第二章第一节：范数

定义2.1.1（向量范数，p41）

$\|\cdot\|_1$，$\|\cdot\|_2$，$\|\cdot\|_\infty$, (p42)

定义2.1.2（矩阵范数，p43）

定理2.1.3（从向量范数诱导出的矩阵范数，p44）

定理2.1.4（矩阵范数例子，$\|\cdot\|_1$，$\|\cdot\|_2$，$\|\cdot\|_\infty$, p46）

定义2.1.4 （A的谱半径，$\rho(A)$, p48）

$$\rho(A) = \max\{|\lambda| : \lambda \in \lambda(A)\},$$

这里$\lambda(A)$代表A的特征值的全体。

定理2.1.6 （谱半径与矩阵范数关系, p49）

定理2.1.7 （A^k收敛的充分必要条件，p50）

$$\lim_{k\to\infty} A^k = 0 \Longleftrightarrow \rho(A) < 1.$$

定理2.1.8 （A的Neumann series, p51）

（1）$\sum_{k=0}^{\infty} A^k$收敛$\Longleftrightarrow \rho(A) < 1$。

（2）当$\sum_{k=0}^{\infty} A^k$收敛时，有

$$\sum_{k=0}^{\infty} A^k = (I-A)^{-1}.$$

且存在范数$\|\cdot\|$使得对于任意自然数m, 有

$$\|(I-A)^{-1} - \sum_{k=0}^{m} A^k\| \leq \frac{\|A\|^{m+1}}{1-\|A\|}.$$

第三章第一节：最小二乘问题

定义3.1.1（最小二乘问题，LS问题，p77）

定理3.1.1，定理3.1.2，推论3.1.1（p78-79）:
讨论了$Ax=b$的解的存在唯一性，和不唯一情形下解的集合表达式。

定理3.1.3（p80）：
线性最小二乘问题存在性，和唯一性的充要条件。

定理3.1.4（p80）：
最小二乘问题的正则化方程。对于正则化方程$A^TAx = A^Tb$和最小二乘解的关系，还可从变分法推出。定理3.1.4证明如下：

若x为最小二乘解，则$x = \min_{y\in\mathbb{R}^n} \|Ay-b\|_2^2$。
对于$\forall y \in \mathbb{R}^n$，我们可以定义$f(a) = \|A(x+ay)-b\|_2^2$，则有$f(a)$在0点取到最小值。若$f(a)$在$a=0$处可导，其必为零。我们计算$f'(0)$如下

$$\begin{aligned}
f'(0) &= \lim_{\epsilon\to 0} \epsilon^{-1}\Big(\|A(x+\epsilon y)-b\|_2^2 - \|Ax-b\|_2^2\Big) \\
&= \lim_{\epsilon\to 0} \epsilon^{-1}\Big(\|A(x+\epsilon y)\|_2^2 + \|b\|_2^2 - 2\langle A(x+\epsilon y), b\rangle - \|Ax\|_2^2 - \|b\|_2^2 + 2\langle Ax, b\rangle\Big) \\
&= \lim_{\epsilon\to 0} \epsilon^{-1}\Big(\|Ax\|_2^2 + \epsilon^2\|Ay\|_2^2 + 2\epsilon\langle Ax, Ay\rangle - 2\epsilon\langle Ay, b\rangle - \|Ax\|_2^2\Big) \\
&= \lim_{\epsilon\to 0} \epsilon\|Ay\|_2^2 + 2\langle Ax-b, Ay\rangle \\
&= 2\langle A^T(Ax-b), y\rangle \\
&= 0.
\end{aligned}$$

所以我们有$\forall y \in \mathbb{R}^n$, $\langle A^T(Ax-b), y\rangle = 0$。这意味着$A^TAx = A^Tb$。

