Lecture 8

Office Hours!

Final Exam Date/time.

12/17 10:30 - 12:30

(See Ed Stem)

Coord Descent = Gradient Descent? => No!

Steepest Descent vs. Gradient Descent.

Does Gradient Descent => following gradient? => Basically, yes, Technically, Can go "gradient Related"

Use 2nd order info to take a better step than gradient?

Richardson Relate to Poisson? Is Richardson used? vs Steepest descent?

Matrix Norms vs. Vector Norms ↳ Vector Norm of Each row/col?

When to pick Coord descent vs. gradient descent

How to compute the rank of a matrix/Sparse matrix efficiently?

$p = \infty$ in $\|x\|_\infty$ vs. $\max_{i=1} |x_i|$

· Properties of the Rayleigh quotient $\frac{x^T A x}{x^T x}$

· How to pick Coordinate j in Coord descent?

Gradient Descent vs. Steepest Descent.

Steepest Descent is an instance of gradient descent where

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

$g(x) = Ax - b$

AND

$x_{k+1} = x_k - \alpha_k g(x_k)$

w/ α_k chosen to minimize $f(x_{k+1})$

Gradient Descent

Any alg when

$x_{k+1} = x_k - \alpha g_k$

when $g_k \approx g(x_k)$.

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

$g(x) = \nabla$ 1st order approx => a vector

1st-ord approx of $g(x)$ $= g(x) = Ax - b$.

2nd order approx.

$f(x+p) = f(x) + p^T g(x) + \frac{1}{2} p^T H(x) p + O(\|p\|^3)$.

Using a 2nd-order idea(=)

Solve $Ax = b$ i.e. go w/ $x = A^{-1} b$.

Newton's method.

Given x_0 = arb.

find p s.t.

$x_0 + p = A^{-1} b$

if $x_k \neq x^*$

Then $x_k - x^* = A^{-1} \underbrace{(A x_k - b)}_{\text{Residual.}}$

Iterative Refinement

So if $B \approx A^{-1}$

Say low-precision approx. for GPU.

Then $x_{k+1} = x_k + p$ when

$p = B(A x_k - b)$

is a good approach

Let $\tilde{A} \approx A$

Then we are solving.

$\tilde{A} p = A x_k - b$.

Coord Descent

(!) aka Coordinate Relaxation!

$x_{k+1} = x_k + \alpha e_j$

Step

coordinate.

$\min_\alpha f(x_{k+1}) = \min_\alpha f(x_k + \alpha e_j) = \ldots$

Coord descent for all Coords 1 to n ≈ Work for 1 step of gradient descent.

What j?

- Cyclic / Sequential
- greedy
- random
- queue / optimized Cyclic

Richardson vs. Steepest Descent

Yes! Richardson doesn't need Symmetry!..

How to Compute rank Efficiently.

1. What is "rank" Computationally?

$$\begin{bmatrix} 1 & 2 \\ 1e\text{-}16 & -1e\text{-}16 \end{bmatrix} \ldots$$

2. In Matlab/Julia rank(A) = # of Sing. values $\geq \sigma_{max} \varepsilon$

when ε = floating point Round off.

→ pivoted QR can be slightly faster to get this than SVD.

$$\max_{x \neq 0} \frac{x^T A x}{x^T x} = \lambda_{max}.$$

if $A \succ 0$, $A = A^T$.

$$\min_{x \neq 0} \frac{x^T A x}{x^T x} = \lambda_{min}$$

if $A \succ 0$, $A = A^T$.

if you fix $\|x\|_2 = 1$, look at $x^T A x$