Lecture 7

⊛ HW2 posted soon!

- Is one norm better than the others for Optimization probs? ✓
- Is norm a distance? A metric? Why non-negative? ✓
- Is grad. descent always better than Richardson ✓
- Why is $f(x+y) \le f(x) + f(y)$ Needed? ✓
- for $A\underline{x} = \underline{b}$ how is $x_k = \sum_{l=0}^{k} (I-A)^l \underline{b}$?
- Relationship between
  Error $\underline{x}^* - \underline{x}_k$
  Residual $b - A\underline{x}_k$
- Is Frobenius norm a matrix norm? ✓
  Is it a p-norm? ✓
  Is it sub mult.? ✓

- Why is $\max_{\underline{x}\neq 0} \frac{\|Ax\|_1}{\|x\|_1} = \max_j \sum_i |A_{ij}|$ ✓

- Are op. induced norm

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

well defined ✓

Why is $\rho(A) \le \|A\|$ (for ~~sub-mult.~~ induced $\|A\|$) → ✓

Can you do gradient descent on concave functions.

Yes but can be very slow.

- Better norms for opt. problem

$\min \|A\underline{x} - \underline{b}\|_2$

vs

$\min \|A\underline{x} - \underline{b}\|_1$

+ Solvable w/ linear system (fast!)

− tends to be "smeary"

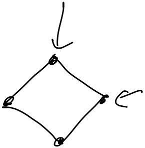
+ best unbiased linear Estimator (for Gaussian)

− very sensitive to outliers

+ Sparsity promotion for Residuals.

− slower to solve.

+ "more Robust" to Outliers.



Metric / Distance are functions that take in 2 points!

A norm measures magnitude

$\text{dist}(x,y) = \|x-y\|$. for any norm.

(normed metric spaces)

Any norm on $x \in \mathbb{R}$

has to be $f(x) = \alpha |x|$

$\alpha \geq 0$.

Frobenius 2-norm.

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

vec(A) as a vector space.

Is sub-mult.

Is not an induced p-norm. ✓

Is a matrix norm. ✓

Any vector norm is a

Matrix norm w/ vec(A).

Induced norm

$$\|A\| = \max_{x \neq 0}$$

IIRC, Ind

norms can onl

for p-norms whe

$p = 1, 2, \infty$,

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

$\|x\|_p = (\sum |x_i$

$$\frac{\|Ax\|}{\|x\|} \quad i$$

eel

be compt

en

/

$_j^p)^{1/p}$

$\|x\|_p \le \|x\|_q$

if $p \ge q$.

· Holder ineq.

CS ineq.

Why is induced 1-norm

$$\max_{\|x\|=1} \sum_i \left| \sum_j A_{ij} x_j \right|$$

(assume $A_{ij} \ge 0$, $x_j \ge 0$.)

$$\max_{\sum x_i = 1} \sum_{ij} A_{ij} x_j.$$

$$\begin{bmatrix} A_{11} x_1 & A_{12} x_2 & \sim \\ \vdots & & \\ A_{m1} x_1 & A_{m2} x_2 & \sim \end{bmatrix}$$

$$\sum_j \boxed{\left(\sum_i A_{ij}\right)} \cdot x_j.$$

$\uparrow$ colsum

$$\sum_j c_j \cdot x_j$$

$$\Downarrow$$

$$\max \sum c_j x_j$$
$$\text{s.t.} \sum x_i = 1$$
$$x_i \geq 0$$

$A_{1n} x_1$

$A_{mn} x_n$

$$\rho(A) = \max_i |\lambda_i|$$

(This is for Induc

$$\rho(A) = K \iff$$

$$Ay = K y \quad ($$

$$\hookrightarrow \frac{\|Ay\|}{\|y\|} = K$$

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

① and ② $\Rightarrow K$

< $\lambda_i \in$ Spectrum(A))

(~~And~~) As an Eigenvalue.

norm.)

(almost).

①

②

$\le \|A\|.$

$L\|A\|_F \le \|A\|_2 \le U\|A\|_F$

$\|A\|_2 \le \|A\|_F$

$\le \sigma_{max} \le \sqrt{\sum \sigma_i^2}$

$L = \frac{1}{n}$ or $\frac{1}{\sqrt{n}}$?

let $\underline{\sigma}$ be the vector of Singular values of A.

Then any vector norm of σ is a matrix norm.

$\|A\|_F = \|\sigma\|_2$

$\|A\|_2 = \|\sigma\|_\infty$

Richards worse than SD?

Richardson works on non-sym!

Can be faster for some structure (but not in general

If $\underline{x}_k \approx \underline{x}^*$ (true soln of $A\underline{x} = \underline{b}$)

or $\underline{x}_k \approx \underline{x}$

Then $\underline{x}_k - \underline{x}^*$ is the Error →

or $\underline{e}_k = x^* - x_k$ is the Error.

Also $\underline{r}_k = b - A\underline{x}_k$ is the Residual. Tha

Then $A\underline{e}_k = \underline{r}_k$ because

$$A(\underline{x}^* - x_k) = \underline{b} - A\underline{x}_k = \underline{r}_k$$

$\underline{f}_k = \underline{x}_k - x^*$

$s_k = \underline{A}\,\underline{x}_k - \underline{b}$

$A \underline{f}_k = \underline{s}_k$

~~Why is~~

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

how is

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

If $\rho(I-A) <$

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

This is the ALG implied by the Neumann series.

1, then

$^{k} = A^{-1}$.

$r_k = b - Ax_k$.

then

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

$g(\underline{x}) = A\underline{x} - \underline{b}$.

2024 CS 515 - Purdue University - Computer Science - David F. Gleich - 2024 CS 515 - Purdue University - Computer Science - David F. Gleich - 2024 CS 515 - Purdue University -