若正则化方程成立，即x满足$A^TAx = A^Tb$，对于$\forall a \in \mathbb{R},\ y \in \mathbb{R}^n$，我们有

$$\begin{aligned}
&\|A(x+ay)-b\|_2^2 \\
&= a^2\|Ay\|_2^2 - 2a\langle Ax-b, Ay\rangle + \|Ax-b\|_2^2 \\
&= a^2\|Ay\|_2^2 - 2a\langle A^T(Ax-b), y\rangle + \|Ax-b\|_2^2 \\
&= a^2\|Ay\|_2^2 + \|Ax-b\|_2^2 \\
&\geq \|Ax-b\|_2^2.
\end{aligned}$$

由a, y的任意性，我们有x为$Ax=b$的最小二乘解。

定理3.1.5, 定理3.1.6（p82-83）:
敏度分析和最小二乘问题条件数。条件数定义为$\kappa_2(A)=\|A\|_2\|A^\dagger\|_2$。

第三章第二节：Householder变换和Givens变换

定义3.2.1 （p84）：
对于$w\in\mathbb{R}^n$且满足$\|w\|_2=1$，定义Householder变换$H\in\mathbb{R}^{n\times n}$如下：

$$H=I-2ww^T.$$

定理3.2.1（p84）：
Householder变换性质，对称性，正交性，对合性，反射性。

定理3.2.2和之后内容（p86-88）:
若要将向量x的第2到n个元素全变为零，选取

$$H=I-\beta vv^T,\ v=x\pm\|x\|_2e_1,\ \beta=2/(v^Tv).$$

考虑更一般的情形，若要将向量x的第$k+1$到j个元素变为零，令

$$H=I-\beta vv^T,\ v=(0,\cdots,0,x_k-\alpha,x_{k+1},\cdots,x_j,0,\cdots,0),\ \alpha^2=\sum_{i=k}^{j}x_i^2,\ \beta=2/(v^Tv).$$

Givens变换，也被称为平面旋转变换，即在(i,k)平面内将x顺时针旋转θ度。

$$G(i,k,\theta)=I+s(e_ie_k^T-e_ke_i^T)+(c-1)(e_ie_i^T+e_ke_k^T),$$

$$c=\cos\theta=x_i(x_i^2+x_k^2)^{-1/2},\ s=\sin\theta=x_k(x_i^2+x_k^2)^{-1/2},$$

则有$y=G(i,k,\theta)x,\ y_i=(x_i^2+x_k^2)^{1/2},\ y_k=0.$

第三章第三节：QR分解定理

定理3.3.1（QR分解，p91）

正交变换法的实现，理论部分（p92），算法部分（p95）。

第四章：古典迭代法

$$
\begin{aligned}
&Ax = b \\
\Longrightarrow &\text{令} A = A_1 + A_2,\ A_1 x = -A_2 x + b \\
\Longrightarrow &x = -A_1^{-1} A_2 x + A_1^{-1} b \\
\Longrightarrow &x = Mx + g.
\end{aligned}
$$

Jacobi迭代，Gauss-Seidel迭代，定理4.2.1，定理4.2.2，定理4.2.3（p105-107）；收敛性，收敛速度。

定义4.2.1，定义4.2.2（p113），定理4.2.8，推论4.2.1（p114-115）：对角占优，可约，可逆，正定这些定义之间的关系。

第五章第一节：最速下降法

$$Ax = b,\ \varphi(x) = x^T Ax - 2b^T x.$$

定理5.1.1（p138）方程组和二次泛函φ的关系。

最速下降法思路和实现：确定方向p_k，确定距离α_k。

$$p_k = r_k = b - Ax_k, \quad \alpha_k = \frac{r_k^T p_k}{p_k^T A p_k}, \quad x_{k+1} = x_k + \alpha_k p_k.$$

迭代的过程为

$$x_k \Longrightarrow (r_k, p_k, \alpha_k) \Longrightarrow x_{k+1} \Longrightarrow (r_{k+1}, p_{k+1}, \alpha_{k+1}) \Longrightarrow \cdots.$$

