Math 5600 6/25/14

- Error Analysis

$Ax = b \rightsquigarrow \hat{x}$

In general, $x \neq \hat{x}$, for example, because of

- round-off errors
- we don't know b exactly
- we don't know A exactly

$\hat{x}$ is the "computed solution"

$e = x - \hat{x}$ is the 'error'

Of course we don't know the error

But we can compute the residual

$$r = b - A\hat{x} = Ax - A\hat{x} = A(x - \hat{x}) = Ae$$

general phenomenon for linear problems:
The error satisfies the same equation as the solution, except that the right hand side is replaced with the residual

- Just replace the matrix A with the appropriate linear operator

- we need to relate the unknown relative error $\frac{\|e\|}{\|x\|}$

to the computable $\frac{\|r\|}{\|b\|}$

- Let's digress into discussing norms for a moment

- we are familiar with

$$\|x\| = \sqrt{\sum_{i=1}^{n} x_i^2} \qquad x \text{ in } \mathbb{R}^n$$

- This is only one of infinitely many norms, and we will henceforth denote it by

$$\|x\|_2 = \sqrt{\sum x_i^2}$$

and call it the 2-norm.

- In general a norm is a function that associates a real number with a vector such that the following properties hold:

$$\left.\begin{array}{l} \|x\| \geq 0 \qquad \|x\| = 0 \Rightarrow x = 0 \\ \|kx\| = |k|\,\|x\| \\ \|x+y\| \leq \|x\| + \|y\| \quad \text{(triangle inequality)} \end{array}\right\} (*)$$

- Examples for norms include $\quad v = \begin{bmatrix} 3 \\ -4 \end{bmatrix}$

$$\|x\|_1 = \sum_{i=1}^{n} |x_i| \qquad \|v\|_1 = 7$$

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p} \qquad \|v\|_2 = 5$$

$$\|x\|_\infty = \max_{i=1,\dots,n} |x_i| \qquad \|v\|_\infty = 4$$

$\|x\|_W = \|Wx\|$ for any vector norm $\| \ \|$ and any non-singular $n \times n$ matrix W

- Norms could be defined similarly for matrices, treating the matrix as a giant vector.

- For example, the Frobenius norm is defined as

$$\|A\|_F = \left(\sum_{i,j=1}^{n} a_{ij}^2 \right)^{1/2}$$

- So we'd require

$\|A\| \geq 0 \qquad \|A\| = 0 \Rightarrow A = 0$

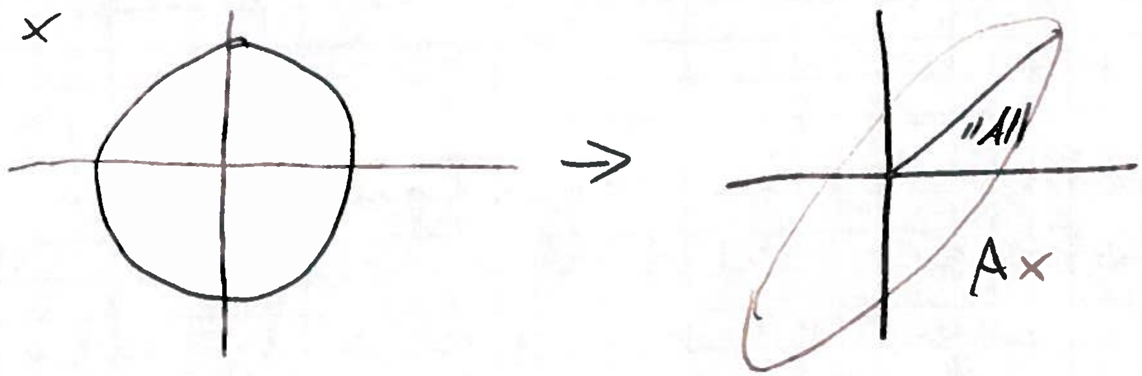
$\|kA\| = |k| \, \|A\|$

$\|A+B\| \leq \|A\| + \|B\|$

- we do require these properties but they are not enough, they don't tell us anything about the norms of matrix products.

- Let $\| \ \|$ be a vector norm. The "induced matrix norm" (or "associated operator norm") is defined by

$$\|A\| = \max_{x \neq 0} \frac{\|Ax\|}{\|x\|} = \max_{\|x\|=1} \|Ax\|$$



- Example

$$\|A\|_\infty = \max_{\|x\|_\infty = 1} \|Ax\|_\infty = \max_i \sum_{j=1}^{n} |a_{ij}|$$

("maximum row sum")

- Simple argument

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

(so k is the index of a row with largest sum. There may be several, then we'll pick one arbitrarily)

- suppose $\|x\|_\infty = \max_{i=1,\dots,n} |x_i| = 1$

- Then, for each $i = 1, \dots, n$

$$|(Ax)_i| = \left|\sum_{j=1}^{n} a_{ij} x_j\right| \le \sum_{j=1}^{n} |a_{ij} x_j| \le \sum_{j=1}^{n} |a_{ij}| \le \rho$$

