Lecture 6

My office Hrs

The goal of this lecture is to understand simple method to solve Ax = b

We'll see how to turn the idea on the Neumann series from last class into an algorithm for Ax=b, and see some conditions for when it works.

Then we'll look at conditions for how to check a solution for a linear system

—> Aside on norms!
* Vector norms
—> all norms are equivalent
* Matrix norms & Submultiplicative prop

Next: other conditions for convergence and other ideas about this SAME alg.

To solve $A\underline{x} = \underline{b}$

Idea: use

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x} \text{ when } |x| < 1$$

and

$$\boxed{\sum_{k=0}^{\infty} A^k = (I-A)^{-1}}$$

To "multiply by A^{-1}"

Pf: Let $S_\ell = \sum_{k=0}^{\ell} A^k$

$(I-A) S_\ell =$

$(I-A)(I + A + A^2 + \dots) =$

$(I-A)I + (I-A)A + (I-A)A^2 + \dots$

$= (I - \cancel{A}) + (\cancel{A} - \cancel{A^2}) + (\cancel{A^2} - \cancel{A^3}) + \cancel{\dots} - A^{\ell+1}$

$= I - A^{\ell+1}$

If $A^{\ell+1} \to 0$, then $S_\ell \to (I-A)^{-1}$

When does $A^\ell \to 0$.

Consider Symmetric A for Convenience

Then $A = V \Lambda V^T$ is the spectral decomposition.

Spectral decomp.

V: Matrix of Eigenvectors ($n \times n$)

Λ: diag. matrix of Eigenvalues.

$A\underline{v} = \lambda \underline{v}$

$\downarrow$

$AV = V\Lambda$

Also $\underline{v}_i^T \underline{v}_j = \begin{cases} 1 & i = j \\ 0 & \text{else} \end{cases}$

$V^T V = I$

$A^\ell = A \cdot A \cdots A$

$= V\Lambda \underbrace{V^T V}_{I} \Lambda \underbrace{V^T V}_{I} \cdots V \Lambda V^T$

$= V \Lambda^\ell V^T$

$$\Lambda^\ell = \begin{bmatrix} \lambda_1^\ell & & 0 \\ & \ddots & \\ 0 & & \lambda_n^\ell \end{bmatrix}$$

So If $A^\ell \to 0$, then $\lambda_i^\ell \to 0$ so $|\lambda_i| < 1$.

Definition

$\rho(A)$ is the Spectral radius of A, defined as $\max_i |\lambda_i(A)|$

$\max \{ |\lambda| \text{ where } \lambda \text{ is an Eigenvalue of } A \}$.

If $\rho(A) < 1$, then $\sum_{k=0}^{\infty} A^k = (I-A)^{-1}$.

We want $Ax = \underline{b}$

We can compute $(I-H)\underline{x} = \underline{b}$ for H via the Neumann series.

What is H s.t. I get $A\underline{x} = \underline{b}$?

I want $(I-H) = A$, so

$H = I - A$.

Alg for $A\underline{x} = \underline{b}$

$\underline{x}_k = \sum_{\ell=0}^{k} (I-A)^\ell \underline{b}$

$= (I-A)\underline{x}_{k-1} + \underline{b}$

where $\underline{x}_0 = \underline{b}$.

In Julia:

x = b

for ...

y = (I-A)*x

x = y + b

y = x - A*x

Why?

If $\underline{x}_{k-1} = \sum_{\ell=0}^{k-1} (I-A)^\ell \underline{b} =$

$(I-A)^{k-1}\underline{b} + \dots$

$(I-A)\underline{b} +$

$\underline{b}$

So $(I-A)\underline{x}_{k-1}$

$(I-A)^k \underline{b} + \dots$

$(I-A)^2 \underline{b} +$

$(I-A)\underline{b}$. so

we are missing $\underline{b}$.

If $\rho(I-A) < 1$, this works!

Now, in theory $k \to \infty$.

We need to know when is k large enough?

Question: When is y a good approx. to $\underline{x}$ where $A\underline{x} = \underline{b}$?

Idea 1: If $y \approx \underline{x}$ then it's close!

Idea 2: If $Ay \approx \underline{b}$

Idea 2': If $Ay - \underline{b} \approx 0$.

Some terms to help!

Error $= y - \underline{x}$

approx. soln. $\uparrow$ $\uparrow$ true soln

$\|\text{error}\| = \|y - x\|_2$

Relative Error $\dfrac{\|y - x\|_2}{\|x\|_2}$

Residual ~~$\underline{r} = Ay - \underline{b}$~~

$\underline{r} = \underline{b} - Ay$

$\|\underline{r}\| = \|b - Ay\|$

rel. Resid

$\|b - Ay\|_2 / \|b\|_2$

We can compute residual.

We would like Error.

Fun fact:

$A \cdot (\text{Error}) = \text{residual}$.

