

# Iterative 2

王胤雅

25114020018

yinyawang25@m.fudan.edu.cn

2025 年 10 月 20 日

## PROBLEM I

1. Prove that the iteration matrix  $G_\omega$  of SSOR, as defined by

$$G_\omega = (D - \omega F)^{-1}(\omega E + (1 - \omega)D)(D - \omega E)^{-1}(\omega F + (1 - \omega)D).$$

can be expressed as

$$G_\omega = I - \omega(2 - \omega)(D - \omega F)^{-1}D(D - \omega E)^{-1}A.$$

2. Deduce the expression below for the preconditioning matrix associated with the SSOR iteration.

$$M_{SSOR} = \frac{1}{\omega(2 - \omega)}(D - \omega E)D^{-1}(D - \omega F)$$

*SOLUTION.* 1. 由于  $\omega E + (1 - \omega)D = -(D - \omega E) + (2 - \omega)D$ ,  $\omega F + (1 - \omega)D = -(D - \omega F) + (2 - \omega)D$ , 考虑:

$$\begin{aligned} G_\omega &= (D - \omega F)^{-1}(\omega E + (1 - \omega)D)(D - \omega E)^{-1}(\omega F + (1 - \omega)D) \\ &= (D - \omega F)^{-1}(-(D - \omega E) + (2 - \omega)D)(D - \omega E)^{-1}(-(D - \omega F) + (2 - \omega)D) \\ &= (D - \omega F)^{-1}(-I + (2 - \omega)D(D - \omega E)^{-1})(-(D - \omega F) + (2 - \omega)D) \\ &= I - (2 - \omega)(D - \omega F)^{-1}D - (2 - \omega)(D - \omega F)^{-1}D(D - \omega E)^{-1}(D - \omega F) \\ &\quad + (2 - \omega)^2(D - \omega F)^{-1}D(D - \omega E)^{-1}D \\ &= I - (2 - \omega)(D - \omega F)^{-1}D(D - \omega E)^{-1}(D - \omega F + (2 - \omega)D + D - \omega E) \\ &= I - \omega(2 - \omega)(D - \omega F)^{-1}D(D - \omega E)^{-1}(D - E - F) \\ &= I - \omega(2 - \omega)(D - \omega F)^{-1}D(D - \omega E)^{-1}A. \end{aligned}$$

2.  $I - G_\omega = M_{SSOR}A$

□

**PROBLEM II** Let  $A$  be a matrix with a positive diagonal  $D$ .

(a) Obtain an expression equivalent to that of (4.13) for  $G_\omega$ , but which involves the matrices

$$S_E \equiv D^{-1/2}ED^{-1/2} \quad \text{and} \quad S_F \equiv D^{-1/2}FD^{-1/2}.$$

(b) Show that

$$D^{1/2}G_\omega D^{-1/2} = (I - \omega S_F)^{-1}(I - \omega S_E)^{-1}(\omega S_E + (1 - \omega)I)(\omega S_F + (1 - \omega)I).$$

(c) Now assume that, in addition to having a positive diagonal,  $A$  is symmetric. Prove that the eigenvalues of the SSOR iteration matrix  $G_\omega$  are real and nonnegative.

**SOLUTION.** 1.  $D$  的对角元均为正, 因此  $D^{1/2}$  与  $D^{-1/2}$  均存在。

注意以下恒等式:

$$\begin{aligned} D - \omega F &= D^{1/2}(I - \omega S_F)D^{1/2}, \\ D - \omega E &= D^{1/2}(I - \omega S_E)D^{1/2}, \\ \omega E + (1 - \omega)D &= D^{1/2}(\omega S_E + (1 - \omega)I)D^{1/2}, \\ \omega F + (1 - \omega)D &= D^{1/2}(\omega S_F + (1 - \omega)I)D^{1/2}. \end{aligned}$$

因此它们的逆可以写作:

$$\begin{aligned} (D - \omega F)^{-1} &= D^{-1/2}(I - \omega S_F)^{-1}D^{-1/2}, \\ (D - \omega E)^{-1} &= D^{-1/2}(I - \omega S_E)^{-1}D^{-1/2}. \end{aligned}$$

