Lecture 7

(3/4th homework! 2:30pm 11/14 11/12)

Finish up Matrix norms.
- Operator induced norms
- Sub-multiplicative norms

That same algorithm (Richardson) converges with norms based on the spectral radius.

A scaled iteration converges in a different regime, and how to pick the right scale.

Convexity and a quadratic function related to Ax=b.

This will let us pick the scale dynamically to get an iteration that works for any symmetric pos. def. matrix!

Theorem: All vector norms are Equivalent.

Let f, g be 2 vector norms, then there exist Constant $C_1 \le C_2$ s.t.

$C_1 f(x) \le g(x) \le C_2 f(x)$ for all x.

- C_1, C_2 Can depend on n.

$f(x) = \|x\|_2$

$g(x) = \|x\|_1$

$f(x) \le g(x) \le n f(x)$

Matrix Norms.

- $A \in R^{m \times n}$

Element-wise/ Frobenius norms.

$\|A\|_F = \sqrt{\sum |A_{ij}|^2} = \|\mathrm{vec}(A)\|_2$

$\mathrm{vec}(A)$ = any Consider way of flattening $A \in R^{m \times n}$ into R^{mn}

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

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

Sub. - Mult.

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

Vector-mult. Matrix mult. mult mult

An important class of norms

Called operator - induced norms

Let f be a vector norm then

$f(A)$ def $\max_{x \neq 0} \frac{f(Ax)}{f(x)}$

Theorem: f is a matrix norm.

Pf: 1. (easy by vector norm)
2. $f(A) = 0$ iff $A = 0$.

3. $f(\alpha A) = \max_{x \neq 0} \frac{f(\alpha A x)}{f(x)}$
$= |\alpha| f(A)$

4. $f(A+B) = \max_{x \neq 0} \frac{f((A+B)x)}{f(x)}$

$\le \max_{x \neq 0} \frac{f(Ax) + f(Bx)}{f(x)} \le \max_{x \neq 0} \frac{f(Ax)}{f(x)} + \frac{f(Bx)}{f(x)}$

$\le \max_{x \neq 0} \frac{f(Ax)}{f(x)} + \max_{x \neq 0} \frac{f(Bx)}{f(x)}$.

Thus is sub-mult. too!

$f(AX) \le f(A) \cdot f(x)$

$f(AB) = \frac{f(ABx)}{f(x)} \le \max_{y \neq 0} \frac{f(Ay)}{f(y)} \cdot \frac{f(Bx)}{f(x)}$

$= f(A) \cdot f(B)$.

If $f(A) = 0$ then $f(Ae_j) = 0$ for all $i = 1$ to m.

So $Ae_i = 0$ by vector norm $\Rightarrow$

$A = 0$. $\Rightarrow$

$A = 0$.

Pf $f(A) = 0$ iff $A = 0$.

Then $\|A\|_2 = \max_{x \neq 0} \sum |A_{ij}|$

$\|A\|_1 = \max_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2}$

$\|A\|_\infty = \max_{x \neq 0} \sum |A_{ij}| + \max_{x \neq 0} \frac{\|Ax\|_\infty}{\|x\|_\infty}$

How to Remember?

$\|A\|_2$ = largest sing. value of A.

Back to $Ax = b$

A: $n \times n$

b: length n

We had:

$x_{k+1} = x_k + r_k$

$r_k = b - Ax_k$.

Let $e_k = x_k - x$ be the error.

$e_{k+1} = x_{k+1} - x$

$= x_k + b - Ax_k - x$

$= e_k - x + b - Ax_k$

$= e_k + Ax - Ax_k$

$= e_k - (A)x_k$

$= (I - A) e_k$

If $\|\cdot\|$ is sub mult.

$\|e_{k+1}\| \le \|(I - A)\| \|e_k\|$

Note $\rho(I - A) \le \|I - A\|$ for any norm.

What if $\rho(I - A) > 1$

Change the problem!

