Lect 9 Solving Linear system (II)

一. 最速下降法: (Steepest decent method)

1. 设要求解 $A \cdot X = b$ A 对称正定 (S.P.D.)

等价的变分形式: 定义

$$\Phi(X) = \frac{1}{2} X^T A X - X^T b$$

则有:

$$-\nabla\Phi = b - AX, \quad \nabla^2\Phi = A$$

故

$$\min_X \Phi(X) \Longleftrightarrow \text{求解 } AX = b$$

2. 最速下降法:

$$\frac{dX}{dt} = -\nabla\Phi(X) \Rightarrow \frac{d\Phi(X(t))}{dt} = -|\nabla\Phi(X(t))|^2 \leq 0$$

且 $= 0 \Longleftrightarrow \nabla\Phi(X(t)) = 0$

3. 离散形式:

$$X^{k+1} = X^k + \alpha_k(-\nabla\Phi(X^k))$$

$$= X^k + \alpha_k r_k$$

$r_k \triangleq b - AX^k$ 称残量. (residual)

α_k的选取使得

$$\min_{\alpha_k} \Phi(X^{k+1}) = \frac{1}{2}(X^{k+1})^T \cdot A \cdot X^{k+1} - b^T \cdot X^{k+1}$$

有

$$\frac{d\Phi(X^{k+1})}{d\alpha_k} = r_k^T A r_k \cdot \alpha_k - r_k^T \cdot r_k = 0$$

$$\Rightarrow \alpha_k = r_k^T r_k \big/ r_k^T A r_k$$

即格式为:

$$\begin{cases} X^{k+1} = X^k + \alpha_k r^k \\ r^k = b - A \cdot X^k \\ \alpha_k = r_k^T r_k / r_k^T A r_k \end{cases}$$

4. 收敛性:

S.D.M 迭代矩阵 $B_{\alpha_k} = I - \alpha_k A$

引理: 设A特征值$\lambda_1 \geq \cdots \geq \lambda_n > 0$, 则当$\alpha_k \equiv \alpha$时,当且仅当

$0 < \alpha < 2/\lambda_1$ 时收敛,且

$$\rho_{opt} = \min_{\alpha}[\rho(B_\alpha)] = \frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n}$$

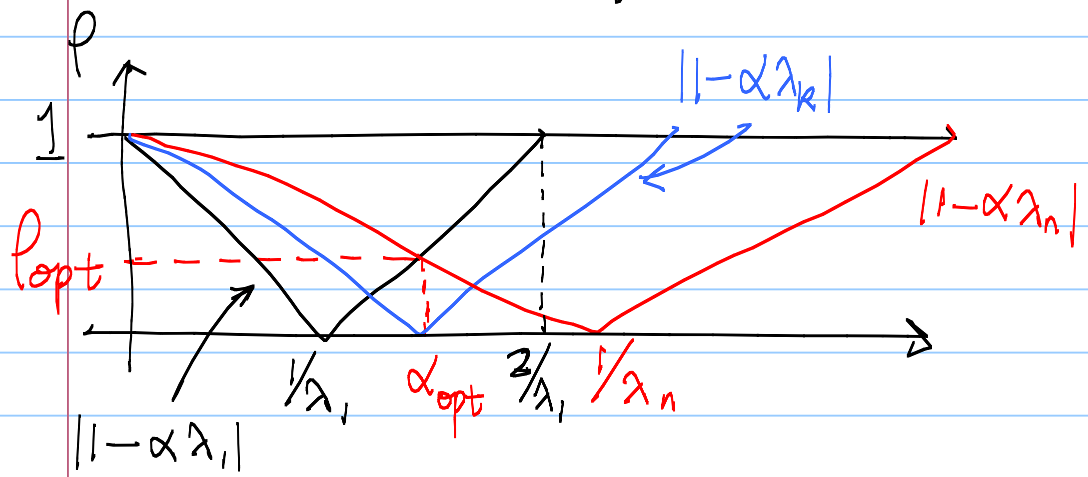
此时 $\alpha_{opt} = 2/(\lambda_1 + \lambda_n)$

09-03

解：收敛 $\Leftrightarrow \rho(B_\alpha)<1$ 即 $\max_i |1-\alpha\lambda_i|<1$

$\Rightarrow 2/\alpha > \lambda_i > 0 \quad i=1,\cdots,n$ 即 $0<\alpha<2/\lambda_1$.