代入  $G_\omega$  得:

$$\begin{aligned} G_\omega &= (D^{-1/2}(I - \omega S_F)^{-1}D^{-1/2})(D^{1/2}(\omega S_E + (1 - \omega)I)D^{1/2}) \\ &= (D^{-1/2}(I - \omega S_E)^{-1}D^{-1/2})(D^{1/2}(\omega S_F + (1 - \omega)I)D^{1/2}) \\ &= D^{-1/2}(I - \omega S_F)^{-1}(\omega S_E + (1 - \omega)I)(I - \omega S_E)^{-1}(\omega S_F + (1 - \omega)I)D^{1/2} \end{aligned}$$

这一形式仅依赖于矩阵  $S_E, S_F$ , 并且是 (4.13) 的等价表达。

2. 由上一小问的结果可立马得到

$$D^{1/2}G_\omega D^{-1/2} = (I - \omega S_F)^{-1}(I - \omega S_E)^{-1}(\omega S_E + (1 - \omega)I)(\omega S_F + (1 - \omega)I).$$

3. 由于  $A$  对称, 那么  $E = F^T$ , 从而  $S_F = S_E^T$ 。由第二问的结论可知,

$$\begin{aligned} &D^{\frac{1}{2}}G_\omega D^{-\frac{1}{2}} \\ &= (I - \omega S_E^T)^{-1}(I - \omega S_E)^{-1}(\omega S_E + (1 - \omega)I)(\omega S_E^T + (1 - \omega)I)(I - \omega S_E^T)^{-1}(I - \omega S_E^T) \\ &= B^{-1}Q^TQB \end{aligned}$$

其中  $Q = (\omega S_E^T + (1 - \omega)I)(I - \omega S_E^T)^{-1}$ ,  $B = I - \omega S_E^T$ 。那么  $G_\omega = (BD^{\frac{1}{2}})^{-1}Q^TQ(BD^{\frac{1}{2}})$ , 即  $G_\omega$  相似于对称半正定矩阵  $Q^TQ$ 。从而  $G_\omega$  有实的非负特征值。

□

PROBLEM III A matrix of the form

$$B = \begin{pmatrix} 0 & E & 0 \\ 0 & 0 & F \\ H & 0 & 0 \end{pmatrix}$$

is called a three-cyclic matrix.

- (a) What are the eigenvalues of  $B$ ? (Express them in terms of eigenvalues of a certain matrix which depends on  $E, F$ , and  $H$ .)
- (b) Assume that a matrix  $A$  has the form  $A = D + B$ , where  $D$  is a nonsingular diagonal matrix, and  $B$  is three-cyclic. How can the eigenvalues of the Jacobi iteration matrix be related to those of the Gauss–Seidel iteration matrix? How does the asymptotic convergence rate of the Gauss–Seidel iteration compare with that of the Jacobi iteration matrix in this case?
- (c) Answer the same questions as in (b) for the case when SOR replaces the Gauss–Seidel iteration.
- (d) A matrix of the form

$$B = \begin{pmatrix} 0 & E_1 & 0 & \cdots & 0 \\ 0 & 0 & E_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 0 & E_{p-1} \\ E_p & 0 & \cdots & 0 & 0 \end{pmatrix}$$

is called a  $p$ -cyclic matrix, where the blocks  $E_1, \dots, E_p$  are conformable (for instance, each  $E_i$  is an  $m_i \times m_{i+1}$  block with  $m_{p+1} = m_1$ ).

SOLUTION. 1. 设矩阵

$$B = \begin{pmatrix} 0 & E & 0 \\ 0 & 0 & F \\ H & 0 & 0 \end{pmatrix},$$

其中各块矩阵的维度使得乘法  $EFH$  有定义。

设  $\lambda \neq 0$  是  $B$  的特征值, 存在非零向量

$$v = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

使得  $Bv = \lambda v$ , 即

$$\begin{cases} Ey = \lambda x, \\ Fz = \lambda y, \\ Hx = \lambda z. \end{cases}$$

从而,  $z = \lambda^{-1}Hx$ ,  $F(\lambda^{-1}Hx) = \lambda y \Rightarrow y = \lambda^{-2}FHx, E(\lambda^{-2}FHx) = \lambda x$ . 因此,  $EFHx = \lambda^3x$ . 若  $x \neq 0$ , 则  $\lambda^3$  为  $EFH$  的特征值。