$Ax = b \Leftrightarrow \alpha A x = \alpha b$

Now we care about $\rho(I - \alpha A) \ll \|I - \alpha A\|$

If $\rho(I - \alpha A) < 1$, then $x_k \to x$

Result (spd)

If A is symm. pos. def. (then all λ_i are pos.)

then any $\alpha \in (0, \frac{2}{\rho(A)})$ will work in the method.

Pf: A is spd $\Rightarrow$

$0 < \lambda_i \le \rho(A)$

$\rho(I - \alpha A) = \max_i |1 - \alpha \lambda_i|$

< 1

A is spd $\Rightarrow$

$\lambda > 0$

But if we want convergence we need $\lambda \in$

$= \text{width} - 2$

$\alpha < \frac{2}{\rho(A)}$ then $\rho(I - \alpha A) < 2$

$-\alpha A$

$I - \alpha A \Rightarrow$ shifts right

We now have a way to solve any Linear sys problem.

Why?

LS: $\min_x \|Ax - b\|$

Now Eq $A^T A x = A^T b$

$A^T A$ is sym. pos. def.

$\rho(A^T A) \le \|A^T A\|_F$

If $\alpha < \frac{2}{\|A^T A\|_F}$, then it will converge.

---

for $Ax = b$

if not use

$A^T A x = A^T b$

in

Why?

LS: $\min_x \|Ax - b\|$

Now Eq $A^T A x = A^T b$

$A^T A$ is sym. pos. def.

$\rho(A^T A) \le \|A^T A\|_F$

If $\alpha < \frac{2}{\|A^T A\|_F}$, then it will converge.

(3/4th This is not the best way! for Ever. Linear sys on Linear Sys.)

CS 519 - Scientific Computing - Computer Science - David F. Gleich - 2019 CS 519 - Purdue University

Let's see the same thing again!

But first.

$f(x) = ax^2 + bx + c : \mathbb{R} \to \mathbb{R}$

Quiz find x to solve

$\arg\min_x f(x)$ (scalar)

If $a \leq 0$ cheat sheet.

If $a > 0$, the min when $f'(x) = 0$

$f'(x) = 2ax + b = 0$

$x = \frac{-b}{2a}$.

Use $f(\underline{x}) = \underbrace{\frac{1}{2}\underline{x}^T A \underline{x}}_{= \sum A_{ij} x_i x_j} - \underline{x}^T \underline{b}$

Goal: understand

$f(\underline{x}) = \frac{1}{2}\underline{x}^T A \underline{x} - \underline{x}^T \underline{b}$

and $A\underline{x} = \underline{b}$ when

A is sym. pos. def.

If A is 1×1

$A = [a]$

then $f(x) = \frac{1}{2}\underline{x}^T A \underline{x} - \underline{x}^T \underline{b}$

$= \frac{1}{2}ax^2 - bx$

$f'(x) = ax - b = 0$

$\frac{\partial f}{\partial x} = 0$.

Generalize to thinking

Set $f'(\underline{x}) = 0$

$f'(\underline{x}) = \begin{bmatrix} \partial f / \partial x_1 \\ \vdots \\ \partial f / \partial x_n \end{bmatrix}$

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

If we set $f'(\underline{x}) = 0$

then $A\underline{x} - \underline{b} = 0$.

What to show?

- $f'(\underline{x}) = 0 \iff$ minimum!
- Minimum is unique!
- Gradient descent works

Idea: show $f(x)$ is convex.

Gradient Descent Alg

$f'(\underline{x}) = A\underline{x} - \underline{b}$

$-\text{grad} = \underline{b} - A\underline{x}$

Gradient Descent

$x_{k+1} = x_k + \alpha \cdot [-\text{grad}]$

$= x_k + \alpha \underbrace{[\underline{b} - A\underline{x}_k]}_{r_k}$

CS 515 - Purdue University - Computer Science - David F. Gleich - 2019