Conjugate Gradient Method & Steepest Descent

15-859N
10/3/16

Goal: Solve $Ax=b$ where A spd & $A\bar{u}=b$

We start with Steepest Descent.

Consider "Eliptical Bowl"

$$M(u) = (\bar{u}-u)^T A(\bar{u}-u)$$

note $\min_u M(u) = \bar{u}$

The major axises are the eigen vectors of A.

Expand

$$M(u) = \bar{u}^T A\bar{u} - 2u^T A\bar{u} + u^T Au$$

$$\underbrace{\bar{u}^T b}_{\text{constant}} - 2u^T b + u^T Au$$

Suffice to minimize $F(u) = \frac{1}{2}u^T Au - u^T b$

Crash course on matrix calculas

Let $F : \mathbb{R}^n \to \mathbb{R}$

Def $\nabla F = \begin{pmatrix} \frac{dF}{du_1} \\ \vdots \\ \frac{dF}{du_n} \end{pmatrix}$

$$\nabla = \frac{1}{2} \nabla u^T A u - \nabla u^T b$$

Claim $\frac{d\, u^T b}{du_i} = b_i$

pf $$\frac{d\, u^T b}{du_i} = \lim_{h \to 0} \frac{(u + he_i)^T b - u^T b}{h}$$

$$= \lim_{h \to 0} \frac{h e_i^T b}{h} = b_i$$

Check that

$$\frac{d\, u^T A u}{du_i} = 2(Au)_i \quad A \text{ sym}$$

Thus $\nabla F(u) = Au - b$

<u>Note</u> Richardson's Alg

$$u^{(n+1)} = u^{(n)} + \lambda(b - Au^{(n)}) \quad \lambda = 1$$

thus RA is steepest descent with step size $\lambda = 1$.

The Exstropolated Method is optimal fixed size method.

<u>Goal!</u> Find a variable size method.

<u>Idea</u> Pick λ to minimize $F(u^{(n)} - \lambda r)$

Set $u = u^{(n)}$

Now

$$F(u-\lambda r) = \frac{1}{2}(u-\lambda r)^T A(u-\lambda r) - b^T(u-\lambda r)$$

$$= \frac{1}{2}u^T A u - \lambda u^T A r + \frac{1}{2}\lambda^2 r^T r - b^T u + \lambda b^T r$$

$$\frac{dF(u-\lambda r)}{d\lambda} = -u^T A r + \lambda r^T r + b^T r$$

$$r^T(b-Au) + \lambda r^T r$$

$$= -r^T r + \lambda r^T r$$

---

set $\frac{\partial F}{\partial \lambda} = 0 = -r^T r + \lambda r^T r$

$$\lambda = \frac{r^T r}{r^T A r} \leftarrow \text{step size!}$$

Steepest Descent

Input: A, b $\quad A$ spd

Init: $u^{(0)} = 0$

$$u^{(n+1)} = u^{(n)} + \lambda r \text{ where}$$

$$r = b - Au$$

$$\lambda = r^T r / r^T A r$$

Convergence Rate:

(Kantorvich Lemma) B spd

λ_M & λ_m for B. then

$$\frac{(x^T B x)(x^T B^{-1} x)}{(x^T x)^2} \leq \frac{(\lambda_M + \lambda_m)^2}{4 \lambda_M \lambda_m}$$

pf See SAAD Chap 5 page 132.

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

$x \perp_A y$ if $x^T A y = 0$

note $x A y = (A^{1/2} x)^T (A^{1/2} y)$

$x \perp_A y$ iff $A^{1/2} x \perp A^{1/2} y$

Thm A spd & $\varepsilon^{(k)} = \bar{u} - u^{(k)}$ for steepest descent then

$$\|\varepsilon^{(k+1)}\|_A \le \left(\frac{\lambda_M - \lambda_m}{\lambda_M + \lambda_m}\right) \|\varepsilon^{(k)}\|_A$$