另一方面, 若  $EFHx = \mu x$  且  $x \neq 0$ , 任取  $\lambda$  满足  $\lambda^3 = \mu$ , 定义

$$z = \lambda^{-1}Hx, \quad y = \lambda^{-2}FHx.$$

则有  $Ey = E(\lambda^{-2}FHx) = \lambda^{-2}EFHx = \lambda^{-2}\mu x = \lambda x$  并且  $Fz = F(\lambda^{-1}Hx) = \lambda^{-1}FHx = \lambda y, Hx = \lambda z$  因此

$$B \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \lambda \begin{pmatrix} x \\ y \\ z \end{pmatrix},$$

即  $\lambda$  为  $B$  的特征值。

综上所述, 矩阵  $B$  的非零特征值  $\lambda$  与矩阵  $EFH$  的特征值  $\mu$  满足

$$\lambda^3 = \mu.$$

2. 由于  $G_J = -D^{-1}B$ , 因此

$$G_J = - \begin{pmatrix} D_1^{-1} & 0 & 0 \\ 0 & D_2^{-1} & 0 \\ 0 & 0 & D_3^{-1} \end{pmatrix} \begin{pmatrix} 0 & E & 0 \\ 0 & 0 & F \\ H & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & D_1^{-1}E & 0 \\ 0 & 0 & D_2^{-1}F \\ D_3^{-1}F & 0 & 0 \end{pmatrix}$$

令  $P := D_1^{-1}ED_2^{-1}FD_3^{-1}H$ , 那么由上一问的结果可知,  $G_J$  的非零特征值  $\lambda$  满足  $-\lambda^3$  为  $P$  的特征值。

设

$$A = D + B, \quad B = \begin{pmatrix} 0 & E & 0 \\ 0 & 0 & F \\ H & 0 & 0 \end{pmatrix}, \quad D = \text{diag}(D_1, D_2, D_3),$$

其中每个  $D_i$  可逆。记 Gauss-Seidel 迭代矩阵为

$$G_{GS} = -(D + L)^{-1}U.$$

设  $\mu$  是  $G_{GS}$  的一个非零特征值, 且

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

是对应的分块特征向量。特征方程为

$$G_{GS}x = \mu x \iff -(D + L)^{-1}Ux = \mu x.$$

两边同乘以  $-(D + L)$  得到等价形式

$$Ux = -\mu(D + L)x.$$

从而,

$$Ex_2 = -\mu D_1 x_1, \quad (1)$$

$$Fx_3 = -\mu D_2 x_2, \quad (2)$$

$$0 = -\mu(D_3 x_3 + Hx_1). \quad (3)$$

从而得到:

$$x_3 = -D_3^{-1} Hx_1, \quad x_2 = \frac{1}{\mu} D_2^{-1} F D_3^{-1} Hx_1, \quad D_1^{-1} E D_2^{-1} F D_3^{-1} Hx_1 = -\mu^2 x_1.$$

故

$$E \left( \frac{1}{\mu} D_2^{-1} F D_3^{-1} Hx_1 \right) = -\mu D_1 x_1.$$

故

$$Px_1 = -\mu^2 x_1.$$

因此, 对于任一使  $x_1 \neq 0$ ,  $\mu$  必满足

$$\mu^2 = -\theta, \quad \text{其中 } \theta \in \text{eig}(P).$$

也就是说,  $G_{GS}$  的非零特征值是矩阵  $-P$  的二次根。

从而  $G_J$  的特征值  $\lambda$ , 总能找到  $G_{GS}$  的特征值  $\mu$  满足  $\lambda^3 = \mu^2$ .

下面进行渐近收敛率比较:

设

$$\Theta := \max\{|\theta| : \theta \in \text{eig}(P)\}.$$

则对于 Jacobi 迭代矩阵  $G_J = -D^{-1}B$ , 有  $\lambda^3 = -\theta$ , 因此

$$\max_{\lambda \in \text{eig}(G_J)} |\lambda| = \Theta^{1/3}.$$

而对于 Gauss-Seidel, 非零特征值满足  $\mu^2 = -\theta$ , 因此

$$\max_{\mu \in \text{eig}(G_{GS})} |\mu| = \Theta^{1/2}.$$

当 Jacobi 收敛, 即  $\Theta < 1$  时, 由  $0 \leq \Theta < 1$  有  $\Theta^{1/2} < \Theta^{1/3}$ , 因此

