Preconditioned Iterative Methods

Basic Idea

Goal: Solve $Ax=b$ A spd or spsd

Pick a method, say Richardson

Idea find a matrix B and solve

$$B^{-1}Ax = B^{-1}b \qquad (B \text{ is called a preconditioner})$$

note $Ax=b$ iff $B^{-1}Ax = B^{-1}b$

This gives
$$\begin{aligned} x^{(n+1)} &= x^{n} + (B^{-1}b - B^{-1}Ax^{(n)}) \\ &= x^{n} + B^{-1}(b - Ax^{(n)}) \end{aligned}$$

Alg

1) $r = b - Ax^{(n)}$
2) solve $By = r$
   (preprocess $B = LU$ or solve recursively)
3) return $x^{(n+1)} = x^{(n)} + y$

Examples of preconditions.

1) Extrapolated method

$$B = \frac{2}{\lambda_{max}(A) + \lambda_{min}(A)} I$$

2) Jacobi $\quad B = \text{Diagonal}(A)$

3) Gauss-Seidel $\quad B = \text{LowerTri}(A)$

4) SSOR

5) Graph theoretic

Convergence Rates

P-Richardson $\quad \text{Rate} \approx \frac{1}{\kappa(B^{-1}A)}$

PCG $\quad \text{Rate} \approx \frac{1}{\sqrt{\kappa(B^{-1}A)}}$

Goal: Find B s.t. we get better convergence rate but solves in step 2 are not too expensive.

Estimating $\kappa(B^{-1}A)$

Note $B^{-1}Ax = \lambda x$ iff $Ax = \lambda Bx$

Def λ & x are called generalized eigenvalues & vectors

Let $\lambda_f(A,B) = \{\lambda : \exists x \; Ax = \lambda Bx \;\&\; x \notin Null(A) = Null(B)\}$

Note $Ax = \lambda Bx$ iff $Bx = (1/\lambda) Ax$

If B spd then iff $B^{-1/2} A B^{-1/2} x = \lambda x$

$$\kappa(B^{-1}A) = \frac{\max \lambda_f(A,B)}{\min \lambda_f(A,B)} = \max \lambda_f(A,B) \max_f(B,A)$$

$$\sigma(A/B) = \operatorname*{argmin}_{\tau} \; \tau B \succeq A$$

Claim $\kappa(B^{-1}A) = \sigma(A/B)\,\sigma(B/A)$

Subgraph Preconditioners

Let $H \subseteq G$ be a subgraph of G.

$B \equiv L_H$ & $A \equiv L_G$

$B \preceq A$ thus eigens of $B^{-1}A \geq 1$

Thus $\kappa(B^{-1}A) \leq \sigma(A/B) = \underset{\tau}{\arg\min}\ \tau B \succeq A$

for any subgraph Precon!

Spanning Tree Preconditioners

Vaidya: H to be max weight spanning tree of G
say T

$\varphi : G \xrightarrow{\text{path}} T$

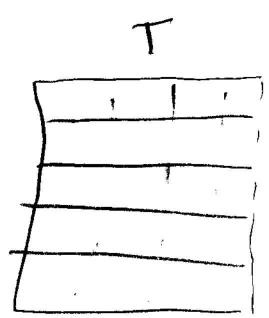
Congestion $\leq m$ # edge in G
Dialation $\leq n$

$\sigma(T/G) \leq 1 \quad (T \subseteq G)$

$\sigma(G/T) \leq m \cdot n$

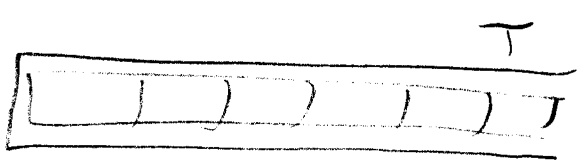
$\kappa(G,T) \leq m \cdot n$

EG



The E tree for square mesh.

$C = \sqrt{n}$
$D = \sqrt{n}$
$C \cdot D = n$



$C = n$
$D = n$
$C \cdot D = n^2$

Thm Max Spanning Tree CG $\rightsquigarrow$
$O(m^{3/2} \cdot n^{1/2})$ time per bit.

1) $O(m)$ time to find T
2) $O(m)$ time per iteration
3) $O((m \cdot n)^{1/2})$ iteration per bit using PCG.

Finding Better Trees

EG C_n be the cycle & P_n a spanning tree.