$A(y - \underline{x}) = Ay - A\underline{x}$

$= Ay - \underline{b}$

$\|\text{Error}\| =$

$\|A^{-1} \text{residual}\|$

If $\|\cdot\|$ is sub-multiplicative

$\|y - x\|_2 \le \|A^{-1}\|_2 \cdot \|b - Ay\|_2$

$\|\text{error}\|_2 \le \text{Const.} \cdot \|\text{resid.}\|_2$

$\Rightarrow$ Make resid as small as possible.

Back to Algs

Idea: guess, check, correct

check, correct. ...

How to check?

$\to$ look at residual!

Silly Idea:

If $\underline{x}_k \approx \underline{x}$ then

$r_k = \underline{b} - A\underline{x}_k$

will be small.

$\|r_k\|$ is small.

Let's try

$\underline{x}_{k+1} = \underline{x}_k + \underline{r}_k$

$\underline{x}_{k+1} = \underline{x}_k + (\underline{b} - A\underline{x}_k)$

$= \underline{x}_k + \underline{b} - A\underline{x}_k$

$= \underline{x}_k - A\underline{x}_k + \underline{b}$

$= (I-A)\underline{x}_k + \underline{b}$

~~Quiz~~ (Not today for time) Let $\underline{x}_0 = \underline{b}$

Use $\underline{x}_{k+1} = \underline{x}_k + \underline{r}_k$

where $\underline{r}_k = \underline{b} - A\underline{x}_k$

Show $\underline{x}_k \to \underline{x}$

(and give conditions when.)

$\underline{x}_0 = \underline{b}$

$\underline{x}_1 = \underline{x}_0 + \underline{b} - A\underline{x}_0$

$= \underline{b} + \underline{b} - A\underline{b}$

$= (I-A)\underline{b} + \underline{b}$

$\underline{x}_2 = \underline{x}_1 + \underline{r}_1$

$= \dots$

$= (I-A)^2 \underline{b} +$

$(I-A)\underline{b} +$

$\underline{b}$,

Ind. Show

$\underline{x}_k = \sum_{\ell=0}^{k} (I-A)^\ell \underline{b}$

So $\underline{x}_k \to (I - (I-A))^{-1} \underline{b}$

$\to A^{-1} \underline{b}$

when $\rho(I-A) < 1$

Norms!

* Vector norms
* Matrix norms.

Norms

- vector norms
- matrix norms

Norms generalize the absolute value.

$|x|$

They tell us size or magnitude (no direction).

"Norms tell us if it's small or big"

We have

$$\|x\|_2 = \sqrt{\sum_{i=1}^{n} |x_i|^2}$$

(2-norm, Euclidean norm)

$$\|\underline{x}\|_2 = \sqrt{\underline{x}^T \underline{x}}$$

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

A generalization:

p-norms

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

Why have multiple norms?

- They make some proofs easier!
- They give different guarantees.

$p = 1$, we get

$$\|x\|_1 = \sum |x_i|$$

$p = \infty$ (via a limit)

$$\|x\|_\infty = \max |x_i|$$

An even bigger class.

a function $f : \mathbb{R}^n \to \mathbb{R}$ is a vector norm if

1. $f(\underline{x}) \geq 0$
2. $f(\underline{x}) = 0$ iff $\underline{x} = 0$
3. $f(\alpha \underline{x}) = |\alpha| f(\underline{x})$ (1-homogeneous)
4. $f(\underline{x} + \underline{y}) \leq f(\underline{x}) + f(\underline{y})$ (tri-ineq.)

Eg. $f(x)$ = sum of the largest 2 entries in magnitude

Why vector norms? We want.

$\underline{x}_k \to \underline{x}$

$\iff$

$\|\underline{x}_k - \underline{x}\| \to 0.$

So does the norm matter?

Theorem: All vector norms are equivalent.

Let f, g be 2 vector norms, then there exist constants $C_1 \leq C_2$

s.t.

$$C_1 f(\underline{x}) \leq g(\underline{x}) \leq C_2 f(\underline{x})$$

for all $\underline{x}$.

- C_1, C_2 can depend on n.

$f(\underline{x}) = \|x\|_\infty$

$g(\underline{x}) = \|\underline{x}\|_1$

$$f(\underline{x}) \leq g(\underline{x}) \leq n f(\underline{x})$$

Matrix Norms

- $A \in \mathbb{R}^{m \times n}$
- Element-wise / Frobenius norms.

$$\|A\|_F = \sqrt{\sum_{ij} |A_{ij}|^2}$$

$$= \|\text{vec}(A)\|_2$$

vec(A) = any consistent way of making $A \in \mathbb{R}^{m \times n}$ into $\mathbb{R}^{mn}$.

- A function $f : \mathbb{R}^{m \times n} \to \mathbb{R}$ is a matrix norm if

1. $f(X) \geq 0$
2. $f(X) = 0$ iff $X = 0$
3. $f(\alpha X) = |\alpha| f(X)$
4. $f(X + Y) \leq f(X) + f(Y)$.

Sub. - Mult.

$$f(XY) \leq f(X) f(Y).$$