$$\max |\mu| < \max |\lambda|,$$

即 Gauss-Seidel 的特征值最大模更小, 故其渐近收敛速度优于 Jacobi。

3. 设  $\rho$  是 SOR 迭代矩阵  $G_\omega$  的一个非零特征值, 并令  $x = (x_1, x_2, x_3)^T$  为其对应的块特征向量, 其中每个  $x_i$  的维度与  $D_i$  相容。特征值方程

$$G_\omega x = \rho x$$

等价于

$$((1 - \omega)D + \omega U)x = \rho(D - \omega L)x.$$

将此关系按块形式展开:

$$(\text{第 1 行}) : \quad \omega E x_2 = (\rho - (1 - \omega)) D_1 x_1,$$

$$(\text{第 2 行}) : \quad \omega F x_3 = (\rho - (1 - \omega)) D_2 x_2,$$

$$(\text{第 3 行}) : \quad (1 - \omega) D_3 x_3 = \rho D_3 x_3 - \rho \omega H x_1.$$

即:

$$D_1^{-1} E x_2 = \frac{\rho - (1 - \omega)}{\omega} x_1, \quad D_2^{-1} F x_3 = \frac{\rho - (1 - \omega)}{\omega} x_2, \quad D_3^{-1} H x_1 = \frac{\rho - (1 - \omega)}{\rho \omega} x_3,$$

从而,

$$P x_1 = \frac{(\rho + \omega - 1)^3}{\omega^3 v} x_1, \quad \text{其中 } P := D_1^{-1} E D_2^{-1} F D_3^{-1} H.$$

由此可见,  $x_1$  是矩阵  $P$  的特征向量, 其对应的特征值为

$$\theta = \frac{(\rho + \omega - 1)^3}{\omega^3 v}.$$

最终的代数关系式:

$$\omega^3 \rho \theta = (\rho + \omega - 1)^3.$$

4. •  $B$  的非零特征值  $\lambda$  满足  $\lambda^p$  为  $E_1 \cdots E_p$  的特征值。
- $G_J$  的非零特征值  $\lambda$  满足  $(-1)^p \lambda^p$  为  $\prod_{i=1}^p D_i^{-1} E_i$  的特征值。
- $G_{GS}$  的非零特征值  $\lambda$  满足  $(-1)^p \lambda^{p-1}$  为  $\prod_{i=1}^p D_i^{-1} E_i$  的特征值。
- $G_\omega$  的非零特征值  $v$  满足  $\frac{(v+\omega-1)^p}{\omega^p v}$  为  $\prod_{i=1}^p D_i^{-1} E_i$  的特征值。

□

**PROBLEM IV** Consider the linear system  $Ax = b$ , where  $A$  is symmetric positive definite. Consider a projection step with  $K = L = \text{span}\{v\}$ , where  $v$  is some nonzero vector. Let  $x_{\text{new}}$  be the new iterate after one projection step from  $x$ , and let

$$d = A^{-1}b - x, \quad d_{\text{new}} = A^{-1}b - x_{\text{new}}.$$

(a) Show that

$$(Ad_{\text{new}}, d_{\text{new}}) = (Ad, d) - \frac{(r, v)^2}{(Av, v)},$$

where  $r = b - Ax$ . Does this equality establish convergence of the algorithm?

(b) (Gastinel's method.) In Gastinel's method the vector  $v$  is selected so that  $(v, r) = \|r\|_1$ ; for example one can take  $v$  with components  $v_i = \text{sign}(e_i^T r)$ , where  $r = b - Ax$  is the current residual. Show that

$$\|d_{\text{new}}\|_A \leq \left(1 - \frac{1}{n\kappa(A)}\right)^{1/2} \|d\|_A,$$

where  $\kappa(A)$  is the spectral condition number of  $A$ . Does this prove that the algorithm converges?

- (c) Compare the cost of one step of this method with that of cyclic Gauss–Seidel (see Example 5.1) and with that of an “optimal” Gauss–Seidel in which at each step  $K = L = \text{span}\{e_i\}$  and  $i$  is chosen to be the index of largest magnitude in the current residual vector.

