Announcements

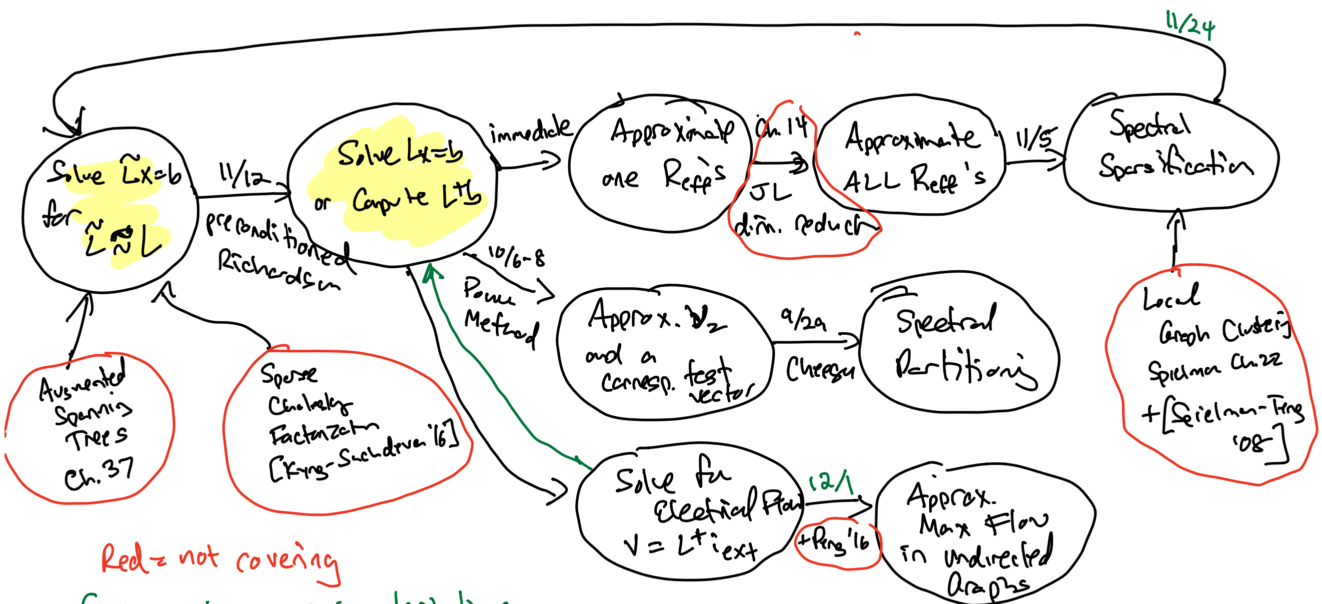
- start recording
- scribe: work w/other scribe to produce one set of notes
- My OH: Mon 12:30–1:30, Thu 9–10
- TF section/OH: Mon 2–3, Tue 8–9, Wed 5–6, Thu 5:30–6:30
- PS 4 posted. Project proposal due Sun 11/8, problems due TUE 11/17
- If Zoom goes down, check Piazza
- Sync whiteboard
- FOCS starting tomorrow! → no class next week
  stay tuned for watch+discussion parties
  we can reimburse the $20 reg fee
- jamboard link in chat

Agenda

- Recap: Richardson iterations
- The A-norm
- Preconditioning

Reductions between randomized nearly linear-time algorithms

time $\tilde{O}(m \cdot \text{polylog } n)$ $m = \#$ edges w/ nonzero weight



Red = not covering

Green = changes since last time

Richardson Iterations

Goal: Solve $Ax = b$ (e.g. A = a Laplacian or normalized Laplacian restricted to $v_1^\perp$)

without Gaussian elimination (only applying A to vectors)

Algorithm: $x^{(0)} = \vec{0}$

$$x^{(t+1)} = (I - \alpha A) x^{(t)} + \alpha b$$

$$= x^{(t)} - \alpha (A x^{(t)} - b)$$

when A is symmetric
gradient descent on
obj function
$f(x) = \frac{1}{2} x^T A x - b^T x$

