

15-859N
10/12/11

Conjugate Gradient Method

Steepest Descent

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

Start with steepest descent

Consider "Elliptical Bowl"

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

$$\min_u G(u) = \bar{u}$$

Major axes are eigenvector of A .

note $G(u) = \bar{u}^T A \bar{u} - 2\bar{u}^T A u + u^T A u$

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

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

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

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

Claim $\frac{d u^T b}{d u_i} = b_i$ pf $\lim_{h \rightarrow 0} \frac{(u + h e_i)^T b - u^T b}{h}$

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

Check $\frac{d u^T A u}{d u_i} = 2(Au)_i$

$\nabla F(u) = Au - b$ residual $r = b - Au$

Note Richardson Alg

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

RA is steepest descent, step size $\lambda = 1$

Goal: Pick λ minimize $F(u^{(n)} + \lambda r)$

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

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

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

$$r^T (A u - b) + \lambda r^T A r$$

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

$$\frac{\partial F}{\partial \lambda} = 0$$

$$-r^T r + \lambda r^T A r = 0$$

$$\lambda = \frac{r^T r}{r^T A r}$$

Steepest Descent

Initial guess $u^{(0)}$

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

$$r = b - Au^{(n)}$$

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

Convergence Rate

(Kantorovich Lemma) B spd (real) λ_M, λ_m for B

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

pt See SAAD Chap 5 page 132

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

$$x \perp_A y \text{ if } x^T A y = 0$$

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

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

Thm A spd, error in steepest descent

$$\varepsilon^{(k)} = \bar{u} - u^{(k)} \text{ then}$$

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

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

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

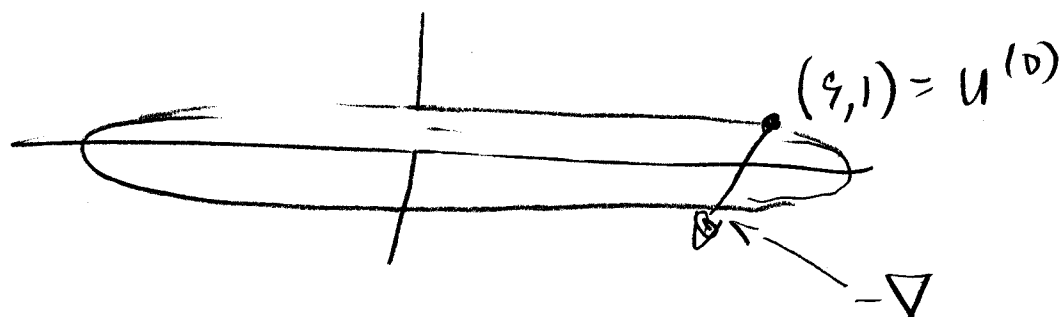
pt SAAID Chap 5 133 p.

Steepest Descent an example

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

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

"Elliptical Bowl"



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

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

$$u_2^{(1)} = 1 - \frac{1810}{1081} = \frac{-729}{1081} \approx -3/4$$

We over shot!

Krylov Subspaces

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

Def n th 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 $x^{(n)} = a_1 b + \dots + a_{n-1} A^{n-1} b$

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

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

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

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

Not clear how to do this!

Next best thing

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

This we can do! It is called Conjugate Gradient.

Conjugate Gradient Iteration

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

Iteration:

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

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

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

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

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

Claim $r_n = b - Ax_n$

pf induct on n $n=0$

assume true for $n-1$

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

$$= b - Ax_{n-1} - \alpha_n A p_{n-1}$$

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

Thm A CG for spd A for prob $Ax=b$ then while $r_n \neq 0$

$$\mathcal{K}_n = \langle X_1, \dots, X_n \rangle = \langle P_0, \dots, P_{n-1} \rangle$$

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

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

$$\& P_n^T A P_j = 0 \quad (j < n)$$

Pf See Trefethen & Bau Chap 38

Thm B $A^{m \times m} \bar{x} = b$, A spd, CG

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

2) $\| \epsilon_n \|_A \leq \| \epsilon_{n-1} \|_A$

3) $\exists n \leq m$ st $\epsilon_n = 0$

pf $x_n \in \mathcal{K}_n$ by previous thm.