SECTION. 1. 设投影在

$$K = L = \text{span}\{v\}$$

上, 令

$$x_{\text{new}} = x + \alpha v,$$

其中  $\alpha$  由正交投影条件知:

$$v^T(b - A(x + \alpha v)) = 0$$

即

$$v^T r - \alpha v^T A v = 0.$$

因此有

$$\alpha = \frac{(r, v)}{(Av, v)}.$$

故  $d_{\text{new}} = A^{-1}b - x - \alpha v = d - \alpha v$ . 下证明恒等式:

$$(Ad_{\text{new}}, d_{\text{new}}) = (A(d - \alpha v), d - \alpha v) = (Ad, d) - 2\alpha(Ad, v) + \alpha^2(Av, v).$$

由于  $Ad = r$ , 代入

$$\alpha = \frac{(r, v)}{(Av, v)}$$

得到

$$\begin{aligned} (Ad_{\text{new}}, d_{\text{new}}) &= (Ad, d) - 2\frac{(r, v)}{(Av, v)}(r, v) + \frac{(r, v)^2}{(Av, v)^2}(Av, v) \\ &= (Ad, d) - \frac{(r, v)^2}{(Av, v)}. \end{aligned}$$

下面说明关于收敛性的讨论: 由上式可见

$$\|d\|_A^2 = (Ad, d)$$

在每一步迭代中, 当  $(r, v) \neq 0$  时  $\|d\|_A^2$  减小, 当  $(r, v) = 0$ , 即  $v$  与残差正交时,  $\|d\|_A^2$  不变。因此该恒等式说明: 每一次投影步骤都不会使误差在范数意义下增大, 只有当  $v$  与当前残差正交时, 能量不再减小。故不能保证算法必然收敛到零误差, 因为如果在某些迭代步中总是选取与残差正交的  $v$ , 则算法可能停滞, 不再进步。因此, 为了保证整体收敛性, 必须对向量  $v$  的选择施加额外条件或采用合适的选取策略, 使得每一步迭代都能有效降低误差。

2.

$$(r, v) = \sum_i r_i \text{sign}(r_i) = \sum_i |r_i| = \|r\|_1.$$

由 (a) 的恒等式:

$$\|d_{\text{new}}\|_A^2 = \|d\|_A^2 - \frac{(r, v)^2}{(Av, v)}.$$

我们要估计  $\|d\|_A^2$  的下界:

分子下界:

$$(r, v)^2 = \left( \sum_i |r_i| \right)^2 \geq \left( \sum_i r_i^2 \right) = (Ad)^T (Ad) = d^T A^2 d,$$

分母上界: 设  $A \in \mathbb{R}^{n \times n}$  为对称正定矩阵, 记其特征值按照非降序为

$$0 < \lambda_{\min} = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n = \lambda_{\max},$$

$v$  的二范数满足

$$\|v\|_2^2 = n$$

故

$$(Av, v) \leq \lambda_{\max} \|v\|_2^2 = n \lambda_{\max}.$$

由此得到

$$\frac{(r, v)^2}{(Av, v)} \geq \frac{d^T A^2 d}{n \lambda_{\max}}.$$

下面估计  $d^T A^2 d$ :

取正交特征向量矩阵  $Q$  (即  $Q^T Q = I$ ), 使得

$$A = Q \Lambda Q^T, \quad \Lambda = \text{diag}(\lambda_1, \dots, \lambda_n).$$

由于  $d \in \mathbb{R}^n$ , 令  $y = Q^T d$ 。由于  $Q$  正交, 有  $\|y\|_2 = \|d\|_2$ 。由此可得

$$d^T A d = (Qd)^T \Lambda Qd = y^T \Lambda y = \sum_{i=1}^n \lambda_i y_i^2,$$

以及

$$d^T A^2 d = y^T \Lambda^2 y = \sum_{i=1}^n \lambda_i^2 y_i^2.$$

因为对所有  $i$  都有  $\lambda_i \geq \lambda_{\min}$ , 于是

$$\lambda_i^2 y_i^2 \geq \lambda_{\min} \lambda_i y_i^2,$$

故

$$\sum_{i=1}^n \lambda_i^2 y_i^2 \geq \lambda_{\min} \sum_{i=1}^n \lambda_i y_i^2.$$