$\Rightarrow \|A\| \le \rho$

- on the other hand, if we pick

$$x = [\operatorname{sign} a_{kj}]_{j=1,\dots,n} \qquad \|x\|_\infty = 1$$

- we get

$$|(Ax)_k| = \sum_{j=1}^{n} a_{kj} \operatorname{sign} a_{kj}$$

$$= \sum_{j=1}^{n} |a_{kj}| = \rho$$

$\Rightarrow \|Ax\|_\infty \geq \rho$

$\Rightarrow \|A\| = \rho$

For example $\begin{bmatrix}-1\\1\end{bmatrix}$

$$A = \begin{bmatrix}1 & -2\\-3 & 4\end{bmatrix}\begin{bmatrix}-3\\7\end{bmatrix}$$

$$\|A\|_\infty = 7$$

- more examples (exercise)

$$\|A\|_1 = \max_j \sum_{i=1}^{n} |a_{ij}| \quad \text{max column sum}$$

$$\|A\|_2 = \sqrt{\rho(A^TA)} \quad \rho: \text{spectral radius}$$

max |eigenvalue|

- Major property of induced matrix norms: the norm of the product is no larger than the product of the norms.

$$\|Ax\| = \left\|A\frac{x}{\|x\|}\right\|\|x\| \leq \max_{y\neq 0}\left\|A\frac{y}{\|y\|}\right\|\|x\| = \|A\|\|x\|$$

$$\|AB\| = \|ABx\| \leq \|A\|\|Bx\| \leq \|A\|\|B\|\|x\| =$$

for some x with $\|x\|=1$

$$= \|A\|\|B\|$$

OK now we can put this all together

$$A x = b \qquad \|b\| \le \|A\|\|x\| \qquad (1)$$

$$A^{-1} b = x \qquad \|x\| \le \|A^{-1}\|\,\|b\| \qquad (2)$$

$$A e = r \qquad \|r\| \le \|A\|\|e\| \qquad (3)$$

$$A^{-1} r = e \qquad \|e\| \le \|A^{-1}\|\|r\| \qquad (4)$$

combine 4 and 1

$$\frac{\|e\|}{\|A\|\|x\|} \le \frac{\|A^{-1}\|\|r\|}{\|b\|}$$

$$\boxed{\frac{1}{\|A\|\|A^{-1}\|}\frac{\|r\|}{\|b\|} \le \frac{\|e\|}{\|x\|} \le \|A\|\|A^{-1}\|\frac{\|r\|}{\|b\|}} \qquad (*)$$

from (2) and (3)   from (1) and (4)

- worth the deepest study.

Notes

$\|A\|\|A^{-1}\|$ is the condition number of A with respect to the underlying vector norm

- (*) works for any vector norm and the induced matrix norm

- large condition number: ill-conditioned
- small condition number: well-conditioned

- $\|A\| \geq \rho(A) = \max\limits_{Ax=\lambda x} |\lambda|$

$$\|A^{-1}\| \geq \frac{1}{\min\limits_{Ax=\lambda x} |\lambda|}$$

$$\|A\| \, \|A^{-1}\| \geq \frac{\max |\lambda|}{\min |\lambda|}$$

- A matrix is ill-conditioned if its eigenvalues are widely spread

- The inequalities (*) are <u>sharp</u>.

For any A and any norm you can find x and e so that one (or the other) is satisfied with equality

- $\frac{\|r\|}{\|b\|}$ usually about τ (round-off unit)

$$\tau \approx 10^{-16}$$

- if $\|A\|\|A^{-1}\| = 10^{p}$ you loose p digits!

- Example: Hilbert Matrix

$$p(x) = \sum_{j=0}^{n} \alpha_j x^j dx \qquad \int_0^1 (f(x)-p(x))^2 dx = 0$$

gives rise to Hilbert matrix

$$H = \left[\frac{1}{i+j+1}\right]_{i,j=0,\dots,n}$$

n	$\|H\|\|H^{-1}\|$ (2 norm, matlab)
4	$5 \cdot 10^5$
9	$1.6 \cdot 10^{13}$
14	$2.25 \cdot 10^{17}$

- As we discussed, this is because the monomials $1, x, x^2, \dots$ all look the same.

- One can do a probabilistic analysis to show that $\frac{\|e\|}{\|x\|}$ is likely to be within a factor 10 of $\|A\|\|A^{-1}\|\frac{\|r\|}{\|b\|}$

- Query: We never compute an inverse. How do we compute the condition number.

- One of the greatest papers in Numerical Analysis

  Cline, Moler, Stewart, Wilkinson

  An Estimate for the Condition Number of a Matrix

  SIAM J Numerical Analysis

  V. 16 (1979) pp. 368 – 375

- online on our home page

  www.math.utah.edu/~pa/5600/CN.pdf

- The above analysis is an example of

  Backward Error Analysis

- you think of $\hat{x}$ as the exact solution of a perturbed problem $A\hat{x} = b + r$ rather than the approximate solution of an exact problem.

- The result is independent of the method by which we solve the linear system
- Consequence: you can't fight ill-conditioning, you have to avoid it.
- Example: use orthogonal polynomials instead of the power form.
- on the other hand, the trig fns
  $\sin kx$ $\cos kx$
  are already orthogonal.