算法5.1.1（p141）最速下降法的算法实现

定理5.1.2（p141）最速下降法的收敛性

第五章第二、三、四、五节：共轭梯度法

与最速下降法的区别在于下山方向和距离的选取。

$$r_{k+1} = r_k - \alpha_k A p_k,\ \alpha_k = \frac{r_k^T r_k}{p_k^T A p_k},\ \beta_k = \frac{r_{k+1}^T r_{k+1}}{r_k^T r_k},$$

$$x_{k+1} = x_k + \alpha_k p_k,\ p_{k+1} = r_{k+1} + \beta_k p_k.$$

迭代的过程为

$$(x_k, p_{k-1}, \alpha_{k-1}) \Longrightarrow (r_k = b - Ax_k,\ \beta_{k-1}) \Longrightarrow (p_k, \alpha_k, x_{k+1}) \Longrightarrow \cdots$$

关于共轭梯度法系数α_k，β_k化简的证明

已知信息：$x_k = x_{k-1} + \alpha_{k-1} p_{k-1}$；$\alpha_k = r_k^T p_k / (p_k^T A p_k)$ $(P140, 5.1.3)$；
$p_k = r_k + \beta_{k-1} p_{k-1}$；$\beta_k = -(r_{k+1}^T A p_k)/(p_k^T A p_k)$.

从β_k信息出发，我们有
$r_k^T A p_{k-1} + \beta_{k-1} p_{k-1}^T A p_{k-1} = 0 \Longrightarrow (r_k^T + \beta_{k-1} p_{k-1}^T) A p_{k-1} = 0 \Longrightarrow p_k^T A p_{k-1} = 0$.

从α_k信息出发，我们有
$r_k^T p_{k-1} = (b - A x_k)^T p_{k-1} = [b - A(x_{k-1} + \alpha_{k-1} p_{k-1})]^T p_{k-1} = (r_{k-1} - \alpha_{k-1} A p_{k-1})^T p_{k-1}$
$= r_{k-1}^T p_{k-1} - \alpha_{k-1} p_{k-1}^T A p_{k-1} = 0$.

由$r_k^T p_{k-1} = 0$，我们有
$r_k^T r_{k+1} = r_k^T (b - A x_{k+1}) = r_k^T [b - A(x_k + \alpha_k p_k)] = r_k^T (r_k - \alpha_k A p_k)$
$= r_k^T (p_k - \beta_{k-1} p_{k-1} - \alpha_k A p_k) = r_k^T p_k - \alpha_k r_k^T A p_k$.
考虑到$r_k = p_k - \beta_{k-1} p_{k-1}$和之前证明过的$p_k^T A p_{k-1} = 0$还有$\alpha_k$定义，我们有
$r_k^T p_k - \alpha_k r_k^T A p_k = r_k^T p_k - \alpha_k p_k^T A p_k + \alpha_k \beta_{k-1} p_{k-1}^T A p_k = 0$.
因此，$r_k^T r_{k+1} = 0$.

对于α_k，从$p_k = r_k + \beta_{k-1} p_{k-1}$，$r_k^T p_{k-1} = 0$出发我们有
$r_k^T p_k - r_k^T r_k = r_k^T \beta_{k-1} p_{k-1} = 0$，
所以$\alpha_k = r_k^T p_k / (p_k^T A p_k) = r_k^T r_k / (p_k^T A p_k)$.

对于β_k，从$x_{k+1} = x_k + \alpha_k p_k$出发，我们有
$r_{k+1} = b - A x_{k+1} = b - A(x_k + \alpha_k p_k) = r_k - \alpha_k A p_k \Longrightarrow A p_k = \alpha_k^{-1} (r_k - r_{k+1})$.
考虑到$r_k^T r_{k+1} = 0$，我们有
$r_{k+1}^T A p_k = \alpha_k^{-1} r_{k+1}^T (r_k - r_{k+1}) = -\alpha_k^{-1} r_{k+1}^T r_{k+1}$.
考虑到$p_k = r_k + \beta_{k-1} p_{k-1}$，$r_k^T p_{k-1} = 0$，$r_k^T r_{k+1} = 0$，我们有
$p_k^T A p_k = \alpha_k^{-1} p_k^T (r_k - r_{k+1}) = \alpha_k^{-1} p_k^T r_k = \alpha_k^{-1} (r_k + \beta_{k-1} p_{k-1})^T r_k = \alpha_k^{-1} r_k^T r_k$.
所以$\beta_k = -(r_{k+1}^T A p_k)/(p_k^T A p_k) = r_{k+1}^T r_{k+1} / (r_k^T r_k)$.