Analysis: Suppose $Ax = b$

Then $x^{(t+1)} - x = (I - \alpha A)(x^{(t)} - x)$

$\|M\| = \max_{y \neq 0} \frac{\|My\|}{\|y\|}$

$\Rightarrow x^{(t)} \to x$ if $\|I - \alpha A\| < 1$

$$\|x^{(t+1)} - x\| \leq \|I - \alpha A\|^{t+1} \cdot \|x\|$$

$t = O\left(\frac{\log(\frac{1}{\varepsilon})}{1 - \|I - \alpha A\|} \right)$ iterations suffices to get

$\|x^t - x\| \leq \varepsilon \cdot \|x\|$

When A symmetric, $\|I - \alpha A\| = \max\{ |1 - \alpha \lambda_1|, |1 - \alpha \lambda_n| \}$

$\lambda_1 \leq \cdots \leq \lambda_n$
e-values of A

Optimizing α: $(1 - \alpha\lambda_1) = -(1 - \alpha\lambda_n)$

$\alpha = \frac{2}{\lambda_1 + \lambda_n}$ $\quad \|I - \alpha A\| = \frac{\lambda_n - \lambda_1}{\lambda_1 + \lambda_n}$

< 1 if $\lambda_1 > 0$

$\#$ iterations $= O\left(\left(1 + \frac{\lambda_n}{\lambda_1}\right) \cdot \log\left(\frac{1}{\varepsilon}\right) \right)$ also measures numerical stability

$\kappa(A)$ "condition number"

Similarly $\underbrace{Ax^{(t+1)} - b}_{\text{"residual"}} = (I - \alpha A)(Ax^{(t)} - b)$

$\Rightarrow$ Same # iterations gives $\|Ax^{(t)} - b\| \leq \varepsilon \cdot \|b\|$.

Another view: $x^{(t)} = p_t(A)b$ for a polynomial p_t

Richardson: $p_t(A) = \alpha \cdot \sum_{j=0}^{t-1} (I - \alpha A)^j \longrightarrow A^{-1}$ as $t \to \infty$

$(\alpha A)^{-1} = (I - (I - \alpha A))^{-1}$

$(1-x)^{-1} \ldots$

$\forall x \quad \|p_t(A)(Ax) - x\| \leq \varepsilon \cdot \|x\|$ where $x^{(t)} = p_t(A)(Ax)$

Richardson: suffices to take degree $t = O\left(\frac{\lambda_n}{\lambda_1} \cdot \log\left(\frac{1}{\varepsilon}\right)\right)$

Chebychev polys: degree $O\left(\sqrt{\frac{\lambda_n}{\lambda_1}} \cdot \log\left(\frac{1}{\varepsilon}\right)\right)$

Analysis of Richardson entirely in spectral norm:

$$I - p_t(A)A = (I - \alpha A)(I - p_{t-1}(A)A)$$

$\Rightarrow \|I - p_t(A)A\| \leq \|I - \alpha A\|^t$

$\forall x \quad \|x - p_t(A)b\| \leq \|I - \alpha A\|^t \cdot \|x\|$

NB: Holds for any submultiplicative matrix norm $(\|MN\| \leq \|M\| \cdot \|N\|)$

A-norm : $\|x\|_A = \sqrt{x^T A x}$ for $n \times n$ psd A

for psd A

$\|M\|_A = \max_{x \neq 0} \frac{\|Mx\|_A}{\|x\|_A}$ for $n \times n$ M

$\|MN\|_A \leq \|M\|_A \cdot \|N\|_A$.