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

$$\Delta x = x_n - x$$

$$\epsilon = \bar{x} - x = \epsilon_n + \Delta x$$

$$\| \epsilon \|_A^2 = (\epsilon_n + \Delta x)^T A (\epsilon_n + \Delta x)$$

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

note $A \epsilon_n = A(\bar{x} - x_n) = b - A x_n = r_n$

$$2 \epsilon_n^T A \Delta x = 2 r_n^T \Delta x = 0 \quad r_n \perp_A \mathcal{K}_n$$

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

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

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.

CG & Polynomial Approx

12

Prob Find poly P_n , $P_n(0)=1$ $\deg(P_n) \leq n$

minimizing $\|P_n(A) \varepsilon_0\|_A$

Thm If $r_{n-1} \neq 0$ for CG then $\exists! P_n$, $P_n(0)=1$

1) X_n has error $\varepsilon_n = P_n(A) \varepsilon_0$ $P_n(0)=1$

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

pf 1) See page 12A.

2) (=) follows from Thm B.

($\leq$) Suppose $\varepsilon_0 = \sum_{j=1}^m \alpha_j V_j$ (eigen expansion of A)

$$P(A) \varepsilon_0 = \sum \alpha_j P(\lambda_j) V_j$$

pf of 1)

By Thm A $X_n \in \mathcal{H}_n$ 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 $P_n(z) = 1 - z \cdot Q(z)$ $\deg(P_n) \leq n$

Claim $\varepsilon_n = P_n(A)\varepsilon_0$ where $\varepsilon_0 = \bar{X}$

$$P_n(A)\varepsilon_0 = (1 - A Q_{n-1}(A))\bar{X}$$

$$= \bar{X} - Q_{n-1}(A)(A\bar{X})$$

$$= \bar{X} - Q_{n-1}(A)b$$

$$= \bar{X} - X_n = \varepsilon_n$$

For: jacobe on quintana

File: people_csail_mit_edu_jacobe_papers_emnlp_unsup_2011_pdf

Printing date: Wed Sep 21 10:24:18 2011

Printer: ghc8206bw / GHC8206BW (interpreter version 3010.107)

ghc8206bw quintana:jacobe Job: people_csail_mit_edu_jacobe_papers_emnlp_unsup_2011_pdf D

ghc8206bw quintana:jacobe Job: people_csail_mit_edu_jacobe_papers_emnlp_unsup_2011_pdf D

ghc8206bw quintana:jacobe Job: people_csail_mit_edu_jacobe_papers_emnlp_unsup_2011_pdf D

ghc8206bw quintana:jacobe Job: people_csail_mit_edu_jacobe_papers_emnlp_unsup_2011_pdf D

$$\| \varepsilon_0 \|_A = \sum_{j=1}^m a_j^2 \lambda_j \quad \| e_n \|_A = \sum_{j=1}^m a_j^2 p_n(\lambda_j)^2 \lambda_j$$

$$\frac{\| \varepsilon_n \|_A}{\| \varepsilon_0 \|_A} \leq \max_j \frac{a_j^2 p_n(\lambda_j)^2 \lambda_j}{a_j^2 \lambda_j} = \max_j |p_n(\lambda_j)|$$

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

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

pf

The Chebyshev polynomial

$$\frac{\| \varepsilon_n \|_A}{\| \varepsilon_0 \|_A} \leq \max_{\lambda \in \lambda(A)} |p_n(A)| \leq \max_{\lambda \in \lambda(A)} |\overline{T}_n(A)|$$

$$\leq 2 \left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1} \right)^n$$