综上，$\alpha_k = r_k^T r_k / (p_k^T A p_k)$，$\beta_k = r_{k+1}^T r_{k+1} / (r_k^T r_k)$.

定理5.2.1 证明（详细版）
对于结论$(1)-(4)$，易知对于$k=1$时都成立。使用归纳法，假设$k=n$时结论$(1)-(4)$成立，我们需要证明对于$k=n+1$时结论$(1)-(4)$成立。

对于（4）前半部分，由$r_{n+1}=p_{n+1}-\beta_n p_n$，有
$\mathrm{Span}\{r_0,\cdots,r_{n+1}\}=\mathrm{Span}\{p_0,\cdots,p_n,p_{n+1}-\beta_n p_n\}=\mathrm{Span}\{p_0,\cdots,p_n,p_{n+1}\}$.

对于（1），由归纳假设，我们有
$p_i^T r_{n+1}=p_i^T(r_n-\alpha_n A p_n)=p_i^T r_n-\alpha_n p_i^T A p_n=0,\ i=0,\cdots,n-1$。对于$p_n^T r_{n+1}$，我们有$p_n^T r_{n+1}=p_n^T(r_n-\alpha_n A p_n)$。由$\alpha_n$定义，得知$p_n^T r_{n+1}=0$。因此，结论（1）成立。

对于（2），由结论（1）可得r_{n+1}与$\mathrm{Span}\{p_0,\cdots,p_n\}$正交。再由结论（4）前半部分，有$r_{n+1}$与$\mathrm{Span}\{r_0,\cdots,r_n\}$正交。因此，结论（2）成立。

对于（3），首先我们有$p_{n+1}=r_{n+1}+\beta_n p_n$，$r_{i+1}=r_i-\alpha_i A p_i$，然后
$p_i^T A p_{n+1}=p_i^T A r_{n+1}+\beta_n p_i^T A p_n=\alpha_i^{-1}(r_i-r_{i+1})r_{n+1}+\beta_n p_i^T A p_n$.
由结论（2）和归纳假设，我们有$p_i^T A p_{n+1}=0$对于$i=0,\cdots,n-1$。
对于$p_n^T A p_{n+1}$，我们有$p_n^T A p_{n+1}=p_{n+1}^T A p_n=(r_{n+1}+\beta_n p_n)^T A p_n=0$。之前等式用到了$\beta_n$定义。因此结论（3）成立。

对于（4）后半部分，首先结论（2）和结论（4）前半部分给出了$\{r_0,\cdots,r_{n+1}\}$和$\{p_0,\cdots,p_{n+1}\}$的线性无关性质。再由$r_{n+1}=r_n-\alpha_n A p_n$，我们有
$\mathrm{Span}\{r_0,\cdots,r_{n+1}\}=\mathrm{Span}\{r_0,\cdots,Ap_n\}=\mathrm{Span}\{p_0,\cdots,p_n,Ap_n\}$
$=\mathrm{Span}\{r_0,\cdots,A^n r_0,Ap_n\}\subset\mathrm{Span}\{r_0,\cdots,A^n r_0,A^{n+1}r_0\}$.
对于$A^{n+1}r_0$，我们有
$A^{n+1}r_0=A(A^n r_0)\in\mathrm{Span}\{Ap_0,\cdots,Ap_n\}=\mathrm{Span}\{Ar_0,\cdots,A^n r_0,Ap_n\}$
$\subset\mathrm{Span}\{r_0,\cdots,A^n r_0,Ap_n\}$.
从而可得$\mathrm{Span}\{r_0,\cdots,A^n r_0,A^{n+1}r_0\}\subset\mathrm{Span}\{r_0,\cdots,A^n r_0,Ap_n\}$。
因此，$\mathrm{Span}\{r_0,\cdots,r_{n+1}\}=\mathrm{Span}\{r_0,\cdots,A^n r_0,A^{n+1}r_0\}$. 结论（4）成立。