由于 $\rho(B_\alpha)=\max_i |1-\alpha\lambda_i| \quad \forall \alpha$ 固定



由上图有 α_{opt} 对应于 $1-\alpha\lambda_n=\alpha\lambda_1-1 \Rightarrow \alpha_{opt}=2/(\lambda_1+\lambda_n)$

此时 $\rho_{opt}=(\lambda_1-\lambda_n)/(\lambda_1+\lambda_n)$

定理：对SDM, A为S.P.D. 有

$$\|e^{k+1}\|_A \leq \frac{K_2(A)-1}{K_2(A)+1}\|e^k\|_A$$

其中 $\|X\|_A \triangleq (X^TAX)^{1/2}=\|A^{1/2}X\|_2$

$A^{1/2}\triangleq P^T\Lambda^{1/2}P, \quad A=P^T\Lambda P, \quad K_2(A)=\mathrm{Cond}_2(A)$.

解：定义 $e_R^{k+1}=(X^k+\alpha_{opt}r^k)-X^*$，则

09-04

$$\|e_R^{k+1}\|_A = \|B_{\alpha_{opt}} e^k\|_A = \|A^{1/2} B_{\alpha_{opt}} e^k\|_2$$

$$\le \|A^{1/2} B_{\alpha_{opt}} A^{-\frac{1}{2}}\|_2 \, \|A^{1/2} e^k\|_2$$

$$= \rho(B_{\alpha_{opt}}) \cdot \|e^k\|_A = \frac{K_2(A)-1}{K_2(A)+1} \|e^k\|_A$$

由 $\Phi(x) - \Phi(x^*) = \frac{1}{2}\|x - x^*\|_A^2$

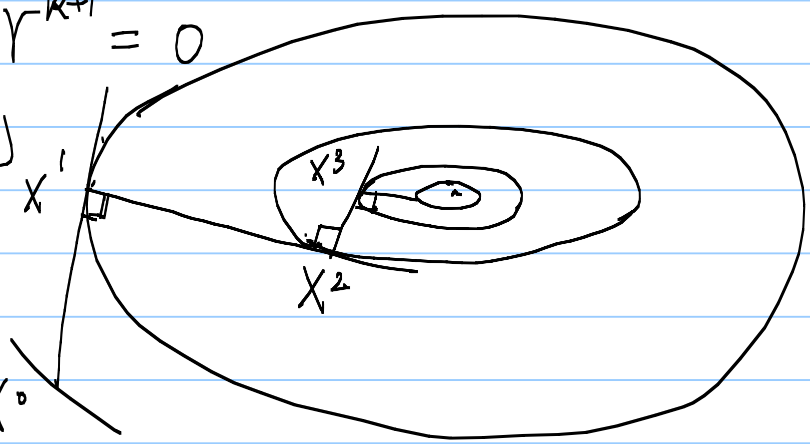
故 $\|e_R^{k+1}\|_A^2 \ge \|e^{k+1}\|_A^2$. #

RMK: ① 由 $r^k = -\nabla\Phi(x^k)$, $r^{k+1} = -\nabla\Phi(x^{k+1})$

$x^{k+1} = x^k + \alpha_k r^k$, 且 $\frac{d}{d\alpha}\Phi(x^{k+1})\Big|_{\alpha=\alpha_k} = 0$

$\Rightarrow \quad r^k \cdot r^{k+1} = 0$

几何直观



② $\lambda_1 \gg \lambda_n$ 时, 收敛慢

二. 共轭梯度法 (Conjugate Gradient Method)

1. 对最速下降法作改进.