$$= \left(\frac{\kappa - 1}{\kappa + 1}\right) \|\varepsilon^{(k)}\|_A \quad \kappa = \frac{\lambda_M}{\lambda_m}$$

$$\approx \left(1 - \frac{2}{\kappa}\right) \|\varepsilon^{(k)}\|_A$$

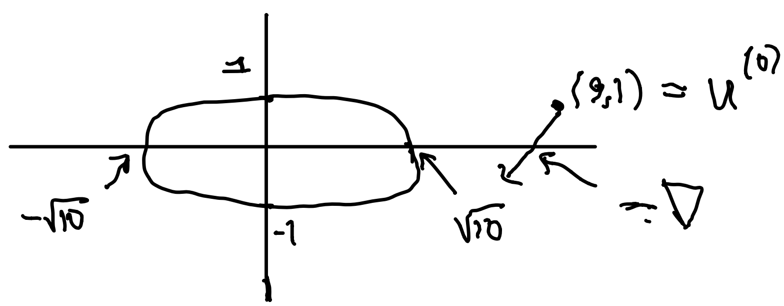
pf SAAD Chap 5 133p

Steepest Descent an Example

$$A\begin{pmatrix} 1 & 0 \\ 0 & 10 \end{pmatrix} \quad b = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \bar{u} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad u^{(0)} = \begin{pmatrix} 9 \\ 1 \end{pmatrix}$$

$$r = b - Au^{(0)} = -\begin{pmatrix} 1 & 0 \\ 0 & 10 \end{pmatrix}\begin{pmatrix} 9 \\ 1 \end{pmatrix} = -\begin{pmatrix} 9 \\ 10 \end{pmatrix}$$

"the Eliptical bowl"



$$\lambda = \frac{r^T r}{r^T A r} = \frac{181}{81 + 1000} = \frac{181}{1081}$$

$$u^{(1)} = \begin{pmatrix} 9 \\ 1 \end{pmatrix} - \left(\frac{181}{1081}\right)\begin{pmatrix} 9 \\ 10 \end{pmatrix}$$

$$u_y^{(1)} = 1 - \frac{1810}{1081} = \frac{-729}{1081} \approx -3/4$$

we over shot $y = 0$!

Krylor Subspaces

$Ax=b$ & $A\bar{x}=b$ initial guess $x^{(0)}$

Def nth Krylov subspace

$$\mathcal{K}_n = \langle b, Ab, A^2b, \dots, A^{n-1}b \rangle$$

If $x^{(0)} \neq 0$ affine space $x^{(0)} + \mathcal{K}_n$ Krylov space

Why is $\mathcal{K}_n$ of interest?

assume $x^{(0)} = 0$

Claim Given an iterative method

$$x^{(n+1)} = \alpha_n x^{(n)} + \beta_n (b - Ax^{(n)}) \quad \text{then}$$

$$x^{(n)} \in \mathcal{K}_n$$

pf Induction on n

$n=1$ then $x^{(1)} = \beta b \in \mathcal{K}_1$

Claim If $x^{(n)} \in \mathcal{K}_n$ then $Ax^{(n)} \in \mathcal{K}_{n+1}$

assume $x^{(n)} = a_1 b + \cdots + a_{n-1} A^{n-1} b$

$$Ax^{(n)} = a_1 Ab + \cdots + a_{n-1} A^n b \in \mathcal{K}_{n+1}$$

$\therefore \; x^{(n+1)} \in \mathcal{K}_{n+1}$

All our methods do is find $x^{(n)} \in \mathcal{K}_n$!

Best would be to find $\underset{x \in \mathcal{K}_n}{\text{Argmin}} \|\bar{x} - x\|_2$

Not clear how to do this!

Next best thing

$$\underset{x^{(n)} \in \mathcal{K}_n}{\text{Argmin}} \|\bar{x} - x^{(n)}\|_A$$

This we can do! It is called Conjugate Gradient.