算法5.2.1（p146）共轭梯度法的算法实现

定理5.2.1（p146）共轭梯度法的性质，Krylov子空间定义。

定理5.2.2（p148）共轭梯度法和Krylov子空间关系。

算法5.3.1（p150）实用共轭梯度法，其本质就是限制迭代次数。

定理5.3.1，定理5.3.2（p150-151）共轭梯度法收敛性，迭代次数和误差估计。

第六章第一节：线性代数知识回顾

特征值，特征向量，特征多项式，谱集，重数，单特征值，重特征值，半单特征值，非亏损

相似变换：$B = XAX^{-1}$，且X非奇异。

定理6.1.1：（Jordan分解定理，p162）

定理6.1.2：（Schur分解定理，p163）

定理6.1.3：（Gerschgorin圆盘定理，p163）

特征值、特征向量条件数：$\|y\|_2$，$\|\Sigma^{\perp}\|_2$。其中$y^TA = \lambda y^T$，y为A的属于λ的左特征向量。

实对称矩阵特征值均为实数。如何证明？

对于任意多项式$P(x) = \sum_{j=0}^{n} a_j x^j$，其友矩阵的特征多项式为$P$本身。如何证明？

第二三节：幂法、反幂法

幂法算法：（6.2.6），定理6.2.1（幂法收敛性），反幂法算法（6.3.1）

第四节：QR方法

迭代算法：（6.4.1），定理6.4.1（收敛性定理）

章节6.4.2：实数情形

实Schur标准形（6.4.11），定理6.4.2（实Schur分解），上Hessenberg矩阵，上Hessenberg分解（p181），隐式QR算法（算法6.4.3）。

定理6.2.1证明补充：

之前的推导为若当型的基本应用，补充部分从(6.2.8)开始，即

$$\lim_{k\to\infty}\frac{1}{\lambda_1^k}A^k u_0 = X_1 y_1.$$

且因为u_0在λ_1的特征子空间上投影不为零，我们有$X_1 y_1 \neq 0$。由以上结果，我们可以得到

$$X_1 y_1 = \lim_{k\to\infty}\frac{1}{\lambda_1^k}A^k u_0 = \lim_{k\to\infty}\frac{A}{\lambda_1}\frac{A^k}{\lambda_1^k}u_0 = \frac{A}{\lambda_1}(X_1 y_1).$$

这意味着$X_1 y_1$为A关于λ_1的特征向量。接下来我们证明u_k的极限就是常数乘以$X_1 y_1$。

关于幂法(6.2.6)，我们需要增加一个额外条件，即当Au_{k-1}存在正负模最大分量时，默认μ_k为正。否则未必能够得到u_k的收敛性。由幂法(6.2.6)得知

$$u_k = \frac{Au_{k-1}}{\mu_k} = \frac{A^k u_0}{\mu_k\cdots\mu_1} = \frac{A^k u_0}{\lambda_1^k}\frac{\lambda_1^k}{\mu_k\cdots\mu_1},$$