即

$$d^T A^2 d \geq \lambda_{\min} d^T A d.$$

记  $A$ -范数为  $\|d\|_A := \sqrt{d^T A d}$ , 则上式可写为

$$d^T A^2 d \geq \lambda_{\min} d^T A d = \lambda_{\min} \|d\|_A^2$$

综上所述可得:

$$\frac{(r, v)^2}{(Av, v)} \geq \frac{\lambda_{\min}}{n \lambda_{\max}} \|d\|_A^2 = \frac{1}{n \kappa(A)} \|d\|_A^2,$$



其中  $\kappa(A) = \lambda_{\max}/\lambda_{\min}$ 。从而，

$$\|d_{\text{new}}\|_A^2 \leq \left(1 - \frac{1}{n\kappa(A)}\right) \|d\|_A^2,$$

两边取平方根得

$$\|d_{\text{new}}\|_A \leq \left(1 - \frac{1}{n\kappa(A)}\right)^{1/2} \|d\|_A.$$

因此，每一步的投影迭代在  $A$ -范数下的收缩因子至多为  $\left(1 - \frac{1}{n\kappa(A)}\right)^{1/2}$ 。只要  $\kappa(A) < 1$  序列收敛到精确解。

3. 每步 Gastinel 的主要代价是一次矩阵向量乘  $O(m)$ ，与一次完整的 Gauss-Seidel sweep 同阶；而“最优”逐坐标选择尽管每次坐标更新的局部效果更好，在选择索引上的额外开销通常总成本更高。

□

**PROBLEM V** Consider the iteration

$$x_{k+1} = x_k + \alpha_k d_k,$$

where  $d_k$  is a nonzero vector called the search direction and  $\alpha_k$  is a scalar. Assume a method that chooses  $\alpha_k$  so that the residual  $\|r_{k+1}\|_2$  is minimized, where  $r_k = b - Ax_k$ .

- (a) Determine  $\alpha_k$  that minimizes  $\|r_{k+1}\|_2$ .
- (b) Show that the residual vector  $r_{k+1}$  obtained in this manner is orthogonal to  $Ad_k$ .
- (c) Show that the residuals satisfy

$$\|r_{k+1}\|_2 \leq \|r_k\|_2 \sin \angle(r_k, Ad_k).$$

- (d) Assume that at each step  $k$  we have  $(r_k, Ad_k) \neq 0$ . Will the method always converge?
- (e) Now assume  $A$  is symmetric positive definite and choose  $d_k \equiv r_k$  at each step. Prove that the method converges for any initial guess  $x_0$ .

**SOLUTION.** 1. 由迭代公式

$$r_{k+1} = r_k - \alpha_k Ad_k,$$

我们希望选取合适的步长  $\alpha_k$  使得残差  $\|r_{k+1}\|_2$  最小。为此，考虑如下标量函数：

$$\phi(\alpha) = \|r_k - \alpha Ad_k\|_2^2.$$

展开：

$$\begin{aligned} \phi(\alpha) &:= (r_k - \alpha Ad_k, r_k - \alpha Ad_k) \\ &= \|r_k\|_2^2 - 2\alpha(r_k, Ad_k) + \alpha^2 \|Ad_k\|_2^2. \end{aligned}$$

对  $\alpha$  求导并令导数为零：

$$\phi'(\alpha) = -2(r_k, Ad_k) + 2\alpha \|Ad_k\|_2^2 = 0.$$

解得：

$$\alpha_k = \frac{(r_k, Ad_k)}{\|Ad_k\|_2^2}.$$

因此，取此  $\alpha_k$  可以保证下一步残差  $\|r_{k+1}\|_2$  在所有  $\alpha$  中最小。

2. 由迭代公式

$$r_{k+1} = r_k - \alpha_k Ad_k,$$

其中  $\alpha_k$  为上一问的结果，

$$\alpha_k = \frac{(r_k, Ad_k)}{\|Ad_k\|_2^2}$$

下面考察残差与方向的关系。

$$(r_{k+1}, Ad_k) = (r_k - \alpha_k Ad_k, Ad_k) = (r_k, Ad_k) - \alpha_k \|Ad_k\|_2^2.$$

将  $\alpha_k$  的表达式代入：

$$(r_{k+1}, Ad_k) = (r_k, Ad_k) - \frac{(r_k, Ad_k)}{\|Ad_k\|_2^2} \|Ad_k\|_2^2 = 0.$$

