matrix A is true definite $\Leftrightarrow$ $\det(A_k) > 0$ for $k = 1, 2, \dots, n$

Exercise Assume A positive definite

Show $\det A_k > 0$ for $k = 1, 2, \dots, n$

(Hints: Use Cholesky decomposition)

Computing A^{-1}

There is usually no good reason for one calculating the inverse

However if one has to calculate A^{-1} find PLU decomposition of A

Solve $Ax_j = e_j$ $\quad e_j$ j^{th} unit vector

$PAx_j = Pe_j$

$LUx_j = Pe_j$

Solve it for all $j = 1, 2, \dots, n$