Conjugate Gradient Iteration

Init: $x_0 = 0$ $r_0 = b$ $p_0 = r_0$

Iteration:

$\alpha_n = r_{n-1}^T r_{n-1} / p_{n-1}^T A p_{n-1}$ step length

$x_n = x_{n-1} + \alpha_n p_{n-1}$ approx solution

$r_n = r_{n-1} - \alpha_n A p_{n-1}$ residual

$\beta_n = r_n^T r_n / r_{n-1}^T r_{n-1}$ improvement

$p_n = r_n + \beta_n p_{n-1}$ search direction

Claim $r_n = b - Ax_n$

pf induct on n $n=0$

assume true for $n-1$

$$b - Ax_n = b - A(x_{n-1} + \alpha_n p_{n-1})$$

$$= (b - Ax_{n-1}) - \alpha_n A p_{n-1} \quad \text{induct}$$

$$= r_{n-1} - \alpha_n A p_{n-1} = r_n$$

Thm A CG for spd A for prob $Ax=b$ then while $r_{n-1} \neq 0$

$$\mathcal{K}_n = \langle x_1, \cdots, x_n \rangle = \langle p_0, \cdots, p_{n-1} \rangle$$

$$\langle r_0, \cdots, r_{n-1} \rangle = \langle b, Ab, \cdots, A^{n-1}b \rangle$$

& $r_n^T r_j = 0 \quad (j<n)$

& $p_n^T A p_j = 0 \quad (j<n)$

---

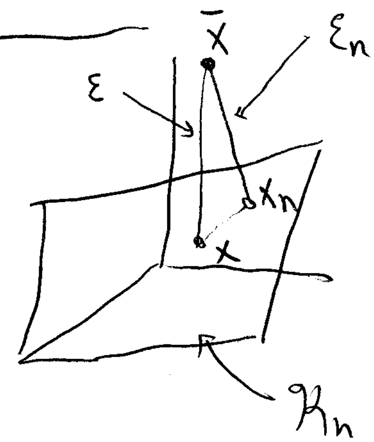
pf See Trefethen & Bau Chap 38

<u>Thm B</u> $A^{m\times m}x=b$, A spd, CG

1) while $r_{n-1}\neq 0$ then x_n unique point in $\mathcal{K}_n$ minimizing $\|\varepsilon_n\|_A$ $\quad(\varepsilon_n=\bar{x}-x_n)$

2) $\|\varepsilon_n\|_A \le \|\varepsilon_{n-1}\|_A$

3) $\exists\, n\le m$ st $\varepsilon_n=0$



<u>pf</u> $x_n\in\mathcal{K}_n$ by Thm A.

Let $x\in\mathcal{K}_n$ be arbitrary.

Set $\Delta x = x_n - x$

$\varepsilon = \bar{x}-x = \varepsilon_n+\Delta x$

$$\|\varepsilon\|_A^2 = (\varepsilon_n+\Delta x)^T A(\varepsilon_n+\Delta x)$$

$$\varepsilon_n^T A\varepsilon_n + (\Delta x)^T A(\Delta x) + 2\varepsilon_n^T A(\Delta x)$$

<u>note</u> $A\varepsilon_n = A(\bar{x}-x_n) = b - Ax_n = r_n$

$2\varepsilon_n^T A\Delta x = 2r_n^T\Delta x = 0$ since $r_n\perp\mathcal{K}_n$

(thus) $\|\varepsilon\|_A^2 = \varepsilon_n^T A \varepsilon_n + (\Delta x)^T A (\Delta x)$

A spd $\Rightarrow$ $x \neq x_n$ then $\|\varepsilon\|_A^2 > \|\varepsilon_n\|_A^2$

---

2) $\mathcal{K}_n \subseteq \mathcal{K}_{n+1} \Rightarrow \|\varepsilon_n\|_A \leq \|\varepsilon_{n-1}\|_A$

---

3) For some $n \leq m$ $\tilde{x} \in \mathcal{K}_n$ and done.