Claim $\kappa(C_n, P_n) = \Theta(n)$

1) $\sigma(P_n / C_n) = 1$ & $\sigma(C_n / P_n) \leq C \cdot D = O(n)$

$X = (x_1 \cdots x_n) \quad x_i = i$

$$\frac{x^T C_n x}{x^T P_n x} = \frac{(n-1)^2 + (n-1)}{(n-1)} \approx n$$

$$\lambda_{max}(P_n^{-1}C_n) \approx n \quad \lambda_{min}(P_n^{-1}C_n) \approx 1$$

What is the average? $\sum \lambda_i$?

Fact: $Tr(A) = \sum \lambda_i$

$Tr(P_n^{-1}C_n)$? Note $Tr(AB) = Tr(BA)$

$A = L_G$ & $B = L_H$

Claim $$Tr(B^{-1}A) = \sum_{e \in A} w_A(e)\, ER_B(e)$$

pf Let $L_{u,v} = (e_u - e_v)(e_u - e_v)^T$ $\quad e_u = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \Big\} u$

$$Tr(B^{-1}A) = Tr\Big(\sum_{e=(u,v) \in A} B^{-1} w_A(u,v) L_{u,v}\Big)$$

$$= \sum w_A(u,v) \sum Tr(B^{-1} L_{uv})$$

$$= \sum w_A(u,v) \underbrace{(e_u - e_v)^T B^{-1} (e_u - e_v)}_{(*)}$$

$$\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = B^{-1}\begin{pmatrix} 0 \\ \vdots \\ +1 \\ 0 \\ \vdots \\ -1 \\ 0 \\ \vdots \end{pmatrix} \text{ ie } B\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ +1 \\ 0 \\ \vdots \\ -1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$$\therefore (*) = ER_B(u,v)$$

<u>Def</u>. Given G, spanning tree T, and $e \in G$

Let $w_1, \dots, w_k$ edge weights on path from u to v. $e = (u,v)$

$$\text{stretch}(e) = w_A(e) \sum_{i=1}^{k} \frac{1}{w_i}$$

<u>Thm</u> $\forall G \; \exists$ spanning tree T st.

$$\sum_{e \in G} \text{stretch}(e) = O(m \log n (\ell\ell n)^2)$$

m = # edges in G.

Back to C_n & P_n

$$\sum_{\lambda_i \in \lambda(C_n, P_n)} \lambda_i = Tr(P_n^{-1} C_n) = \sum_{e \in C_n} ER_T(e)$$

note: $\forall e \in P_n \quad ER_T(e) = 1$

For $e \notin P_n \quad ER_T(e) = n-1$

$$\sum \lambda_i = 2(n-1)$$

Homework: 1 of size n 1 of size 0.

$n-2$ of size 1

Consider $P(X) = (1-X)(1-X/n)$

After 3 iterations PCG on (C_n, P_n) will converge.

<u>Thm</u> Suppose $\lambda_1, \dots, \lambda_n \in \mathbb{R}^+$, $\alpha \le \lambda_i$, at most k st $\lambda_i > \beta$.

$\forall t \geq k \; \exists P(x)$ of degree t s.t.

1) $P(0) = 1$

2) $\forall i \quad |P(\lambda_i)| \le 2 \left(\frac{\sqrt{\beta/\alpha} - 1}{\sqrt{\beta/\alpha} + 1} \right)^{t-k}$

<u>pf</u> Let $r(x)$ be the Chebychev poly of deg $t-k$ st

$$|r(\lambda)| \le 2 \left(\frac{\sqrt{\beta/\alpha} - 1}{\sqrt{\beta/\alpha} + 1} \right)^{t-k} \quad \forall \alpha \le \lambda \le \beta$$

Set $P(x) = r(x) \prod_{\lambda_i > \beta} (1 - x/\lambda_i)$

$$|P(\lambda_i)| = \begin{cases} 0 & \text{if } \lambda_i > \beta \\ \le |r(\lambda_i)| & \text{if } \lambda_i \le \beta \end{cases}$$

Thm Using Low stretch trees PCG

$O(m^{1/3} \log n)$ iterations per bit.

pf Set $\beta = Tr(L_T^{-1} L_G)^{2/3}$ then $K = Tr(L_T^{-1} L_G)^{1/3}$

note $\alpha \geq 1$

$2K$ iterations is $2\sqrt{\beta}$

$$|p(\lambda_i)| \leq 2\left(1 - \frac{1}{\sqrt{\beta}}\right)^{2\sqrt{\beta} - \sqrt{\beta}} = 2\left(1 - \frac{1}{\sqrt{\beta}}\right)^{\sqrt{\beta}} \approx \frac{1}{e}$$

$O(Tr(L_T^{-1} L_G)^{1/3})$ iter per bit

$O(m^{1/3} \log n)$ iter per bit