且$\frac{\mu_k\cdots\mu_1}{\lambda_1^k}$为$\frac{A^k u_0}{\lambda_1^k}$的模最大分量。由红字部分可知当$\frac{A^k u_0}{\lambda_1^k}$拥有正负模最大分量时，$\frac{\mu_k\cdots\mu_1}{\lambda_1^k}$符号为确定的。因此，由$\frac{A^k u_0}{\lambda_1^k}$的收敛性可以得到$\frac{\mu_k\cdots\mu_1}{\lambda_1^k}$的收敛性。记$\zeta = \lim\frac{\mu_k\cdots\mu_1}{\lambda_1^k}$，我们有

$$\lim_{k\to\infty} u_k = X_1 y_1/\zeta.$$

已证出u_k收敛到A关于λ_1的一个特征向量。再由

$$\zeta = \lim_{k\to\infty}\frac{\mu_{k+1}}{\lambda_1}\frac{\mu_k\cdots\mu_1}{\lambda_1^k} = \lim_{k\to\infty}\frac{\mu_{k+1}}{\lambda_1}\zeta$$

可得$\lim_{k\to\infty}\mu_k = \lambda_1$。

第七章第一节：关于对称矩阵的性质

定理7.1.1：谱分解（对角化）。
定理7.1.2：极大极小定理（关于特征值）。
定理7.1.3，定理7.1.4：特征值特征向量敏感性。
定理7.1.5：SVD分解。

第二节：QR方法
三对角化$\Longrightarrow$隐式对称QR迭代$\Longrightarrow$隐式对称QR算法

第三节：Jacobi方法（和Givens变换有相似之处）
旋转曲面的选择和旋转角度的选取（7.3.2），（7.3.6），（7.3.7）。

第四节：二分法（针对实对称三对角阵，证明较难）
顺序主子式估计：定理7.4.1。
符号变化次数和特征值个数的估计：定理7.4.2，推论7.4.1，算法7.4.1。

第五节：分而治之法（分块，针对实对称三角阵）
如何分块：（7.5.3）
粘合：定理7.5.1，定理7.5.2。
过程简述：p231-232。

第六节：奇异值分解计算（非方阵，算法7.6.3）

P240习题2：

$\forall k$, 我们要证D_k中包含特征值。若$\alpha_{k,k}$为特征值，则结论显然成立。

考虑$\alpha_{k,k}$不为特征值情形。有$A-\alpha_{k,k}I$可逆，且

$$(A-\alpha_{k,k}I)e_k=[\alpha_{1,k};\alpha_{2,k};\cdots;0;\alpha_{k+1,k};\cdots;\alpha_{n,k}].$$

因为A是实对称的，考虑分解$A=Q^T\ diag(\lambda_1,\cdots,\lambda_n)\ Q$，其中$Q$为正交阵，将分解代入上述等式，有

$$\begin{aligned}&Q^T\ diag(\lambda_1-\alpha_{k,k},\cdots,\lambda_n-\alpha_{k,k})\ Qe_k=[\alpha_{1,k};\alpha_{2,k};\cdots;0;\alpha_{k+1,k};\cdots;\alpha_{n,k}]\\ \Longrightarrow &e_k=Q^T diag((\lambda_1-\alpha_{k,k})^{-1},\cdots,(\lambda_n-\alpha_{k,k})^{-1})Q[\alpha_{1,k};\alpha_{2,k};\cdots;0;\alpha_{k+1,k};\cdots;\alpha_{n,k}].\end{aligned}$$

在以上等式两端同时取2范数，且考虑到$\|Q\|_2=\|Q^T\|_2=1$，则有

$$\begin{aligned}1=\|e_k\|_2&\le\|diag((\lambda_1-\alpha_{k,k})^{-1},\cdots,(\lambda_n-\alpha_{k,k})^{-1})\|_2\ \|[\alpha_{1,k};\alpha_{2,k};\cdots;0;\alpha_{k+1,k};\cdots;\alpha_{n,k}]\|_2\\ &=\max\{|(\lambda_j-\alpha_{k,k})^{-1}|:j=1,\cdots,n\}\ (\sum_{j\ne k}\alpha_{j,k}^2)^{1/2}.\end{aligned}$$

由上可得，$(\sum_{j\ne k}\alpha_{j,k}^2)^{1/2}\ge\min\{|(\lambda_j-\alpha_{k,k})|:j=1,\cdots,n\}$。证毕。