(Cayley-Hamilton)

CG & Polynomial Approx

<u>Prob:</u> Find poly P_n st $P_n(0)=1$ & $\deg(P_n) < n$ to

minimize $\|P_n(A)\|_A$

<u>Thm C:</u> While $r_{n-1} \neq 0$ for CG

$\exists! \, P_n$, $P_n(0)=1$, $\deg(P_n) < n$

1) $\varepsilon_n = \tilde{x} - x_n = P_n(A)\,\varepsilon_0$

2) $$\frac{\|\varepsilon_n\|_A}{\|\varepsilon_0\|_A} = \inf_{P \in P_n} \frac{\|P(A)\,\varepsilon_0\|_A}{\|\varepsilon_0\|_A}$$

$$\leq \inf_{P \in P_n} \max_{\lambda \in \lambda(A)} |P(\lambda)|$$

<u>pf of 1) thm C</u>

By Thm A $X_n \in \mathcal{K}_n$ where $X_0 = 0$

thus $X_n = Q_{n-1}(A)b$ some polynomial $Q_{n-1}(z)$

$\deg(Q_{n-1}(z)) \leq n-1$

Consider $P_n(z) = 1 - z \cdot Q(z)$ $\quad \deg(P_n) \leq n$

<u>Claim</u> $\mathcal{E}_n = P_n(A)\mathcal{E}_0$ where $\mathcal{E}_0 = \bar{X}$

$$
\begin{aligned}
P_n(A)\mathcal{E}_0 &= (1 - A\,Q_{n-1}(A))\bar{X} \\
&= \bar{X} - Q_{n-1}(A)(A\bar{X}) \\
&= \bar{X} - Q_{n-1}(A)b \\
&= \bar{X} - X_n = \mathcal{E}_n
\end{aligned}
$$

<u>Pf</u> Thm C 2)

($\Rightarrow$) Follows from Thm B

($\Leftarrow$) $\varepsilon_0 = \sum a_j V_j$ (eigen expansion of A)

$V_i^T V_i = 1$

$\forall$ poly P $\quad P(A)\,\varepsilon_0 = \sum a_j P(\lambda_j) V_j$

thus

$$\|\varepsilon_0\|_A^2 = \sum a_j^2 \lambda_j$$

$$\|P(A)\,\varepsilon_0\|_A^2 = \sum a_j^2 \lambda_j P(\lambda_j)^2$$

thus $$\frac{\|P(A)\,\varepsilon_0\|_A^2}{\|e_0\|_A^2} \leq \max_{\lambda \in \lambda(A)} |P(\lambda)|^2$$

Thm CG to solve $Ax=b$, A spd

$$\frac{\|\varepsilon_n\|_A}{\|\varepsilon_0\|_A} \le 2\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^n$$

pf

$$\frac{\|\varepsilon_n\|_A}{\|\varepsilon_0\|_A} \le \max_{\lambda\in\lambda(A)} |P_n(A)| \le \max_{\lambda\in\lambda(A)} |P(x)|$$

any $P(x)$ st $\deg(P) \le n-1$

$P(0)=1$

Call $Ax=b$ $\quad m=\lambda_m \le \lambda(A) \le \lambda_M = M$

$G_\alpha = (I - \alpha A) \qquad -\gamma \le \lambda(G_\alpha) \le \gamma$

for $\gamma = \dfrac{M-m}{M+m}$

$$P(x) = T_n\left(\gamma^{-1} - \frac{2x}{M-m}\right) \Big/ T_n(\gamma^{-1})$$ $T_n \equiv$ Chebyshev Poly

$$= T_n\left(\frac{M+m-2x}{M-m}\right) \Big/ T_n(\gamma^{-1})$$

Note $P(0)=1 \quad \deg(P) \le n-1$

$P(x) \le 1/T_n(\gamma^{-1})$ for $m \le x \le M$

Thus $P(\lambda) \le 2\left(\dfrac{\sqrt{k}-1}{\sqrt{k}+1}\right)^n \quad \forall \lambda \in \lambda(A)$

$k = M/m$