09-05

最速下降法中，搜索方向 $p^k = r^k$

CGM 中， $p^k \in \mathrm{span}\{p^{k-1}, r^k\}$ $k \geq 1$

寻找 (ξ, η) s.t.

$$\min_{(\xi,\eta)} \psi(\xi,\eta) = \Phi(x^k + \xi r^k + \eta p^{k-1})$$

由 $\frac{\partial \psi}{\partial \xi} = \frac{\partial \psi}{\partial \eta} = 0$ 在 (ξ^*, η^*) 处

$$\xi^* (r^k)^T A p^{k-1} + \eta^* (p^{k-1})^T A p^{k-1} = 0$$

$$\Rightarrow \frac{\eta^*}{\xi^*} = -\frac{(r^k)^T A p^{k-1}}{(p^{k-1})^T A p^{k-1}}, \quad p^k = r^k + \frac{\eta^*}{\xi^*} p^{k-1}$$

2. 离散形式：

Step1：取 x^0，$p^0 = r^0 = b - Ax^0$，$k = 0$

Step2：设已有 x^k, p^k, r^k，则

Step2.1 求 $\alpha_k = (r^k)^T p^k / (p^k)^T A p^k$

Step2.2 计算

$$x^{k+1} = x^k + \alpha_k p^k$$

$$r^{k+1} = b - A x^{k+1}$$

（或 $r^{k+1} = r^k - \alpha_k A p^k$）

step2.3 计算 $\beta_k = -\frac{(r^{k+1})^T A p^k}{(p^k)^T A p^k}$, $p^{k+1} = r^{k+1} + \beta_k p^k$ Ⓐ

3. 基本性质及收敛性:

定理: 由CGM得到的向量组满足

① $(p^i)^T \cdot r^j = 0$, $0 \le i < j \le k$

② $(r^i)^T \cdot r^j = 0$, $0 \le i, j \le k$, $i \ne j$

③ $(p^i)^T A p^j = 0$, $0 \le i, j \le k$, $i \ne j$

④ $\mathrm{span}\{r^0, \cdots, r^k\} = \mathrm{span}\{p^0, \cdots, p^k\} = K_{k+1}(A; r^0)$

$K_{k+1}(A; r^0) \triangleq \mathrm{span}\{r^0, Ar^0, \cdots, A^k r^0\}$ 称Krylov子空间.

证: 采用归纳法; $k=1$时

$p^0 = r^0$, 由Ⓐ $(p^k)^T A p^{k+1} = 0$ $\forall k = 0, 1, \cdots$ Ⓑ

故 $(p^0)^T A p^1 = 0$

又由 $\nabla \Phi(x^1) \cdot r^0 = 0 \Rightarrow (r^1)^T \cdot r^0 = (r^1)^T \cdot p^0 = 0$

设 $\le k$时上式都成立, 则 $r^{k+1} = r^k - \alpha_k A p^k$ Ⓒ

① $(p^j)^T r^{k+1} = (p^j)^T r^k - \alpha_k (p^j)^T A p^k = 0$, $0 \le j \le k-1$

又 $(p^k)^T r^{k+1} = (p^k)^T r^k - \dfrac{(r^k)^T p^k}{(p^k)^T A p^k} (p^k)^T A p^k = 0$

② 由①知 $r^{k+1} \perp \mathrm{span}\{p^0, \cdots, p^k\}$，由假设④知②成立.

③ 由 $p^{k+1} = r^{k+1} + \beta_k p^k$ 及Ⓒ有

$$(p^j)^T A p^{k+1} = (p^j)^T A (r^{k+1} + \beta_k p^k) = 0$$

且 $(p^{k+1})^T A p^k = (r^{k+1} + \beta_k p^k)^T A p^k = 0$

故③成立.