Exercise for Breakouts : Fill in the missing matrices

1) $\|x\|_A = \|A^{1/2} x\|$

$\frac{\|Mx\|_A}{\|x\|_A} = \frac{\|A^{1/2} M A^{-1/2} (A^{1/2} x)\|}{\|A^{1/2} x\|}$

2) $\|x\|_A = \|Ax\|_{A^{-1}}$

3) $\|M\|_A = \|A^{1/2} M A^{-1/2}\|$

4) $\|I - ZA\|_A \leq \overset{1-\delta}{\varepsilon} \iff \underset{\delta A^{-1}}{(1-\varepsilon)A^{-1}} \preceq Z \preceq \underset{(2-\delta) A^{-1}}{(1+\varepsilon)A^{-1}}$

for a symmetric Z

$\|A^{1/2}(I - \overbrace{ZA) A^{-1/2}}^{A^{1/2} Z A^{1/2}}\| \leq \varepsilon$

$-\varepsilon I \preceq \underbrace{A^{1/2}(I - \cancel{Z}A) A^{-1/2}}_{= I} \preceq \varepsilon I$

Preconditioning

Given $A (= L_G)$ ($B = Z^{-1}$ is the "preconditioner")

Find a matrix Z such that

(1) easy to apply Z

(2) condition number of $\underline{ZA}$ is small.

$\Rightarrow$ solve $Ax = b$ by applying an iterative method

to solve $\underline{(ZA)x = (Zb)}$

Example: preconditioned Richardson with $\underline{\alpha = 1}$

$$x^{(t)} = (I - ZA)x^{(t-1)} + Zb$$
$$= p_t(ZA)\, Zb$$

$\cancel{\varepsilon = 1 - \delta\, \lambda_A(B)}$

$$\| I - p_t(ZA) \cdot ZA \|_A \le \| I - ZA \|_A^t \le (1-\delta)^t \le \varepsilon$$

$\varepsilon = 1 - \delta$ in (3) above

for the largest value δ s.t. $\delta \cdot A^{-1} \preccurlyeq Z \preccurlyeq (2-\delta) \cdot A^{-1}$ $\quad \Theta(\frac{1}{\delta})$

iterations to get w/in ε

$$= O\left(\frac{1}{\delta} \cdot \log\left(\frac{1}{\varepsilon}\right)\right)$$

Up to scaling $B = Z^{-1}$, $\frac{1}{\delta} = \Theta(\kappa(A, B))$,

where $\kappa(A,B)$ = smallest ratio $\frac{\beta}{\alpha}$ s.t. $\alpha A \preccurlyeq B \preccurlyeq \beta A$

$$= \frac{\lambda_{max}(B^{-1}A)}{\lambda_{min}(B^{-1}A)} = \frac{\lambda_{max}(B^{-1/2} A B^{-1/2})}{\lambda_{min}(B^{-1/2} A B^{-1/2})}$$

$A = L_G$

$Z L_G$

Goal: given G, construct Z s.t.

(1) $L_G^+ \preceq Z \preceq c \cdot L_G^+$ for "small" c

L_H^+

(2) Can apply Z in nearly linear time

$\Rightarrow$ Can solve linear systems in L_G in time

$$O(m \cdot \text{polylog}\, n \cdot c \cdot \log 1/\varepsilon))$$

can be $\sqrt{c}$ using Chebyshev/CG

Subgraph preconditioners : $Z = L_H^+$ for connected subgraph H

Then (1) becomes $L_H \preceq L_G \preceq c \cdot L_H$ free

(2) true for trees

Prop: if H is a spanning tree of $G = (V, E)$, then $L_G \preceq c \cdot L_H$

for $c = \underbrace{\sum_{e=(a,b) \in E} \frac{1}{r_e} \cdot \sum_{f \in \text{path}_H(a,b)} r_f}_{\text{stretch of edge } e \text{ wrt } H}$ $\qquad r_e = \frac{1}{w_e}$

Proof sketch: $c = \lambda_{max}(L_H^+ L_G) \leq \text{Tr}(L_H^+ L_G)$

$$= \sum_{e=(a,b) \in E} w_e \cdot \text{Tr}(L_H^+ (\delta_a - \delta_b)(\delta_a - \delta_b)^T)$$

$$= \sum_{e=(a,b) \in E} \frac{1}{r_e} \cdot R_{eff}^H(a,b)$$

Low-stretch Spanning trees: $c = O(m \cdot \log n \cdot \log\log n)$

$\Rightarrow$ Solve systems in L_G in time $\tilde{O}(m^{3/2})$

Next time (11/24): construct Z with $c = \text{polylog}(n)$

assuming spectral sparsification in near linear time.