因此得出结论：

$$(r_{k+1}, Ad_k) = 0,$$

即新的残差  $r_{k+1}$  与  $Ad_k$  正交。

3. 由前面 (a) 中的公式可得残差平方范数的更新：

$$\begin{aligned} \|r_{k+1}\|_2^2 &= \|r_k - \alpha_k Ad_k\|_2^2 \\ &= \|r_k\|_2^2 - 2\alpha_k (r_k, Ad_k) + \alpha_k^2 \|Ad_k\|_2^2 \\ &= \|r_k\|_2^2 - \frac{(r_k, Ad_k)^2}{\|Ad_k\|_2^2}. \end{aligned}$$

令

$$\theta = \angle(r_k, Ad_k),$$

注意

$$(r_k, Ad_k) = \|r_k\|_2 \|Ad_k\|_2 \cos \theta.$$

代入上式得到：

$$\|r_{k+1}\|_2^2 = \|r_k\|_2^2 (1 - \cos^2 \theta) = \|r_k\|_2^2 \sin^2 \theta.$$

因此有：

$$\|r_{k+1}\|_2 = \|r_k\|_2 |\sin \theta| \leq \|r_k\|_2 \sin \angle(r_k, Ad_k),$$

4. 前面的公式可得：

$$\|r_{k+1}\|_2^2 = \|r_k\|_2^2 - \frac{(r_k, Ad_k)^2}{\|Ad_k\|_2^2} < \|r_k\|_2^2,$$

因此序列  $\{\|r_k\|_2\}$  是单调递减的，并且显然有下界 0，于是  $\{\|r_k\|_2\}$  一定收敛到某个极限

$$L \geq 0.$$

这并不意味着  $L = 0$ 。当矩阵  $A$  为 SPD 且在第  $k$  步选择搜索方向

$$d_k = r_k,$$

此时生成的搜索空间正好是 Krylov 子空间：

$$K_{k+1}(A, r_0) = \text{span}\{r_0, Ar_0, A^2r_0, \dots, A^k r_0\}.$$

由于迭代完全在 Krylov 子空间中进行，因此可以保证收敛到线性方程组的真解。

5. 在此情形下，每步迭代的搜索方向取为残差自身，即

$$d_k = r_k.$$

根据 (a) 的公式，步长为

$$\alpha_k = \frac{(r_k, Ar_k)}{\|Ar_k\|_2^2}.$$

由 (c) 式（将  $d_k = r_k$  代入）可得：

$$\|r_{k+1}\|_2^2 = \|r_k\|_2^2 \left(1 - \frac{(r_k, Ar_k)^2}{\|r_k\|_2^2 \|Ar_k\|_2^2}\right).$$

若  $A$  的最大，最小值特征值为

$$0 < \lambda_{\min} \leq \lambda_{\max},$$

设条件数  $\kappa = \lambda_{\max}/\lambda_{\min}$ ，则由 Kantorovich 不等式，

$$\frac{(r_k^T Ar_k)^2}{(r_k^T r_k)(r_k^T A^2 r_k)} \geq \frac{4\kappa}{(\kappa + 1)^2}.$$

即可得到

$$\|r_{k+1}\|_2^2 \leq \|r_k\|_2^2 \left(1 - \frac{4\kappa}{(\kappa + 1)^2}\right) = \|r_k\|_2^2 \frac{(\kappa - 1)^2}{(\kappa + 1)^2}.$$

取平方根：

$$\|r_{k+1}\|_2 \leq \frac{\kappa - 1}{\kappa + 1} \|r_k\|_2.$$

由于  $\frac{\kappa-1}{\kappa+1} \in [0, 1)$ ，这给出了一个与  $k$  无关的严格收缩因子，从而保证残差范数以线性速率收敛到 0。

又因为  $r_k = Ae_k$ ，其中  $e_k = x^* - x_k$ ，且  $A$  可逆， $r_k \rightarrow 0$  等价于  $x_k \rightarrow x^* = A^{-1}b$ 。因此，在  $A$  SPD 且  $d_k = r_k$  的情形下，对任意初始  $x_0$  都能保证收敛，其收敛速度由条件数  $\kappa(A)$  决定。

□