④ 由Ⓐ有 $\mathrm{span}\{p^{k+1}, \cdots, p^0\} = \mathrm{span}\{r^{k+1}, p^k, \cdots, p^0\}$

$$= \mathrm{span}\{r^{k+1}, r^k, \cdots, r^0\}$$

由Ⓒ有 $\mathrm{span}\{r^{k+1}, \cdots, r^0\} = \mathrm{span}\{Ap^k, r^k, \cdots, r^0\}$

$$= \mathrm{span}\{Ap^k, A^k r^0, \cdots, r^0\}$$

$$= \mathrm{span}\{A^{k+1} r^0, \cdots, r^0\} = K_{k+2}(A, r^0).$$ #

由上可简化 α_k, β_k 的表达式.

09-08

$$\alpha_k=\frac{(r^k)^T p^k}{(p^k)^T A p^k}=\frac{(r^k)^T r^k}{(p^k)^T A p^k}$$

$$\beta_k=\frac{-1/\alpha_k (r^{k+1})^T r^{k+1}}{1/\alpha_k \ (r^k)^T \cdot r^k}=(r^{k+1})^T r^{k+1} / (r^k)^T r^k .$$

RMK: A.由定理知CGM理论上最多n步收敛，但实际计算中由于舍入误差，n步后如还未收敛，则需重启动；

B. CGM无可调参数.

定理：收敛性：

$$\|e^k\|_A \le 2\left(\frac{\sqrt{K_2(A)}-1}{\sqrt{K_2(A)}+1}\right)^k \|e^0\|_A .$$

4. 预处理共轭梯度法（PCG）

一如 $K_2(A) \gg 1$ 则收敛极慢，故引出预处理：

$$Ax=b \longrightarrow C^{-1}AC^{-1}\cdot Cx=C^{-1}b$$

$$\hat{A}\cdot\hat{x}=\hat{b}$$

C为S.P.D，C的选取使 $\hat{A}$ 为良态矩阵，最优的 $C=A^{1/2}$.

应用CG：

$$\alpha_k=\frac{(\hat{r}^k)^T\hat{r}^k}{(\hat{p}^k)^T A\cdot\hat{p}^k},\quad \hat{x}^{k+1}=\hat{x}^k+\alpha_k\hat{p}^k$$

$$\hat{r}^{k+1}=\hat{r}^k-\alpha_k\hat{A}\hat{p}^k$$

$$\beta_k=\frac{(\hat{r}^{k+1})^T\hat{r}^{k+1}}{(\hat{r}^k)^T\hat{r}^k},\quad \hat{p}^{k+1}=\hat{r}^{k+1}+\beta_k\hat{p}^k$$

令 $\hat{x}^k=C\cdot x^k$，$\hat{r}^k=C^{-1}r^k$，$\hat{p}^k=Cp^k$

令 $M=C^2$，得

$$w^k=Ap^k,\quad \alpha_k=\rho_k/(p^k)^T\cdot w^k$$

$$x^{k+1}=x^k+\alpha_k p^k,\quad r^{k+1}=r^k-\alpha_k w^k$$

$$z^{k+1}=M^{-1}r^{k+1},\quad \rho^{k+1}=(r^{k+1})^T\cdot z^{k+1}$$

$$\beta_k=\rho^{k+1}/\rho^k,\quad p^{k+1}=z^{k+1}+\beta_k p^k$$

$$r^0=b-Ax^0,\ z^0=M^{-1}r^0,\ \rho^0=(r^0)^T\cdot z^0,\ p^0=z^0$$

通常对M选为：

09-10

1) 对角预优：

$$M = \mathrm{diag}(a_{11}, \cdots, a_{nn})$$

(如对角元素相差较大)

2) 不完全 Cholesky 分解：$A = L \cdot L^T + R$

取 $M = L \cdot L^T$.

详见：Y. Saad, Iterative methods for sparse linear systems, SIAM.